Information Field Theory with INTEGRAL/SPI data

Mahsa Ghaempanah

München 2017

Information Field Theory with INTEGRAL/SPI data

Mahsa Ghaempanah

Dissertation der Fakultät für Physik
der *Ludwig–Maximilians–Universität* München
für den Grad des
Doctor rerum naturalium

vorgelegt von
Mahsa Ghaempanah
aus Iran

München, 20.01.2017

Erstgutachter: PD Dr. Torsten Enßlin
Zweitgutachter: PD Dr. Martin Kerscher
Tag der mündlichen Prüfung: 15.03.2017

Contents

List of Figures

List of Tables

Acronyms

In the following is a list of acronyms frequently appearing in the thesis:

D^{3}PO	Denoising, Deconvolving and Decomposing Photon Observations
IFT	Information Field Theory
WF	Wiener Filter
MLE	Maximum likelihood estimate
MEM	Maximum Entropy method
ACS	Anticoincidence shield
FWHM	Full width at half maximum
INTEGRAL	International Gamma-Ray Astrophysics Laboratory
ISM	Inter-Stellar Medium

Zusammenfassung

Diese Arbeit beschäftigt sich mit bildgebenden Verfahren auf dem Gebiet der Gammastrahlungs-Astrophysik. Bestehende Methoden der Bildgebung werden weiter entwickelt um eine neue und hoffentlich bessere Himmelskarte der Emission der galaktischen Positronenannihilation 511 keV zu generieren. Der dabei entwickelte Algorithmus basiert auf der Informationsfeldtheorie (IFT). IFT erlaubt neben der alleinigen Analyse der Daten auch das Nutzen von A-priori-Information über das Signal in Form einer A-priori-Wahrscheinlichkeit. Beide, die A-priori-Information und die Information der Messdaten, werden in Form von Wahrscheinlichkeitsverteilungen formuliert und zu einer A-posteriori-Wahrscheinlichkeitsverteilung zusammengefasst. In dieser Arbeit benutzen wir zwei solcher Algorithmen, den Wiener-Filter (WF) und D^{3}PO, um die gemessenen Daten des SPI Spektrometers des INTEGRAL-Satelitten zu analysieren.

Schon während der Entwicklung dieser Methoden werden diese kontinuierlich an beispielhaften Datensätzen getestet. Dabei konnte gezeigt werden, dass die entwickelte Technik die 511keV Linie der Emission der galaktischen Positronenannihilations trotz eines starken Untergrunds an Detektorereignissen auflösen kann.

Weiterhin werden beide Verfahren, WF und D^{3}PO, auf den gesamten elfjährigen Datensatz von SPI in einem Energieband von 510.5 511 keV angewendet. Diese erlauben die Rekonstruktion der Emission des gesamten Himmels. Dabei wird die Morphologie der galaktischen Emission aufgrund von Positronenannihilation bei 511 keV enthüllt. Die rekonstruierte Karte zeigt, dass die Emission am stärksten innerhalb einiger Grad rund um das galaktische Zentrum ist. Die logarithmische Darstellung dieser Karte deutet auf einen geringen Beitrag der galaktischen Scheibe hin.

Durch Bildgebung mittels eines gröeren Energiebandes, von 500 bis 520 keV, können wir diese Entdeckungen bestätigen. Die Ergebnisse des D^{3}PO-Algorithmus zeigen eine starke Emission innerhalb des galaktischen Zentrums und nur ein sehr schwache aus der galaktischen Scheibe, welche in Richtung der galaktischen Ebene ausgedehnter ist verglichen mit der rekonstruierten Emission innerhalb des Energieband von 510.5 bis 511 keV.

Sowohl während der Tests mit simulierten Datensätzen als auch mit echten SPI Daten wurde zur Modellierung von Untergrundereignissen das Modell von Siegert et al. (2016) benutzt. Im abschlieenden Teil dieser Arbeit wurde durch einen modifizierten D^{3}PO Algorithmus simultan die zeitliche Veränderung des Untergrundes des SPI Detektors sowie die Himmelskarte bei 511 keV Emission rekonstruiert, ohne dabei zeitliche Entwicklung von Siegerts Model zu nutzen. Dabei konnte der Überschuss an positronenannihilations Emission aus dem galaktischen Zentrum bestätigt werden.

Abstract

This thesis addresses imaging techniques in the domain of γ-ray astronomy. We advance imaging to provide a new and hopefully improved sky map of the Galactic positron annihilation emission at 511 keV. To this end, we derive our algorithm within the probabilistic language of Information Field Theory (IFT). Within the IFT framework, the prior knowledge on the signal is added to the measurement information. Both, our previous knowledge and the information from our measurement, are formulated as probability distributions and they are merged to a posterior probability distribution. In this thesis, we use two algorithms that result from such information imaging, the Wiener Filter and D^{3}PO to process the data from the SPI spectrometer on board of the Space telescope INTEGRAL.

In the process of building up our data analysis, we perform a series of mock data simulations to test the performance of methods. Results from the simulations show that the technique is able to infer the 511 keV positron annihilation emission of the Galaxy, despite this weak signal is overlayed by a dominant instrumental background.

We further apply both imaging methods to the 11 years of all sky data from SPI within energy range of 510.5-511 keV. We uncover the morphology of the Galactic positron annihilation emission at 511 keV. The reconstructed map shows the emission is brightest in a region of a few degrees around the Galactic center. However a logarithmic scaled version of the map reveals a faint contribution of the Galactic disk.

We confirm these findings by using in a wider range of energies, from 500 to 520 keV. The result by D^{3}PO shows that the emission is intense in the Galactic center and has a faint Galactic disk contribution, which is more extended toward the plane compared to the emission reconstructed only from the energy range of 510.5-511 keV.

In all mock and real reconstructions performed so far, a background model by Siegert et al. (2016) is used. In the final work of this thesis, through the use of a modified version of the D^{3}PO, we reconstruct simultaneously the time component of the SPI background and the sky map of 511 keV emission, independently of the time evolution of the Siegert's background model. We therby confirm the presence of the Galactic center positron excess emission.

Chapter 1

Introduction

1.1 Gamma-ray production and positron annihilation in the Galaxy

Cosmic γ-rays inform us about high energy phenomena that are related to non-thermal processes in the Universe, including acceleration, propagation and radiation of relativistic particles. Cosmic γ-rays permit the study of compact objects (e.g., pulsars, neutron stars, microquasar), galactic diffuse emission, jets, lobes of radio galaxies, and galaxy clusters. Photons in the energy range above 100 keV are referred as γ-rays. While the higher energies are correlated to the highest energy particles observed among the cosmic rays, the lower bound is associated with the region of nuclear γ-ray lines (Aharonian, Bergström & Dermer, 2013). For the analysis of INTEGRAL/SPI data (10 keV-8 MeV), we concentrate our study to the low energy domain of the cosmic γ-ray electromagnetic radiations. The astrophysical phenomena typically investigated in the INTEGRAL/SPI energy ranges are nucleosynthesis of heavy elements related to supernovae, γ-ray burst, solar flares, interactions of sub-relativistic cosmic rays with the interstellar gas and dust, production and annihilation of positrons and more.

Processes which result in γ-ray production can be listed as follows:

1. Accelerated charged particles: When an energetic electron passes through a magnetic field (which fills entirely the Inter-Stellar Medium (ISM)), photons are produced as the result of the magnetic deflection. For the γ-ray regime, this deflection might happen in the vicinity of a highly magnetized sources such as neutron stars. The produced emission is called magnetic bremsstrahlung, a synchrotron radiation.

2. Compton scattering: photons traverse a long path within the interstellar space from their sources to the observers. During their journey, photons might gain or lose part of their energies, depending on the energy of the plasma filling the ISM. Typically γ-rays are more energetic than the plasma, so γ-rays will scatter off electrons and lose part of their energies to the plasma. The process of the production of lower energy photon is called normal Compton scattering. But if the plasma is more energetic than the incident photons, they are promoted in energy. This is the Inverse Compton scattering process which can

convert low energy photons into γ-rays.

3. De-exicitation of atomic nuclei: nucleons have quantized states of energy with typical energy difference of $\sim$ MeV. Many transition between levels involve γ-rays emission.

4. Pair production: in the regime below 1.022 MeV, γ-rays in presence of a sufficient strong electromagnetic field are produced in the important process of pair annihilation, in which an electron-positron pair annihilates producing two photons. Due to the momentum conservation, these photons move in opposite directions in the rest frame of the reaction with a characteristic rest mass energy of the positron and electron, 511 keV.

The first prediction of the positron dates back to 1931, when Dirac proposed an antiparticle for the electron with the same mass but opposite charge (Dirac, 1931). The first discovery of positrons in our Galaxy was in 1972 when Johnson in a balloon experiment detected the 511 keV line emission from the Galactic center (Johnson, Harnden & Haymes, 1972).

Galactic positrons might annihilate into γ-ray emission, after losing their initial energies through interactions with the ISM. However before cooling down to lower energies, positrons might annihilate directly in-flight with ISM electrons. According to the conservation laws, two γ-rays will be produced that fly in opposite directions in the rest frame of the annihilation. This in-flight annihilation has a spectrum peaking at 511 keV and leads mostly to a line component of the γ-ray emission.

In the indirect annihilation, the positrons cool down to thermal energies, and form an unstable bound system with ISM electrons, the positronium, prior to annihilation. The positronium was predicted in 1934 by a Croatian physicist, Mohorovicic. The positronium has a mass of 1.022 MeV and is similar to hydrogen, but with the proton replaced by a positron. Depending on the relative spin orientation of the electron and positron, the positronium takes two states: para-positronium (p-Ps) with anti-parallel and total spin of S = 0, which constitutes 25% of the positronium, and ortho-positronium (o-Ps) with parallel spins and S = 1 that forms 75% of the positronium . The p-Ps decays into two γ-rays of 511 keV energy in opposite directions, contributing to the line component of the γ-ray emission, and o-Ps result in three photons which form continuum emission. The corresponding lifetimes before annihilation are 1.2×10^{-10} sec and 1.4×10^{-7} sec for p-Ps and o-Ps, respectively. As positrons annihilate in the ISM via positronium, a superposition of a narrow line at 511 keV and a continuum emission below 511 keV is expected. Because of the poor spectral resolution of the initially used Nal detectors and their low signal-to-noise ratios, these two contributions could not be separated in the data of the first detection of balloon flights and were reported as a single emission peak at the energy of $\sim$ 490 keV (Prantzos et al., 2011). However according to later analysis (Leventhal et al., 1982) the separation of the line and continuum component of the emission became possible and comparison of their intensities implies that 92% of the annihilation occurs after the formation of the positronium and only 8% comes from the direct annihilation.

In a steady state and ignoring positron transportation, the number of positrons annihilated in the Galactic disk and more intense in the bulge should be equal to the number of produced positrons in the Galaxy. In other words, the spatial morphology of the γ-ray emission should reflects the spatial distribution of the sources provided that positrons an-

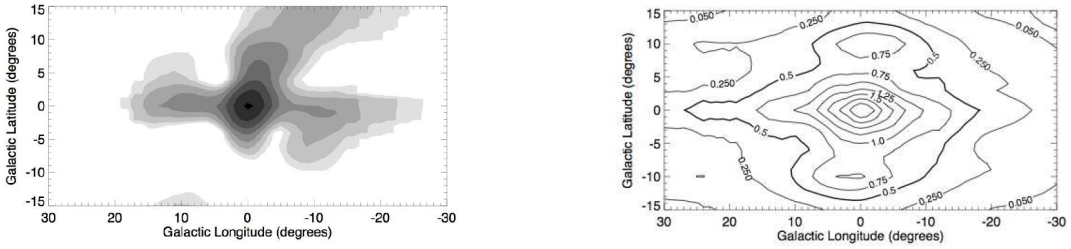

Figure 1.1: OSSE/CGRO 511 keV line map of the Galactic center region (left panel) and corresponding exposure map (right panel) by Purcell et al. (1997)

nihilate close to their production sites. Therefore, the imaging of the Galaxy at 511 keV emission is the best way to identify the Galactic positron sources. The spectral properties of the 511 keV line measured by SPI (Jean et al., 2006) suggests that a large fraction of the positrons generated in compact objects should propagate in the ISM before annihilation. The sources of the Galactic positrons are uncertain, but they are likely contained within the observed emission region. Nevertheless studies of Ferrière, Gillard & Jean (2007); Jean et al. (2006) suggest that typically the Galactic positrons do not diffuse more than about 100 kpc from their sources . Although the physical processes responsible for producing positrons are quite understood, it is hard to find a model that explains both the high observed rate of annihilation and spatial distribution in the Galaxy.

Until the advent of CGRO/OSSE (Compton Gamma Ray Observatory/Oriented Scintillation Spectrometer Experiment on GRO satellite), the map of the spatial distribution of the Galactic 511 keV emission was mainly based on theoretical expectation and poorly constrained. However, nine years of OSSE observations drastically improved this situation. The first constraints on the Galactic γ-ray emission at 511 keV was performed in 1990s through the OSSE/CGRO measurements, which showed that the emission is strongly concentrated toward the Galactic bulge. Purcell et al. (1997) mapped the 511 keV line emission for the first time (Fig. 1.1) revealing a symmetrical emission, centered on the Galactic bulge, and contribution of the Galactic plane. A third component was detected at Galactic coordinates of $l \sim -2$ °and $b \sim 12$ °. It is also interesting to note the work of Diehl (1994) through CGRO/COMPTEL observations, which suggest the central peak of the Galactic 511 keV emission is located at $l \sim 2$ °. This is interpreted as the slight asymmetry of the emission.

Other models for the spatial distribution of the Galactic positron annihilation emission were proposed by Prantzos (1993); Kinzer et al. (1996); Purcell et al. (1997); Kinzer et al. (2001); Milne et al. (2001). Each proposed model had in common a two component emission: a spheroid located in the inner Galaxy and the extended Galactic disk. Despite the significant progress which was made by the CGRO/OSSE observations, the data did not constrain the morphology of the 511 keV emission well enough to clarify the underlying source(s) population.

The INTEGRAL (INTErnational Gamma Ray Astrophisics Laboratory) observatory

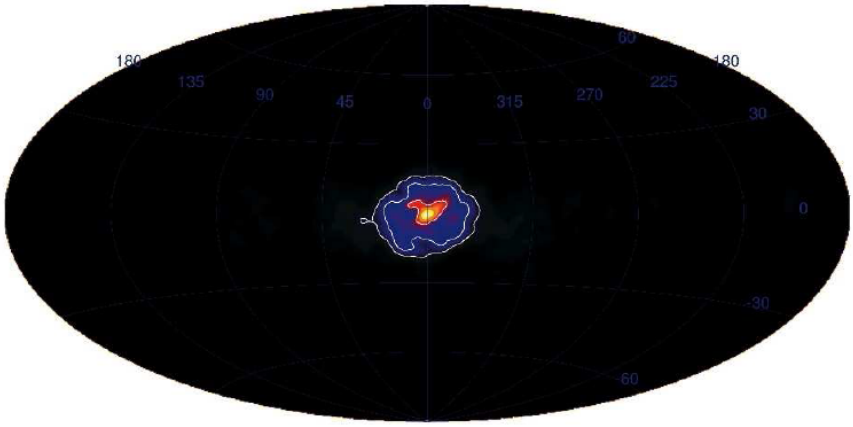

Figure 1.2: 511 keV line map from one year study of INTEGRAL/SPI data (Knödlseder et al., 2005).

provided a new opportunity to study the positron annihilation with the advanced high-resolution spectroscopy. The INTEGRAL orbit is eccentric, with an apogee of 153 000 km, a perigee of 9000 km, and a 72 hour period (one revolution). The INTEGRAL satellite, launched from Baikonur on October 17, 2002, is an ESA mission in cooperation with Russia and the United States. The 511 keV γ-ray line observed by the INTEGRAL/SPI (Spectrometer on INTEGRAL) is consistent with the production of 1.2×10^{43} ph/sec by positron in the center of the Galaxy within $\sim$ 1 kpc of the Galactic center. The emission is strongly concentrated towards the inner Galaxy and only a weaker emission is detected from the Galactic disk.

The first 511 keV line all-sky map from INTEGRAL data was presented by Knödlseder et al. (2005). He presented an all-sky map of the 511 keV γ-ray line emission (Fig. 1.2) with the goal of determining the morphology of the emission in the Galaxy and of searching for previously unknown sources of the emission. He found that this emission is significantly detected towards the Galactic bulge regions, confirming previous works and a possible detection of the Galactic disk. Previously, there was no evidence for positive latitude enhancement from CGRO/OSSE data (Milne et al., 2001). Annihilation rates are reported to amount $1.5 \pm 0.1 \times 10^{43}$ ph/sec and $0.3 \pm 0.2 \times 10^{43}$ ph/sec for the bulge and disk components, respectively. The bulge emission is spherically symmetric (Knödlseder et al., 2005) and centered toward the Galactic center with an extension of $\sim$ 8 degree (FWHM, Full Width at Half Maximum). The bulge morphology is dominated by the 511 keV line emission suggesting an old stellar population as the main Galactic positron sources. In contrast, the faint disk emission is well explained by the release of positrons during the radioactive decay of ^{26}Al originating from massive stars (Tomlinson et al., 2015). Although this result using the SPI data was consistent with previous analysis from CGRO/OSSE, its extreme bulge-to-disk ratio imposes severe constrains on the Galactic positron sources. Despite of possible explanations for this high γ-ray line contribution to the Galactic center, there remain a large uncertainty about the spatial distribution of the emission.

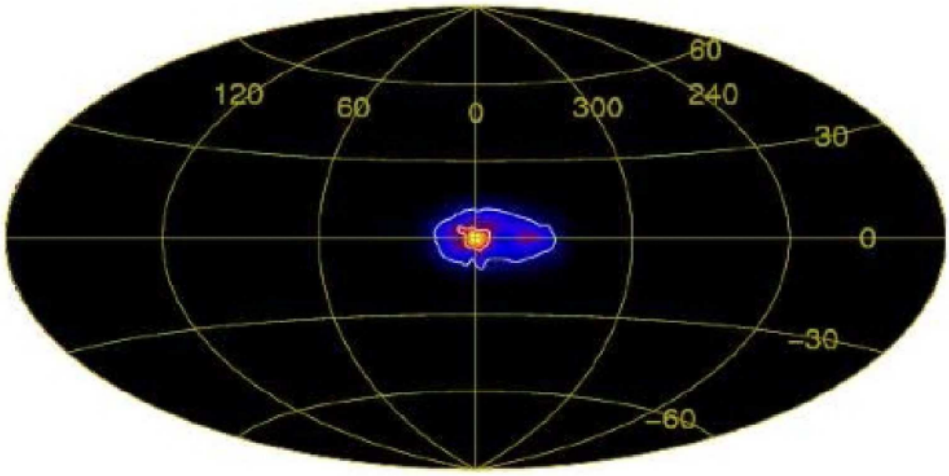

Figure 1.3: 511 keV line map derived from 5 years study of INTEGRAL/SPI data (Weidenspointner et al., 2008b).

Based on the 5 years of INTEGRAL/SPI observations, an all-sky map at 511 keV (Fig.1.3) revealed the extended faint emission along the Galactic plane (Weidenspointner et al., 2008b). However disk and spheroid components fitted to the data did not lead to a unique explanation. The map shows an asymmetric structure of the emission toward the positive Galactic longitude. However, a different analysis of the same SPI data found no evidence on this asymmetry (Bouchet et al., 2008). Confirming the symmetric or asymmetric nature of the Galactic positronium emission was a major goal of studies of the Galactic 511 keV line emission.

Table 1.1 summarizes main features of potential candidates for positron production (Prantzos et al., 2011). In the bottom line of this table through SPI data, the constraints from γ-ray emission at 511 keV are shown. From the SPI data the positron production rate is identified by amount of $\sim 2 \times 10^{43}$ ph/sec. From SPI measurement, the quantity of the bulge/disk can also calculated. This quantity shows the ratio of the Galactic bulge flux to the Galactic disk flux. According to the SPI data, this ratio exceeds 1.4, which represents a much stronger contribution of the source(s) in the Galactic bulge compared to those in the disk. Several potential candidates for positron production are classified in table 1.1. For each candidate four main characteristics are specified: the process in which positron are produced, the initial energy of the produced positron, the rate of positron production and the ratio of the bulge/disk, which was explained above. We can compare the characteristics of each source to that of the observational constrains from SPI. Although some of these potential candidates are more consistent with the observational constrains, none of the sources is in complete agreement with the observation. Among the compact objects that we know, type Ia supernovae (SNIa) is a good candidate for the positron production.

Source	Process	$E(e^+)$[a] (MeV)	e^+ rate[b] $\dot{N}_{e^+}(10^{43}\ s^{-1})$	Bulge/disk[c] B/D	Comments
Massive stars: ^{26}Al	β^+ decay	~1	0.4	<0.2	$\dot{N}$, B/D: Observationally inferred
Supernovae: ^{24}Ti	β^+ decay	~1	0.3	<0.2	$\dot{N}$: Robust estimate
SNIa: ^{56}Ni	β^+ decay	~1	*2*	<0.5	Assuming $f_{e^+,\mathrm{esc}} = 0.04$
Novae	β^+ decay	~1	*0.02*	<0.5	Insufficient e^+ production
Hypernovae/GRB: ^{56}Ni	β^+ decay	~1	?	<0.2	Improbable in inner MW
Cosmic rays	p-p	~30	0.1	<0.2	Too high e^+ energy
LMXRBs	γ-γ	~1	*2*	<0.5	Assuming $L_{e^+} \sim 0.01\ L_{\mathrm{obs},X}$
Microquasars (μQs)	γ-γ	~1	*1*	<0.5	e^+ load of jets uncertain
Pulsars	γ-γ/γ-γ_B	>30	*0.5*	<0.2	Too high e^+ energy
ms pulsars	γ-γ/γ-γ_B	>30	*0.15*	<0.5	Too high e^+ energy
Magnetars	γ-γ/γ-γ_B	>30	*0.16*	<0.2	Too high e^+ energy
Central black hole	p-p	High	?		Too high e^+ energy, unless $B > 0.4$ mG
	γ-γ	1	?		Requires e^+ diffusion to ~1 kpc
Dark matter	Annihilation	1 (?)	?		Requires light scalar particle, cuspy DM profile
	Deexcitation	1	?		Only cuspy DM profiles allowed
	Decay	1	?		Ruled out for all DM profiles
Observational constraints		<7	2	>1.4	

[a]Typical values are given.
[b]e^+ rates: in roman: observationally deduced or reasonable estimates; in italic: speculative (and rather close to upper limits).
[c]Sources are simply classified as belonging to either young ($B/D < 0.2$) or old (< 0.5) stellar populations.

Table 1.1: Properties of candidate positron sources in the Milky way (Prantzos et al., 2011)

One problem in this case is that the positron rate observed by SPI in the Galactic bulge is more than what we would expect from the supernovae distribution. Therefore SNIa cannot account for the entire 511 keV emission. A possible solution has been proposed by Higdon, Lingenfelter & Rothschild (2009), who suppose that a large amount of positrons are generated in the Galactic disk by supernovae explosions and are transported to the bulge following the Galactic magnetic field lines. However studies show that the propagation of positrons to the distance more than 100 kpc is impossible (Ferrière, Gillard & Jean, 2007). Another candidate for positron sources are the low mass X-ray binaries (LMXBs), proposed by Weidenspointner et al. (2008c) as a class of positron sources that could explain the 511 keV emission line measured by INTEGRAL. This idea is supported by two arguments: first, the LMXBs are old systems concentrated in the Galactic bulge, where we detect the main amount of annihilation radiation; second the SPI results show an asymmetric 511 keV emission morphology with a spatial shape that is consistent with the LMXBs distribution above 20 keV. In the X-ray binaries, electron-positron pair are might created in the vicinity of the compact objects, either in the hot inner accretion disk, or in the X-ray corona surrounding the disk, or at the base of jets. The jet might channel a fraction of positron-electron pairs out of the system (Prantzos et al., 2011). However, the amount of produced positron flux in the Galactic bulge is not enough to reach the amount seen by SPI.

Open questions remain: What is the true spatial distribution of the Galactic γ-ray emission at 511 keV? Is the disk emission indeed asymmetric or not? How large are the contributions by the Galactic bulge and disk? Answers to these problems can improve our knowledge about the astrophysical positron source(s), the escaping fraction of positrons in supernovae type Ia, and maybe also the large-scale configuration of the Galactic magnetic field which might affects the way of positron propagation in the Galaxy. They also provide information on the properties of the ISM turbulence as it influences the propagation of positrons after production. A deeper γ-ray observations of the Galaxy and a better under-

standing of the observational background are required to tackle those questions. In order to shed light on these astrophysical questions, and to image the positronium emission we propose to use probability theory, applied to the unknown emission field.

1.2 Probability theory

Probability theory is used to tackle the ill-posed inverse problem of imaging based on the observed data. Information theory provides a robust mathematical framework to address the question what is the inferred signal for a given set of data (Shannon, C. T, 1949). We apply the information theory on fields in the framework of probability theory, allowing a robust flux reconstruction of the detected celestial sources.

The data are usually corrupted by noise and also limited by instrument specifications. Noise is an unwanted quantity originating from random processes and signal is defined to be any unknown quantity of interest. The detected astrophysical events are corrupted by effects due to instrumental complexity. The function which identifies how the events are recorded in the detectors with respect to the hardware characteristics of the instruments is called response function. The details of the SPI response will be discussed in chapter 3. Extracting information from the data is a challenging task. From the data we want to infer the properties of a signal. The application of the response function to the astrophysical signal is called the signal response (Rs). There are three problems that relate to the signal inference:

1) The noise deviate the data from the ideal signal response of an instrument. In this project, the observed data recorded by INTEGRAL/SPI are noisy because they are corrupted by background events, statistical fluctuations like Poissonian shot noise, and calibration effects. The background is mainly due to cosmic rays, i.e., high energy particles traveling through the Milky Way .

2) Data incompleteness: SPI observes the sky in the γ-ray domain and provides us with counts as a function of instrument pointing, detector and photon energy. In this project our signal is the diffuse celestial γ-ray emissivity of electron-positron annihilation. This emissivity is a function of celestial position and therefore a field over the sphere. The signal space is continuous with degrees of freedom approximate to infinity. However in practice, the observed data are limited: SPI data do not cover all energy ranges, times, and positions in the sky. Thus, not all aspects of the signal are recorded by the instrument, producing missing information.

3) Ill-posedness of the inverse problem: The solution of the signal inference for a given data set is not unique and many different signal configurations would provide the same data. This problem is inverse because one starts from the observed data to deduce the signal. The problem is ill-posed because not a unique solution is assured.

For these reasons we use probability theory for the data analysis. Probability theory permits us to calculate the expectation value and the uncertainty of a signal field given the data and other information. If A is an astrophysical quantity (e.g., signal), then P(A) is the probability of the mentioned quantity that occurs or not. The outcome is a real

number between zero and one, 0< P(A) < 1, and the more plausible is A, the more P(A) is closer to 1.

In 1946 Richard Cox derived the rules of probability theory (Jaynes, E. T. and Bretthorst, G. L, 2003). The first rule is called the sum rule and it relates the probability $P(A|I)$ of an event to happen to that of the event not to happen $P(\overline{A}|I)$

$$P(A \mid I) + P(\overline{A} \mid I) = 1. \tag{1.1}$$

$\overline{A}$ denotes the proposition that A is false (or does not occur) and "$\mid I$" means given that I assumed to be true. The sum rule states that the probability that A is true plus the probability that A is false is equal to one. From Eq. 1.1, all given probabilities will be conditioned over background information I, i.e., it consists of all information we had before the event A could happen.

The second rule is the product rule. The joint probability of A and B events given I is stated by the product rule to be:

$$P(A, B|I) = P(A|B, I) \times P(B|I) = P(B|A, I) \times P(A|I). \tag{1.2}$$

Here the joint probability $P(A, B|I)$ is the probability that both A and B happen.

Commonly, two different approaches are employed for signal inference: Bayesian and frequentist methods. Both approaches are based on the principles of probability theory but with different philosophical promises. For the analysis of the INTEGRAL/SPI data both approaches have been used (Bouchet, Jourdain & Roques, 2015; Siegert et al., 2016). In subsections 1.2.1 and 1.2.2, a brief summary of these two approaches will be discussed.

1.2.1 Bayes' theorem

Bayesian theory is based on probability theory to quantify the belief in statements. Therefore, Bayes' theorem is able to deal explicitly with the uncertainty associated to any statement. This uncertainty originates from the incomplete nature of the observed data and it is based on the product rule (Eq. 1.2). For signal inference, the Bayes' theorem can be expressed as follows

$$P(s|d) = \frac{P(d, s)}{P(d)} = \frac{P(d|s)\, P(s)}{P(d)}. \tag{1.3}$$

Here s is the signal, any quantity that we are interested to study, and d is the measured data. The normalization factor in the denominator of Eq. 1.3 is called evidence. The evidence is defined as

$$P(d) = \int P(d, s)\, \mathcal{D}s, \tag{1.4}$$

where $\mathcal{D}s$ integrates over all possible configurations for the signal. $P(d, s)$ is the joint probability of data and signal, that can be decomposed by $P(d|s)$ and $P(s)$. $P(s)$ is the prior representing our knowledge about the signal before any data were taken. For

example, assume the signal is the diffuse emission from a celestial source and we are interested in studying how the emission is as a function of position. We know that diffuse emissions change smoothly across different locations in the sky. Therefore, we use a suitable probability density function (pdf) expressing this mathematically as our prior. Therefore, we encode our knowledge on the signal with a suitable prior pdf which expresses the physical properties of the signal. An appropriate prior pdf for the diffuse emission is the Gaussian distribution pdf, which allows one to correlate the signal emissivity with nearby regions.

The likelihood, $P(d|s)$, is the probability of the observed data given the signal. Using the likelihood, several signal configurations, that produce the observed data, are tested.

After identifying the prior and the likelihood pdfs, the posterior $P(s|d)$ is derived (Eq. 1.3). Bayes' theorem relates the likelihood and the prior to the posterior, $P(s|d)$. The posterior pdf states how probable the signal is given a specific data set.

Often a point estimate for the signal is desired. One possibility is to use the maximum a posterior (MAP) estimate which provides the most probable signal configuration given the data and other information.

Often the signal depends on several parameters. For instance if the signal is weather conditions, then the signal would consist of pressure, temperature, humidity and etc. Each of these parameters reflects different aspects of the signal. By use of the Bayesian theory, we can focus on one parameter and marginalize the others. In the mentioned example suppose that there are two parameters ω and α in the signal, and we just want to concentrate on the information of ω in the data, irrespectively of α. Then the probability of the parameter ω given the data is (Gregory, 2005)

$$P(\omega|d) = \int P(\omega, \alpha|d)\, d\alpha, \tag{1.5}$$

where $P(\omega|d)$ is the marginal posterior distribution for ω. We can also expand the Eq. 1.4 by using Bayes' theorem

$$P(\omega, \alpha|d) = \frac{P(d|\omega, \alpha)\, P(\omega, \alpha)}{P(d)} = \frac{P(d|\omega, \alpha)\, P(\omega)\, P(\alpha|\omega)}{P(d)}. \tag{1.6}$$

If we can assume that the priors for both parameters are independent, $P(\alpha|\omega) = P(\alpha)$, we have

$$P(\omega, \alpha|d) = \frac{P(\omega)\, P(\alpha)\, P(d|\omega, \alpha)}{P(d)} \tag{1.7}$$

and

$$P(\omega|d) \propto P(\omega) \int P(\alpha)\, P(d|\omega, \alpha)\, d\alpha. \tag{1.8}$$

In summary, the Bayes' theorem provides two main advantages:

1) It is possible to incorporate prior information into the inference. Mathematically the prior information describes our previous knowledge modeled with a statistical distribution.

2) The marginalization in Bayes' theorem reduce the complexity of the signal inference, since a lower number of parameters are included in the models. However in practice, marginalization can lead to awkward multidimensional integrals over the parameters space, which can be tackled by numerical methods.

1.2.2 Frequentist inference

Frequentist theory constructs parametrized models for the sky. Various data samples are produced by application of the response function on these models to test the hypotheses. The procedure of the hypotheses testing is repeated over and over based on the long-run frequency behaviour. Notice that all these data samples are produced in the same circumstances such as the application of the same response and the same noise. This is similar to calculate the likelihood, $P(d|M)$, in Bayesian theory if M is an known model.

Since the nature of the astrophysical events are always uncertain, random variables are used as argument to calculate the probability of the events (Jaynes, E. T. and Bretthorst, G. L, 2003). The frequentist approach provides only the probability of statements about those arguments, that are random variables. These arguments take various values. Suppose to perform an astronomical observation with several measurements. Outcomes of this observational experiment in different measurements are not the same. During the transformation of these collected photons into the recorded instrument, there is the possibility that some information is missing. In fact, collected photons are influenced by many factors such as cosmic rays, secondary particles produced by instrumental interactions and etc. The recorded astrophysically interesting photon counts are integer numbers. In other words, the real counts are the signal that happen in the sky and random variables are data that are obtained from the instrument. We can only associate random variables to such quantities that are not fixed. The reason why we call these parameters random stems from the intrinsic uncertainty of the nature. The word random is used for those experiments that the details of the operation are not either completely understood or too complicated to handle in detail. For this reason it is not possible to consider directly the prior, $P(s)$, in the framework of the frequentist approach in the procedure of the signal inference.

The application of just the random variables also prevents us to implement the posterior in the frequentist theory. If we want to define the posterior pdf in the framework of the frequentist theory, the condition d in $P(M|d)$ has to be identified at a fixed value. While in this approach different datasets come out from different hypothesis tests and the goal is the comparison of these datasets. So the condition in the posterior in this framework is not fixed at a particular point, and the posterior can not be derived readily. Therefore specifying the posterior is not possible in the frequentist theory. However, the likelihood $P(d|M)$ is extensively used and calculated in this method. Because the likelihood provides the probability of the data if a model M is given, and the model can be specified at a particular fixed point. In this approach information provided by observed data are the only quantities that can be related to the probability theory.

As it was mentioned before, frequentist approach is based on the hypothesis testing, and in this procedure just rejection or confirmation of a assumed model comes out. In this

test the likelihood, $P(d|M)$, would be derived for a particular model, M. This likelihood can be used to answer the question whether the assumed model is consistent with optimal value of the observed data or not. Consider we want to infer an astronomical event and the optimal value for data is equal to d_α. In the frequentist framework there are only two possibilities for $P(d|M)$: 0 or 1, depending on whether $d = d_\alpha$ or not. It is true that many other hypotheses can be tested again, but the process of the calculation should be repeated agian. While in the Bayesian theory many different models are involved in inference calculations simultaneously, and probabilities of models are described by only real numbers between 0 and 1.

1.3 Concepts of Information Field Theory

1.3.1 Information Theory

Information theory studies the way how to use, extract and transform information coming through a noisy data channel. The theory was developed by Shannon (1948), and was applied to many data analysis and signal inference problems subsequently. A crucial quantity in this theory is the entropy, given by

$$\mathcal{H} = -\sum_i p_i \log(p_i). \tag{1.9}$$

Here i labels the possible events and the p_is are the probabilities of the events, i. $\mathcal{H}$ is so-called the Shannon entropy. The entropy quantifies the amount of uncertainty expressed by p_i. Information theory stands on probability theory.

1.3.2 Information Field Theory

The application of information theory to fields is called information field theory (IFT) (Enßlin, Frommert & Kitaura, 2009). In this framework, a field is any physical quantity which lives on some continuous spaces. To infer the properties of a signal field, IFT provides mathematical tools to extract the information on the signal field from the observed data and to develop working algorithms for the signal inference, based on Bayes's theorem. The results of the signal inference can be an image of the sky, and for this reason the procedure of signal inference is also called image reconstruction.

IFT is capable of image reconstruction in the case of linear and non-linear relations between signal and data. There are ambiguities for the reconstruction problems. A dataset observed by an instrument could be described from many different signal configurations, which one should be chosen as the best reconstruction.

One of the strength of IFT is that it can explain the smoothness of a field to the benefit of a better image reconstruction. For practical reasons, observed data are always limited in size, while the number of degrees of freedom for signal fields is infinite. Therefore, it is impossible to solve uniquely for the signal from the data since the number of unknown

quantities exceeds largely the number of constrains. IFT uses the smoothness of a field to decrease effectively the freedom of the unknown parameters. The mathematical details of this procedure are discussed in Sec. 2.3.

1.3.3 Signal and data spaces

In this project, the signal field is the photon flux of the γ-ray at 511 keV originating from electron-positron annihilation as a function of sky position. Both the signal and the observed data can be defined in a continuous or discrete space. However, since we are dealing with digital analysis of the data, the discrete space is preferred for the data. INTEGRAL/SPI data depend on three coordinates: pointing, energy and detector.

The data can be regarded as a finite dimensional vector. The signal field is a vector as well, but from an infinite dimensional Hilbert space. On a computer, we represent it as a finite dimensional vector of pixel values. In IFT, we just require that for smooth fields the resolution of the pixels can virtually reach the continuum.

For spatially smooth signals, we expect correlations between various pixels of the space over which the signal is defined on. This requires to define a continuous space for the signal field. We are dealing with a high-dimensional discrete space that can be interpreted as a continuous space.

We extend the usual function scalar product for fields in a signal space to their discrete representations. For clarity, let us suppose a discrete set of pixel positions $\{x_i\}$, being the subset of some continuous field space $\Omega : \{x_i\} \subset \Omega$. The scalar product of the two discrete function-vectors $f = (f_i)$ and $g = (g_i)$ is defined by (Enßlin, Frommert & Kitaura, 2009)

$$g^\dagger f \to g^\dagger f \equiv \sum_{i=1}^{N_{pix}} V_i {g_i}^* f_i, \tag{1.10}$$

where the star indicates the complex conjugate and the dagger the adjoined of a vector or matrix, and V_i is a volume-size attributed to the set of x_i . This scalar product has the desired continuous limit

$$g^\dagger f \to \int \mathrm{d}x\, \mathrm{g}^*(x)\, \mathrm{f}(x). \tag{1.11}$$

1.3.4 Basic formalism of IFT

The general data model for the linear measurements is

$$d = R(s) + n, \tag{1.12}$$

where d, s and n denote the data, signal and noise, respectively. R is the response operator which applied on s and transforms the sky emissivity into the data. For simplicity we assume s to live here only over one dimension. The data model is given by

$$d_i = \int R_i(x)\, s(x)\, dx + n, \tag{1.13}$$

in which i labels data vector entries which collected signal from different locations x in the sky as specified by the response $R_i(x)$. The data are not necessarily influenced only from a single location in the sky but various sky locations might have contributed to one single data point. For this reason, the response connects a particular data point i to different locations in the sky, x. While the response connects data and signal, we assume here the noise to be independent of the signal.

As it was explained earlier in Sec. 1.2, for the goal of signal inference we can adopt the MAP method which maximizes the posterior $P(s|d)$ with respect to s. Here, we want to see how the MAP method works in IFT. The posterior is defined as follows

$$P(s|d) = \frac{P(d,s)}{P(d)} \equiv \frac{1}{Z} e^{-H(d,s)}, \tag{1.14}$$

where $Z \equiv Z_d$ is the partition function and $H(d,s)$ is the information Hamiltonian

$$\begin{aligned} H(d,s) &= -\ln(P(d,s)), \\ P(d,s) &= P(d|s)\, P(s). \end{aligned} \tag{1.15}$$

The evidence of the data is

$$P(d) \equiv \int \mathcal{D}s\, P(s|d)\, P(s) = \int \mathcal{D}s\, e^{-H(d,s)} \equiv Z, \tag{1.16}$$

where the integration occurs over all possible signal configurations. Although different configurations for the signal are possible, we can calculate the posterior mean of the signal, i.e., the most probable configuration for the signal. By use of MAP method, when the optimal value for the posterior is obtained, the expectation value of the signal is derived as follows

$$m = \langle s \rangle_{P(s|d)} = \int s\, P(s|d)\, \mathcal{D}s. \tag{1.17}$$

This mean that m provides an optimal field over a continuous space and is called the reconstructed map.

The uncertainty map for the posterior mean is defined as

$$D \equiv \left(\langle (s-m)(s-m)^\dagger \rangle \right)_{P(s|d)}. \tag{1.18}$$

This quantity is also known as Wiener variance (Enßlin, 2013).

Depending on the different physical and instrumental circumstances, different priors and likelihoods have to be adopted, respectively. Consequently, the resulting posterior $P(s|d)$ will differ as well as any signal map derived from it. This is equivalent to say that corresponding to different physical situations, different mathematical operators or filters

are applied to the observed data. The details of two specific filters are discussed in this thesis: Wiener Filter (Enßlin, 2013) and D^{3}PO (Selig & Enßlin, 2015). These two filters will be discussed in further detail in Chapter 2, therefore we will have the opportunity to compare these IFT filters to the previously used filters in the framework of the frequentist theory.

By use of this modern information processing theory, we hope to find the unsolved emission for 511 keV line from the measured INTEGRAL/SPI data. This could help to find out about the unusual behavior of currently known sources of Galactic positrons which are located in the Galactic disk plane, and the annihilation radiation generated mainly in a central bulge regions.

1.4 Outline

In the continuing chapters of this thesis, we detail the inference problems individually, depending on the physical conditions we are considering. In Chapter 2, two common methods in the signal inference and image reconstruction will be discussed in details: maximum likelihood and maximum entropy methods. Outstanding results on the sky imaging of the Galactic positron annihilation for individual methods will be illustrated. Details of the filters in framework of IFT will be investigated.

In chapter 3, INTEGRAL and SPI instrument are described, including performance and properties. The instrumental response is discussed in details. The collected and employed data are explained.

Results from several simulations are presented in Chapter 4, needed for the verification of the response function applied in the framework of IFT algorithms.

Results obtained within IFT algorithms on the INTEGRAL/SPI data are presented in Chapter 5. First the result of the image reconstruction by use of Siegert et al. (2016)'s background model in one energy bin of 511 keV is shown. Later the Galactic γ-ray emission in a wider range of energies from 500-520 keV is investigated, which is implemented by use of a spectral response function. In the end of Chapter 5, the challenging reconstruction of the background changes with respect to the exposure time are studied to image the sky of the Galactic positron annihilation at 511 keV, independent of the time variations of the Siegert et al. (2016) 's background model.

The conclusion and remarks are presented in Chapter 6. We conclude and interpret the physical concepts from the maps. We investigate the contribution of the Galactic bulge and disk in individual maps. The total flux of the all-sky maps are calculated and they are compared to that of previous obtained images, which used the same data, but with different data analysis methods. In the end, we discuss the critical role of the used background to reconstruct the sky image of the γ-ray emission at 511 keV.

Chapter 2

Theory of image reconstruction

2.1 Introduction

Statistics studies ensembles of data. As it was previously discussed in Sec. 1.2, the observed data are part of a large population, and they do not cover all aspects of a celestial objects. Therefore, the information provided by the data on celestial signals is usually incomplete. In statistical inference, a sample (the observed data) from a larger set (compared to the sample) is studied to infer the properties of the celestial objects.

2.2 Different methods for the signal inference

The imprint of the sky signal on the data generated by a detector is usually described by response function R. If we can define the inverse response function, R^{-1}, then it is possible to apply this inverse function on the observed data and recover the sky. But in practice observed data provids only incomplete information and therefore the R^{-1} is not uniquely defined (Skilling, J. and Bryan, R. K., 1984). For this reason various inverse filters are used to overcome this difficulty by modifying R^{-1}. The different ways how scientists modify the R^{-1} create different signal inference methods.

In the frequentist point of view, as it was discussed earlier in chapter 1, many data samples are produced by considering various models. These samples are compared to the real data. The way this comparison is performed produces different frequentist signal inference methods.

In the following two subsections, the main features of two common approaches will be discussed: the maximum likelihood estimation and the maximum entropy method. The former approach is based on the frequentist statistics while the second one on Bayesian probability theory.

2.2.1 Maximum likelihood estimate

For all-sky signals, observational data are usually obtained through several observations targeting different sky directions so that the whole sky is covered eventually. Suppose our plan is to study the morphology and emissivity of an emission in a particular energy. Since we do not know the real signal, various models (M) could be considered for it. The likelihood function of a hypothesis M, $P(d|M)$ is given by

$$P(d|M) = \prod_{i=1}^{N} P(d_i|M), \tag{2.1}$$

where i indicates N independent observations. The model M here can include various parameters of the sky emissivity. This function is not a probability density for the model parameters, but it is the probability function for the data. The likelihood provides information on the validity of a supposed model. If the assumed model is consistent with the observational data, the model will be regarded to be valid. By implementation of the response function on the various considered models, different signal-responses are produced. The noise is also modeled, depending on the observational data. When the modeled noise is added to the different signal-responses, several simulated data are created. To estimate the validity of an assumed model we will need a tool to compare these simulated data to the observational ones (real data). It is more convenient and computationally more stable if to consider the logarithm of the likelihood. The so-called log-likelihood (Bohm, G. and Zech, G., 2010) is given by

$$\ln(P(d|M)) = \sum_{i=1}^{N} \ln(P(d_i|M)). \tag{2.2}$$

The likelihood function contains all information the observation provides. The maximum likelihood estimate (MLE) chooses among all available models or hypotheses one whose likelihood is maximum. To find the maximum of the likelihood function, first we derive the log-likelihood with respect to the model parameters and set this to zero

$$\frac{d\ln(P(d|M))}{dM}|_M = 0. \tag{2.3}$$

An example for the usage of MLE from the image reconstruction of the Galactic positron annihilation is the work of Knödlseder et al. (2005). In this work an all-sky map of the 511 keV γ-ray line emission is presented based on one year INTEGRAL data starting from Dec 10, 2004. To determine a model for the 511 keV γ-ray line the Richarson-Lucy algorithm was employed (Richardson (1972) and Lucy (1974)). The resulting all-sky image of the 511 keV line emission is shown in Fg. 2.1.

Another example of a MLE application to map the Galactic 511 keV emission is the work of Bouchet, Roques & Jourdain (2010), who estimated the γ-ray emission shown in Fig. 2.2. The 6 years of data which were used in this investigation are in the $508.25-513.75$ keV energy range, so that the 511 keV γ-ray line is at the center of the energy range. Based

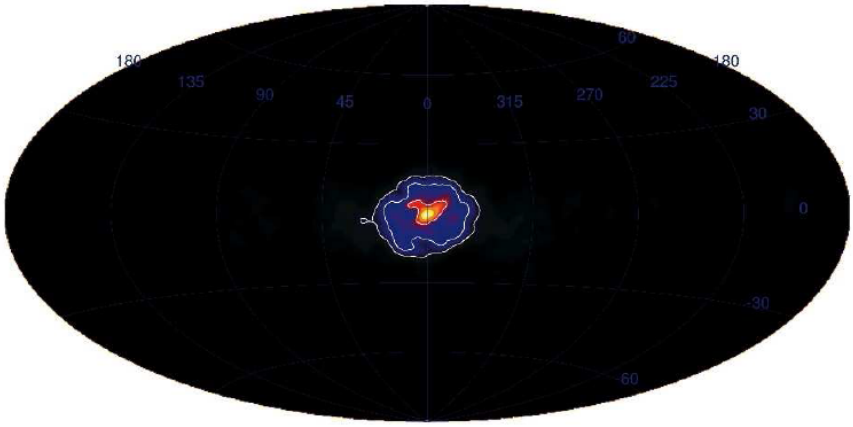

Figure 2.1: 511 keV line map from one year study of INTEGRAL/SPI data (Bouchet, Roques & Jourdain, 2010).

on the observed data the longitude and latitude extension of the emission for the utilized models are quantified. In this analysis method, point sources are modeled iteratively from starting with the strongest ones. Then successively weaker sources are added to the model (However in the end, no point sources of the annihilation radiation was detected). At each iteration of the algorithm supposed sources are fitted to the recorded data, then those sources are used as baseline knowledge for the next search (Bouchet et al., 2008).

In principle after maximization of the likelihood, it is also possible to derive the uncertainty and the residual of the derived map for individual points (the residual is the difference of the simulated data to the real data). The obtained uncertainties are called confidence intervals, often be assummed to be symmetric.

As it was explained, prior, $P(s)$, does not have a direct role in the MLE parameter method. Non-parametric MLE imaging is known to be unstable. The Richardson-Lucy (RL) algorithm is often regularized by stopping the iterations before convergence. This can be regarded as a hidden prior, as the use of RL method decides, based on his/her expectation on the image, when is the best moment to truncate the algorithm. However this issue will be different for the next approach that is going to be discussed.

2.2.2 Maximum Entropy method

The Maximum Entropy method (MEM) introduced by the work of Jaynes (1957). Based on the maximum entropy (MaxEnt) principle, to determine the probability distribution for the signal we should select that distribution which has the largest entropy, consistent with our constraints. If s_i indicates the intensity of the individual pixel in the sky image, the entropy of the image is defined as $S = -\sum_i s_i \ln s_i$. In the γ-rays regime, as we count the photons, the Poisson distribution is for the likelihood. Contrary to the MLE, the Bayesian framework which MEM is built on it, requires the explicit description of the prior probability. The entropy for the signal and the log likelihood are combined in the Q

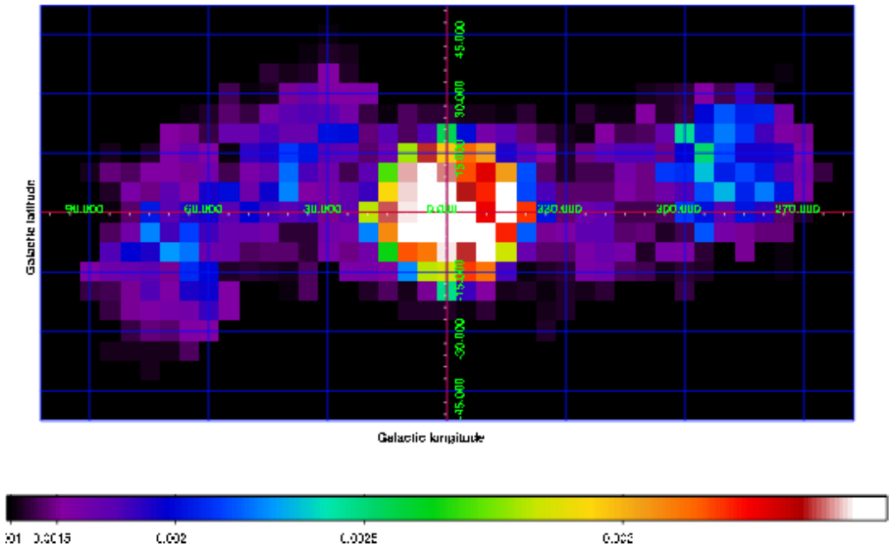

Figure 2.2: 508.25-513.75 keV INTEGRAL/SPI smoothed intensity map in photon cm^{-2} sec^{-1}. Pixel size is $5° \times 5°$ (Bouchet, Roques & Jourdain, 2010).

function

$$Q = S - \lambda L, \tag{2.4}$$

where λ is an undetermined coefficient specifying the relative weight of the log likelihood to that of the entropy, the log-prior of MEM. In this method the entropy is maximized over all possible signal configurations which are consistent with the data. To maximize the constrained entropy S, the derivative of Q has to be taken,

$$dQ = dS - \lambda dL = 0. \tag{2.5}$$

If we compare the Eq. 2.4 to the Bayesian formula, the first term represents the log-prior and second one is the negative of the log-likelihood, if Q is considered as the posterior. Thus

$$Q(s) = -\lambda L - \alpha \int dx\, s(x)\, \ln \frac{s(x)}{s_0}, \tag{2.6}$$

where α is still an undetermined coefficient which has to be identified and s_0 is a normalization factor. We have

$$\ln P(s) = -\alpha \int dx\, s(x)\, \ln \frac{s(x)}{s_0} \tag{2.7}$$

so that,

$$\begin{aligned} P(s) &= e^{-\alpha \int dx\, s(x)\, \ln \frac{s(x)}{s_0}} \\ &= \prod_x e^{-\alpha\, V\, s_x \ln(\frac{s_x}{s_0})}, \end{aligned} \tag{2.8}$$

in which V is pixel volume factor. In the end the MEM prior is

$$P(s) = \prod_x \left(\frac{s_x}{s_0}\right)^{-\alpha\, V\, s_x}. \tag{2.9}$$

This probability distribution decreases exponentially with increasing s. If s represents the intensity of pixels in the sky image, this distribution basically forbids bright sources. This renders the MEM prior to be inadequate for the reconstruction of the point-like sources.

An example for MEM image reconstruction of the 511 keV γ-ray line was produced by Cheng et al. (1997). In their MEM application, a large data set based on CGRO/OSSE observations was processed by the MEMSY5 software. This algorithm is based on the iterations, starting from a map which has the highest probability for the prior, which is usually assumed to be a flat surface. After each iteration more constraints were considered for the prior, regarding if that prior fits better or not to the data. This leads to more structures in the map. As well after each iteration, the prior probability went down a little while the likelihood increased. Thus, the posterior becomes more influenced by the likelihood rather than the prior. The algorithm was continued until the posterior reaches the maximum state. The map of their reconstruction of the 511 keV line emission is shown in Fig. 2.3 and suggests two components: a central bulge and a Galactic plane component. The Galactic plane is extended toward the negative Galactic longitudes.

In general, the MEM has three main advantages for the sky imaging:

1) The first advantage of MEM is the presence of a prior explicitly into the inference formula, while the prior is not considered directly in the frequentist approach. There are infinite number of signal configurations which are compatible with the observed data. To constrain these configurations we incorporate our prior knowledge about the signal into the process of the signal inference. However this prior is less restricted to limit to a few models compared to the frequentist approach. Therefore in MEM more signal configurations con contribute to the prior.

2) The second benefit of the MEM approach for the signal inference is related to the positivity of the γ-ray flux. In the γ-ray domain, where the wavelength of the photons are very short, we are dealing with the counts of the photon, which hit the detectors. Suppose the sky image is pixelized and each pixel contains a number of photons. Eq. 2.9 shows the image entropy can be expressed in terms of a probability , which states a probable intensity of individual pixels has a definite value. Regarding the sky intensity as a set of positive numbers is discrete space, this probability is equal to

$$P_i = \frac{s_i}{\sum s}. \tag{2.10}$$

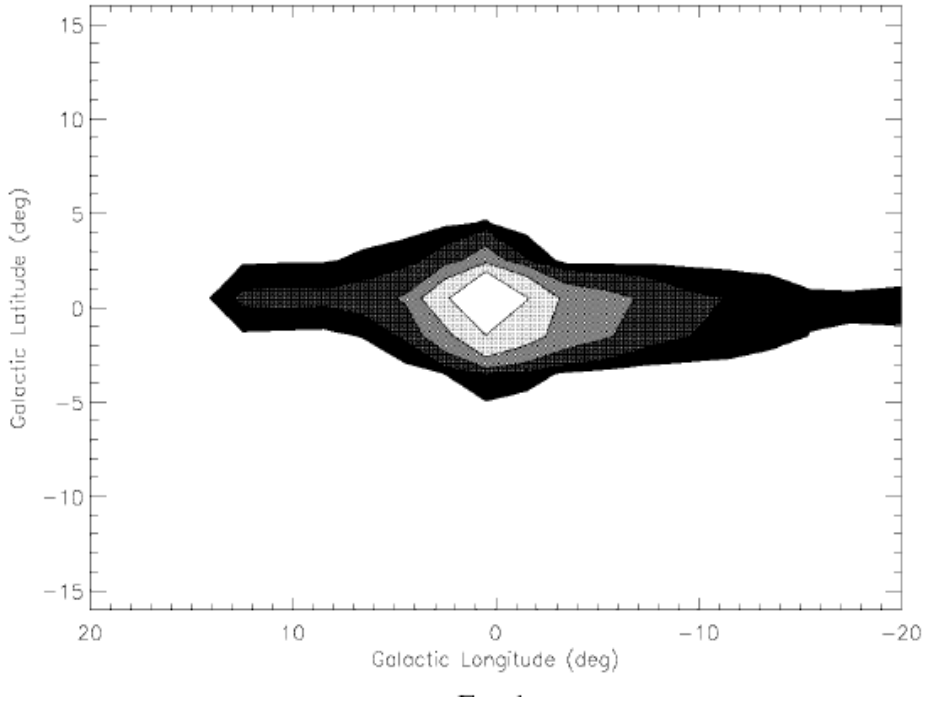

Figure 2.3: The reconstructed maps of GC 511 keV line radiation distribution from a simulated data set generated from a model composed of a central bulge component and a disk component (Cheng et al., 1997).

Since the image entropy does not exist for negative values of s, this requires the flux to be positive everywhere. Positivity ($s_i \geq 0$) is a physical important requirement, is enforced automatically by the MEM (Skilling, J. and Bryan, R. K., 1984).

3) The third advantage of the MEM is related to the correlation exists between the sky pixels. The reason why MEM images look usually smooth is that the supere-exponential form of the prior punishes bright pixels so that any observed flux likes to be spread across the pixels of the sky image. To investigate in details, we turn to refer to the MEM which used for the sky mapping by Bouchet, Jourdain & Roques (2015). The energy range of this investigation is not at 511 keV, but in range of 1805-1813 KeV which is related to the Galactic ^{26}Al emission. The entropy function of the signal used by Bouchet, Jourdain & Roques (2015) is

$$Q(s) = \alpha \sum_{i=1}^{N} \left(s_i - m_i - s_i \ln \left(\frac{s_i}{m_i} \right) \right) - \frac{1}{2} \left(\sum_{i=1}^{M} \frac{\left(\sum_{j=1}^{N} R_{ij} s_j - b_i - d_i \right)^2}{{\sigma_i}^2} \right), \tag{2.11}$$

where m_i is a sky model of an individual pixels, which the intensity reconstruction s should adopt in case of lack of data. b refers to the background and α is an undetermined parameter that controls the weight of the prior term with respect to the likelihood. N and M count the pixels in the signal and the data space, respectively. Bouchet, Jourdain & Roques (2015) maximizes Q to obtain a model for the sky brightness in ^{26}Al. Here in Eq. 2.11 the first term is the log prior, $\ln P(s)$, and second one is the log likelihood, $\ln P(d|s)$. If one wants to maximize the entropy with the data constraints, the sum of both terms has to be maximized. Suppose that we only consider the maximum of the log prior

$$\ln P(s) = \sum_{i} \left(s_i - m_i - s_i \ln \frac{s_i}{m_i} \right). \tag{2.12}$$

If one takes the exponential for both side one obtains

$$P(s) = \prod_{i} (e^{s_i} e^{-m_i} \sum_{i} \quad (\frac{s_i}{m_i})^{-s_i}). \tag{2.13}$$

For the prior image m Bouchet, Jourdain & Roques (2015) decided to insert

$$m_j = \text{const}(j) = \exp \left(\frac{\sum_{j=1}^{N} s_j \ln s_j}{\sum_{j=1}^{N} s_j} \right). \tag{2.14}$$

If Eq. 2.14 is substituted in Eq. 2.13, the prior would be

$$P(s) = \prod_{i} \left\{ e^{s_i} \exp \left[- \exp \left(\frac{\sum_{j=1}^{N} s_j \ln s_j}{\sum_{j=1}^{N} s_j} \right) \right] \sum_{i} {s_i}^{-si} \sum_{i} \left[\exp \left(\frac{\sum_{j=1}^{N} s_j \ln s_j}{\sum_{j=1}^{N} s_j} \right) \right]^{s_i} \right\}. \tag{2.15}$$

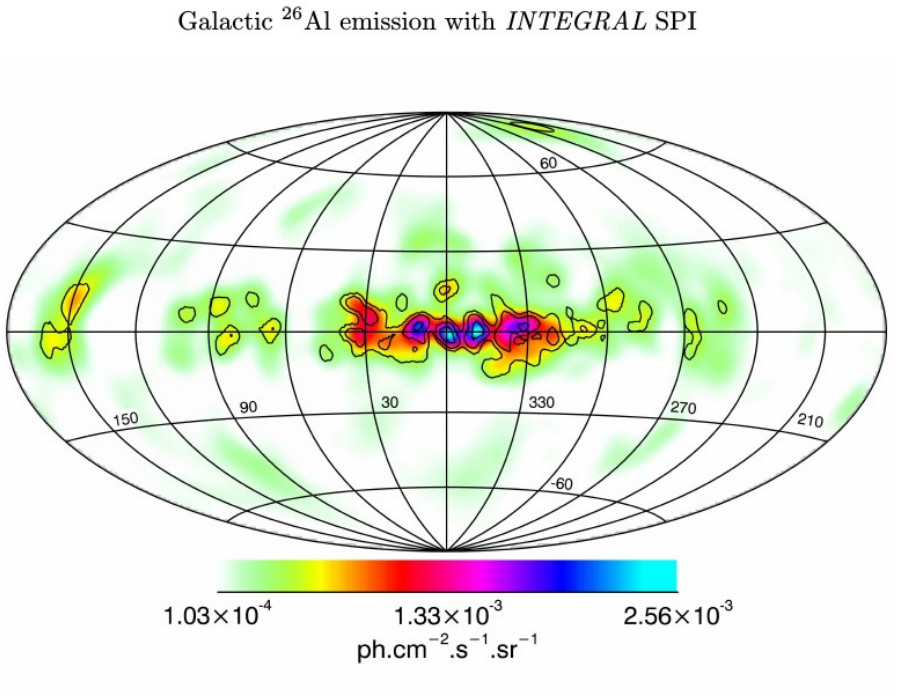

Figure 2.4: Image of the ^{26}Al line in range of 1805-1813 keV (Bouchet, Jourdain & Roques, 2015).

So the probability of the signal for the particular model Eq. 2.15, couples the intensity of all different pixels i and j, which expresses correlation between the pixels of the sky image. This is a positive point for MEM, since most of the astrophysical events are smooth in the sky. The image derived from this smooth prior, is illustrated in Fig. 2.4. However this correlation between pixels does not depend on the distances of the pixels. It means this correlation does not distinguish the location of the individual pixel.

2.3 Filters in Information Field Theory

2.3.1 Wiener Filter

In principle, the Wiener Filter (WF) is the simplest filter in IFT, and it is used here to introduce the theory. For the WF, the prior for the signal is modeled as a zero-mean Gaussian distribution,

$$P(s) = \mathcal{G}(s, S) = \frac{1}{\sqrt{|2\pi S|}} \exp(-\frac{1}{2} s^\dagger S^{-1} s). \tag{2.16}$$

S describes the typical fluctuations of the signal in terms of correlations between different locations (see 2.4.6)

$$S(x, y) = \langle s(x)s(y)\rangle_{(s)} = \langle ss^\dagger\rangle_{(s)} = \int \mathcal{D}s\, s(x)s(y)P(s). \tag{2.17}$$

Here, a correlated Gaussian distribution is chosen for the prior because many of the astrophysical signals are continuous, so that the correlations between different locations in the sky image are expected. The Gaussian distribution models correlates very well.

At this moment, we assume the noise to be signal-independent and Gaussian,

$$P(n) = \mathcal{G}(n, N) = \frac{1}{\sqrt{|2\pi N|}} \exp(-\frac{1}{2} n^\dagger N^{-1} n), \tag{2.18}$$

where $N = \langle nn^\dagger\rangle$ is the noise covariance matrix assumed here to be known. According to Eq. 1.12 and 2.18 and also considering that the noise is signal-independent, the likelihood is

$$P(d|s) = \mathcal{G}(d - Rs, N). \tag{2.19}$$

Since our goal is the signal inference, all s-independent terms can be ignored. From Eq. 1.17, the posterior is

$$P(s|d) \propto P(d|s)\, P(s) = \mathcal{G}(d - Rs, N)\, \mathcal{G}(s, S). \tag{2.20}$$

If Eq. 2.20 is substituted into Eq. 1.17 and the integral is solved, the formula for the map results in

$$m = (S^{-1} + R^\dagger N^{-1} R)^{-1} (R^\dagger N^{-1} d). \tag{2.21}$$

This filter is called Wiener Filter (WF) and it is composed of different operators applied to the data (Enßlin, Frommert & Kitaura, 2009). From the right side of Eq. 2.21 and in IFT context, $R^\dagger N^{-1} d$ is known as the information source and is denoted by j. The $(S^{-1} + R^\dagger N^{-1} R)^{-1}$ is known as the information propagator operator, D. It consists of an inversion of two terms: S^{-1} and $R^\dagger N^{-1} R$. For well observed regions by the instrument, the response is powerful and noise values decrease. Therefore $R^\dagger N^{-1} R$ is dominant compared to S^{-1}, so that automatically in IFT algorithms just the response is undone in this regime by $m \approx (R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} d$ being the data processed by the noise weighted pseudo inverse of R. On the other hand, $R^\dagger N^{-1} R$ vanishes for areas where the instrument did not spend observing time. In this case S^{-1} is large compared to $R^\dagger N^{-1} R$ and $D \approx S$ propagates any available image information in j over distances of the prior signal correlation length. The reconstructed map can be written as

$$m = Dj. \tag{2.22}$$

The information are gained from the data and are projected back into the signal space by the constructing of the source information, $j = R^\dagger N^{-1} d$. Then the propagator, D, as it was explained above, propagates information from locations, where well observed to locations which were not observed, if these points are a prior assumed to be correlated, as encoded in S. In general, if we are interested in the signal field value at location y, then the propagator operator D_{xy} transports all available and appropriate information source from all locations x, to the locations of y. Therefore:

$$m(y) = \int dx D(x, y) j(x). \tag{2.23}$$

By use of Eqs. 1.15 and 2.20, and also considering just the s-dependent terms for the Hamiltonian, the posterior is given by a Gaussian distribution,

$$P(s|d) = \mathcal{G}(s - m, D), \tag{2.24}$$

with $r = s - m$ indicating the residual error. Therefore the posterior over the signal has the Gaussian distribution around the mean m, and the full width at half maximum (FWHM) is roughly equal to the propagator, $\sqrt{D}$. Therefore the uncertainty is determined by $\sqrt{D}$ where

$$D = \langle r r^\dagger \rangle_{(s|d)}. \tag{2.25}$$

The uncertainty derived for the signal and noise Gaussian distributions will result in a symmetric uncertainty.

As it was mentioned before, the MAP is the maximum of the posterior. If we can find the most probable value for the signal if the data are known, we perform signal inference.

From Eq. 1.15, the MAP is equivalent to minimize the Hamiltonian with respect to s, when $s = m\,(\mathcal{MAP})$. The minimum of the Hamiltonian helps us to find the MAP

$$\frac{\delta H(d,s)}{\delta s}|_{s=m(\mathcal{MAP})} = 0. \tag{2.26}$$

The second derivative of the Hamiltonian with respect to s also determines the local curvature and it is known as the Hessian matrix. The posterior uncertainty is given by

$$(D^{\mathcal{MAP}})^{-1} = \frac{\delta^2 H(d,s)}{\delta s^2}|_{s=m(\mathcal{MAP})}. \tag{2.27}$$

2.3.2 Power spectrum

Correlations between pixels in the sky image are characterized by a power spectrum in harmonic space for fields that are statistical homogenous. Including available correlation information into the field estimation leads to a smooth reconstructed map and reduces effectively the number of free parameters of the signal field. The exploitation of correlations is necessary, because otherwise infinitely many unknown field values of the signal can not be specified given only the finite amount of measurement data constraining the field. In sky regions where not probed by the instrument, no direct constraints on the field values exist. However, in case we have knowledge about the spatial correlations of the field and also knowledge the field values in some areas, it is possible to extrapolate there into regions insufficient probed by the instrument.

As it was mentioned before from Eq. 2.16, S describes the typical fluctuations of the signal. In other words, it specifies the smoothness of the field. As mentioned before, for a statistical homogenous fields s, S is best specified in Fourier space, since it is diagonal in this basis. For the Fourier transformation of a field, we will use the following consideration

$$s(k) = \int e^{ik\cdot r} s(r) d^2 r. \tag{2.28}$$

$s(k)$ is the complex amplitude of the field at wave vector k with corresponding wavelengths: $\lambda = \frac{2\pi}{k}$. The correlation between different field amplitudes is diagonal, if the corresponding real space field is statistically homogenous

$$S_{kk'} = \langle s(k)s(k')\rangle = \iint d^2 r d^2 r' e^{ik\cdot r} e^{-ik'\cdot r'} \langle s(r)s(r')\rangle_{(s|d)}. \tag{2.29}$$

into which we insert the transformation $r' = r + a$,

$$\begin{aligned} S_{k'k} &= \iint d^2 r d^2 a e^{-ik\cdot a} e^{(k-k')\cdot r} \langle s(r)s(r+a)\rangle_{(s|d)} \\ &= \delta^D(k-k')(2\pi)^2 \int e^{-ik'\cdot a} \xi(a) d^2 a \\ &= (2\pi)^2 \delta^D(k-k') P(k'), \end{aligned} \tag{2.30}$$

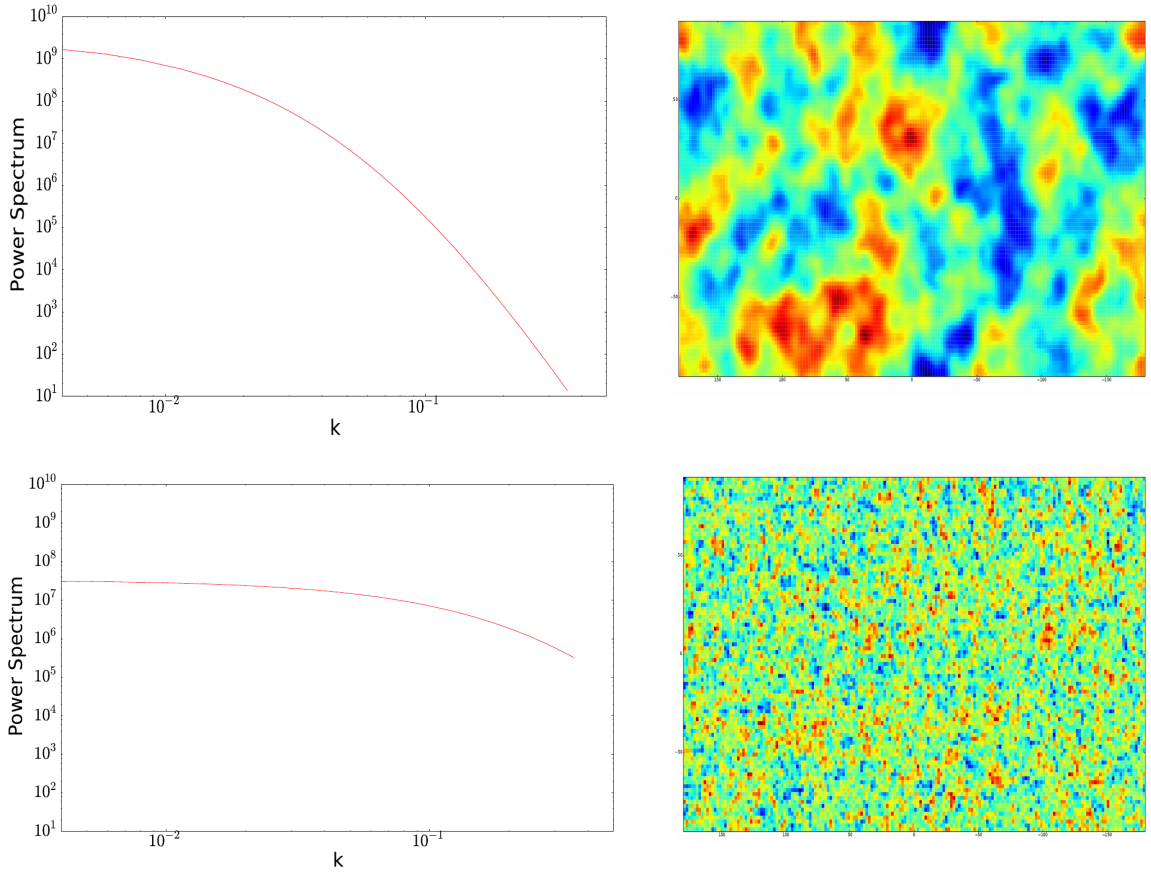

Figure 2.5: Statistically homogenous and isotropic Gaussian random fields drawn from these power spectra

where we defined the power spectrum $P(k')$ to be the Fourier transform of the correlation function $\xi(a)$. The power spectrum becomes 1-D in case of isotropy, since then $P(\overrightarrow{k}) = P(k)$.

In Fig. 2.5, two different power spectra and corresponding Gaussian random field realizations are shown. The power spectrum in the top left panel has more power at low values of k compared to high k values, and this means less power on small spatial scales. Therefore the random field drawn from this power spectrum is smooth. But the second power spectrum shown in the left bottom panel stays roughly constant also at high k and therefore small λ. This means more small scale structures and fluctuations are expected for its corresponding Gaussian random field realizations (right bottom panel).

In the next subsection, the principles of the second IFT filter, D^{3}PO will be discussed and we will see how this algorithm addresses the inference problems.

2.3.3 D^{3}PO

In astrophysical observations, different source types appear in the sky. The various sky contributions imprint diffuse and point-like sources into the image information taken by an instruments. The diffuse emission shows a smooth morphology across large fractions of a sky image. In contrast, point-like sources are well localized sources that one expects to occupy one individual pixel in the image.

There are two major difficulties in the reconstruction of the sky brightness from photon events:

1) The contribution of the diffuse and point-like source emissions are often superimposed on each other which renders the separation of these two components to be a challenging task, especially when the noise is not negligible.

2) γ-ray data is usually subject to Poissonian noise, which is more severe in the regions where the count rates are low. This noise can blur the difference between point-like sources and diffuse emission, which makes the imaging procedure more challenging.

To overcome these difficulties, the D^{3}PO algorithm, has been developed by Selig & Enßlin (2015) in the framework of IFT. D^{3}PO stands for deconvolving, decomposing and denoising of photon observations and it aims at the simultaneous reconstruction of the diffuse and point-like sources. Denoising means removing the noise, deconvolving means removing the effect of the instrument, and decomposing refers to the decomposition of the two emission components.

The signal s is classified by two categories: the diffuse (s) and point-like source (u) emission. The flux originating from these sources is modeled in the form of log-normal distributions

$$\rho = \rho_{\text{diffuse}} + \rho_{\text{point-like}} = \rho_0(e^s + e^u), \tag{2.31}$$

in which the total flux ρ is the superposition of two types of sources e^s and e^u. The flux is defined as the average counts per area per time interval in a particular energy. So given in an energy bin the signal unit is ph/cm^2 sec deg^2. The exponential form of our signal parametrization enforces the physical requirement of positivity for the flux. From observational measurements, one expects to detect a certain amount of the counts, called expectation value, λ

$$\lambda = R_0\rho = R_0\rho_0(e^s + e^u) = R(e^s + e^u). \tag{2.32}$$

Notice that the coefficient ρ_0 plays an important role in adjusting λ, and that in the D^{3}PO algorithm it is absorbed into the response R.

Now it is the time to choose appropriate priors of both signals in D^{3}PO, $P(s)$ and $P(u)$. In D^{3}PO, log-normal distribution is used for the different flux priors. Since for the diffuse emission (s), the flux goes smoothly across various pixels in the image, one expects the correlation between pixels of the sky image. To implement such correlations the Gaussian distribution is used for the prior of the diffuse emission given by

$$P(s|S) = \mathcal{G}(s, S) = \frac{1}{\sqrt{\mid 2\pi S \mid}} \exp(-\frac{1}{2} s^\dagger S^{-1} s). \tag{2.33}$$

The difference between the Gaussian prior in D^{3}PO (Eq. 2.33) and that one which is used in the WF (Eq. 2.16) is that here S is unknown, while in the WF case it was assumed that we know S. Therefore, D^{3}PO provides the possibility to infer the power spectrum from which s originates. The power spectrum has to be positive and the diagonal of S. To qualify these requirements S is expressed as

$$S = \sum_k e^{\tau_k} S_k, \tag{2.34}$$

where τ_k and S_k are spectral parameter and projection operator, respectively. e^{τ_k} enforces the power spectrum, to be positive. The projection operator is diagonal in Fourier space

$$S_k = \delta_{kk'} P(k' \epsilon b_k | k, b_k), \tag{2.35}$$

where b_k is the set of all Fourier modes belonging to a spectral band k, for example all $\overrightarrow{k}$ vectors with the same length. This ensures us to have S as a diagonal operator, which has the assumption of statistical homogeneity and isotropy of the signal build in. Therefore it is enough to specify τ_k for the power operator, S.

The power spectrum is unknown. When we do not have any physical preference for a quantity, it is better to assume that the quantity could be within a broad range of possibilities. This leads to assume a logarithmically uniform prior for each element of the power spectrum, including τ_k. Since the exponential distribution for the prior has singularity at low values, an inverse-Gamma distribution for the prior of the spectral parameter, τ_k, is chosen (Selig & Enßlin, 2015)

$$\begin{aligned} P_{un}(e^\tau | \alpha, q) &= \prod_k I(e^{\tau_k}, \alpha_k, q_k) \\ &= \prod_k \frac{q_k^{\alpha_k - 1}}{\Gamma(\alpha_k - 1)} e^{-(\alpha_k \tau_k + q_k e^{-\tau_k})}, \end{aligned} \tag{2.36}$$

where α and q are parameters that specify the power law index and a low amplitude cut off, respectively, and Γ represents the Gamma function. By use of Eq. 2.36 and the probability mass conservation, which states

$$P(e^\tau)\, de^\tau = P(\tau)\, d\tau, \tag{2.37}$$

the following equation results

$$P_{un}(\tau | \alpha, q) \propto \exp(-(\alpha - 1)^\dagger \tau - q^\dagger e^{-\tau}). \tag{2.38}$$

There is a problem which arises often in the reconstruction of the power spectrum. The reconstructed power does not turn out smooth, even in the case of a smooth underlying

power spectrum. Therefore the spectral smoothness of the power spectrum should be enforced in the power reconstruction (Oppermann et al., 2013) via an additional prior term

$$P_{\mathrm{sm}}(\tau|\sigma) \propto \exp(-\frac{1}{2}\tau^\dagger T\tau), \tag{2.39}$$

with

$$\tau^\dagger T\tau = \int d(\log k)\, \frac{1}{(\sigma_k)^2}\, (\frac{\partial^2 \tau_k}{\partial(\log k)^2})^2, \tag{2.40}$$

where $\sigma = [\sigma_k]_k$ are Gaussian standard deviations controlling the roughness of the inferred spectrum. Combining this power smoothness power Eq. 2.39 with Eq. 2.38 gives us

$$\begin{aligned} P(\tau|\alpha, q, \sigma) &= P_{\mathrm{un}}(\tau|\alpha, q)\, P_{\mathrm{sm}}(\tau|\alpha) \\ \propto\ &\exp(-(\alpha-1)^\dagger\tau - q^\dagger e^{-\tau})\, \exp(-\frac{1}{2}\tau^\dagger T\tau). \end{aligned} \tag{2.41}$$

After modeling the prior of the diffuse emission, the aim is to specify an appropriate statistical description of the point-like sources. These sources are mostly distant objects and typically appear in single pixels of the sky image. For the point-like sources D^{3}PO assumes that there is no correlation between different pixels. Therefore the Gaussian distribution is not anymore an appropriate model for the prior of the point-like sources. Here we suppose that the Universe is homogenous and isotropic everywhere. So in this case the observed number of sources below some distance is proportional to the volume of a sphere with this radius and consequently to the cubed distances from the source location: $N \propto r^3$. On the other hand, the brightness of the point-like sources decreases with the square root of the distance: $B \propto r^{-2}$. The relation between the number of point sources and their brightness tells us that the probability of finding point-like source of a given brightness B falls off with some power of B ($\beta = 3/2$ in case of an Euclidean, infinite Universe). Since this has a singularity at zero, it would be a good choice to assume the inverse-Gamma distribution for the prior. Combination of this idea with probability mass conservation leads us to have the following distribution for the point-like prior

$$P(u|\beta, \eta) \propto \exp(-(\beta-1)^\dagger u - \eta^\dagger e^{-u}), \tag{2.42}$$

where β and η are shape and scale parameters. From Eqs. 2.33, 2.38, 2.39 and 2.42 the prior for the diffuse and point-like sources includes three items: the diffuse component of the signal, its power spectrum and one for the point-like sources emission. What remains to formulate the posterior is the likelihood, $P(d|s, u)$. Since the γ-rays data consist of counts, the Poisson distribution is used

$$P(d|\lambda) = \prod_i P(d_i, \lambda_i) = \prod_i \frac{1}{d_i!}\lambda_i^{d_i} e^{-\lambda_i}, \tag{2.43}$$

where $\lambda = R(e^s + e^u)$ describes the expectation value of counts given the sky model. When all prior terms and likelihood are combined, the posterior is

$$P(s,\tau,u|d) = \frac{P(d|s,u)\,P(s|\tau)\,P(\tau|\alpha,q,\sigma)\,P(u|\beta,\eta)}{P(d)}. \tag{2.44}$$

To get the MAP when the posterior is obtained, the Hamiltonian also can be used

$$\begin{aligned} H(s,\tau,u|d) &= -\log\, P(s,\tau,u|d) \\ &= H_0 + 1^\dagger R(e^s + e^u) - d^\dagger \log\left(R(e^s + e^u)\right) + \frac{1}{2}\log\left(\det[S]\right) \\ &+ \frac{1}{2}s^\dagger S^{-1} s + (\alpha - 1)^\dagger \tau + q^\dagger e^{-\tau} \\ &+ \frac{1}{2}\tau^\dagger T \tau + (\beta - 1)^\dagger u + \eta^\dagger e^{-u}. \end{aligned} \tag{2.45}$$

If the derivative of the Hamiltonian is taken with respect to individual parameter s, u, and τ, it would be possible to infer the diffuse emission, point-like sources emission, and also the power spectrum of the diffuse flux, respectively. This inference for the diffuse emission is stated by (Selig & Enßlin, 2015)

$$\frac{\partial H}{\partial s}|_{\min} = 0 = (1 - \frac{d}{\lambda})^\dagger R * e^{m^{(s)}} + (S^\star)^{-1} m^{(s)}, \tag{2.46}$$

where λ is the expectation value of the data, $\lambda = R(e^{m^{(s)}} + e^{m^{(u)}})$ and $S^\star$ follows Eq. 2.34. Also $*$ denotes componentwise multiplication. The same strategy also is implemented for the inference of the point-like sources emission

$$\frac{\partial H}{\partial u}|_{\min} = 0 = (1 - \frac{d}{\lambda})^\dagger R * e^{m^{(u)}} + \beta - 1 - \eta * e^{-m^{(u)}}, \tag{2.47}$$

In Eq. 2.46, there is an implicit coefficient ρ_0 which is introduced before in Eq. 2.32. This coefficient identifies the mean of the prior. Therefore by the mean of the prior and our previous knowledge, we can identify the point which the prior fluctuates around. To show this issue, we starts from Eq. 2.33. This prior varies around zero. If we want to change the mean of the fluctuation to a in the prior, Eq. 2.33 becomes

$$P(s|S,a) = \mathcal{G}(s-a,S) = \frac{1}{\sqrt{\mid 2\pi S \mid}} \exp(-\frac{1}{2}(s-a)^\dagger S^{-1}(s-a)). \tag{2.48}$$

Since the goal is to study the changes with respect to s, all s-independent terms are dropped

$$P(s|S,a) = \mathcal{G}(s-a,S) \propto \frac{1}{\sqrt{\mid 2\pi S \mid}} \exp(-\frac{1}{2}s^\dagger S^{-1} s + s^\dagger S^{-1} a). \tag{2.49}$$

If we take into account the contribution of this prior to the derivative of the Hamiltonian with respect to s, we get

$$\frac{\partial H}{\partial s}|_{\min} = 0 = (1 - \frac{d}{\lambda})^{\dagger} R_0 * e^{m^{(s)}} + (S^{\star})^{-1}(m^{(s)} - a). \tag{2.50}$$

Introducing $m^{(s)'} = m^{(s)} - a$ we get

$$\frac{\partial H}{\partial s}|_{\min} = 0 = (1 - \frac{d}{\lambda})^{\dagger} R_0 e^{a} * e^{m^{(s)'}} + (S^{\star})^{-1} m^{(s)'}, \tag{2.51}$$

in which $R = R_0 e^a = R_0 \rho_0 = R$. The mean of the prior appears as e^a and it is called ρ_0 in D^{3}PO. By choosing this factor we can specify the point around which the prior fluctuates.

The second derivative of the Hamiltonian is the curvature around the minimum, and its inverse is approximately the uncertainty covariance,

$$\frac{\partial^2 H}{\partial s\, \partial s^{\dagger}}|_{\min} \approx D^{(s)^{-1}}, \quad \frac{\partial^2 H}{\partial u\, \partial u^{\dagger}}\_\min \approx D^{(u)^{-1}}. \tag{2.52}$$

By taking the first derivative of the Hamiltonian with respect to τ of Eq. 2.45, it is possible to derive the formula for the reconstructed power spectrum as follows

$$e^{\tau} = \frac{q + \frac{1}{2}\left(\mathrm{tr}[m^s (m^s)^{\dagger} S_k^{-1}]\right)_k}{\gamma + T\tau}, \tag{2.53}$$

where $\gamma = (\alpha - 1) + \frac{1}{2}(\mathrm{tr}\,[S_k S_k^{-1}])_k$.

In the mentioned reconstructions, just the inference with respect to one parameter (for example only s or u) was considered. However, D^{3}PO is able to reconstruct simultaneously every two or three parameters and this issue is really the powerful point of the algorithm.

D^{3}PO simultaneously reconstructs the diffuse and point-like sources emission from the observed data. As well, it is able to infer the power spectrum of the diffuse emission, which makes the reconstruction procedure more precise. Moreover, this algorithm chooses the Poisson distribution for the likelihood for the data. Concerning the priors, more physical and precise approximations are made compared to the WF. As a proof of this statement, we will see in the next chapter simulation results higher accuracy of D^{3}PO. These two filters and their results are going to be shown and compared with each other in Chapter 4.

Chapter 3

INTEGRAL/SPI Instrument, data, response and background

3.1 Introduction

γ-ray radiation is the most energetic part of the electromagnetic spectrum (Fig. 3.1) photon energies above 100 keV. γ-ray studies help us to identify compact objects, e.g., neutron stars, pulsars, quasars and even bursts, since their emissions have their peak luminosity at γ-ray energies. In the γ-ray domain the wavelength of the radiation is shorter than the distances between atoms and molecules in the interstellar medium. Therefore, the particle description of the γ-ray is dominant to the wave behavior of the radiation. For this reason we encounter the counts in γ-ray observations.

Fig 3.2 shows the Earth's atmosphere transparency against the cosmic electromagnetic radiation in different wavelengths (Diehl,R., 2001). From the figure in the region of less energetic γ-ray, atoms and molecules existing in the interstellar medium interact with γ-ray (however it is rare), and the processes including absorption and scattering prevent the astronomical observations performed in the ground. For this reason, γ-ray telescopes should be operated outside the atmosphere at high altitude.

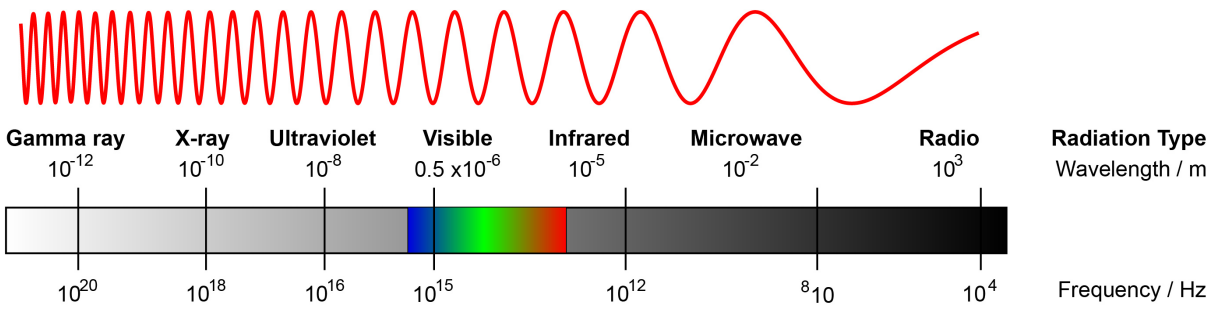

Figure 3.1: Electromagnetic spectrum

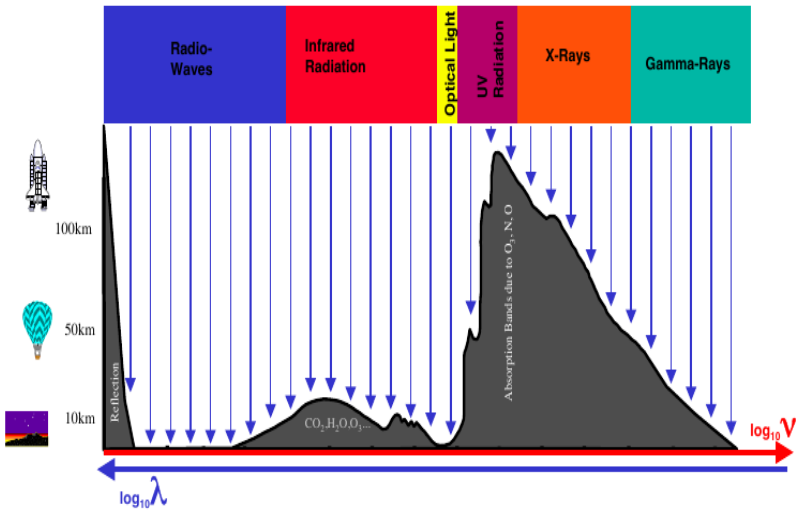

Figure 3.2: The Earth's atmospheric opacity against cosmic electromagnetic radiation over the range of frequencies (Diehl,R., 2001)

There are two problems make γ-ray observations challenging:

1. The γ-ray photons originate from distant objects so that the emission happens rarely and their fluxes are low. Therefore material used for the detectors required to be sufficiently dense to have a high probability of interaction with the γ-ray.

2. The detector count values are dominated by the background mostly originate from cosmic rays while hitting the instrument. Cosmic rays are very energetic particles propagating at the speed of light. Due to this high background, the γ-ray instruments have anticoincidence shield to reduce the background counts to high degree.

During recent years many attempts are implemented to study γ-ray radiations. NASA's Compton Observatory had two telescopes: COMPTEL and EGERT. Their energy range was between 1 MeV and 1 GeV. There were also other telescopes like OSSE and BATSE on the Compton Observatory and SIGMA on the Russian mission GRANAT. At present the Fermi γ-ray Space telescope, launched on June 2008, observes the cosmos using the highest-energy form of light in energy range of 8 keV to 300 GeV. Also the European Space Agency (ESA) dedicates one of its important mission to detect this high-energy radiations. This satellite is called INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL) and it was successfully launched in 2002 to observe the celestial radiations in the energy range of hard x-ray and γ-ray.

INTEGRAL has two instuments on board: IBIS (Imager) and SPI (spectrometer). Both instruments use coded-aperture masks for imaging purposes. IBIS observes rays from energy range of 15 keV (hard x-ray) to 10 MeV (γ-ray). The SPI operates in energy range of 20 keV to 8 MeV. Details about SPI will be described in Sec .3.2. Fig. 3.3

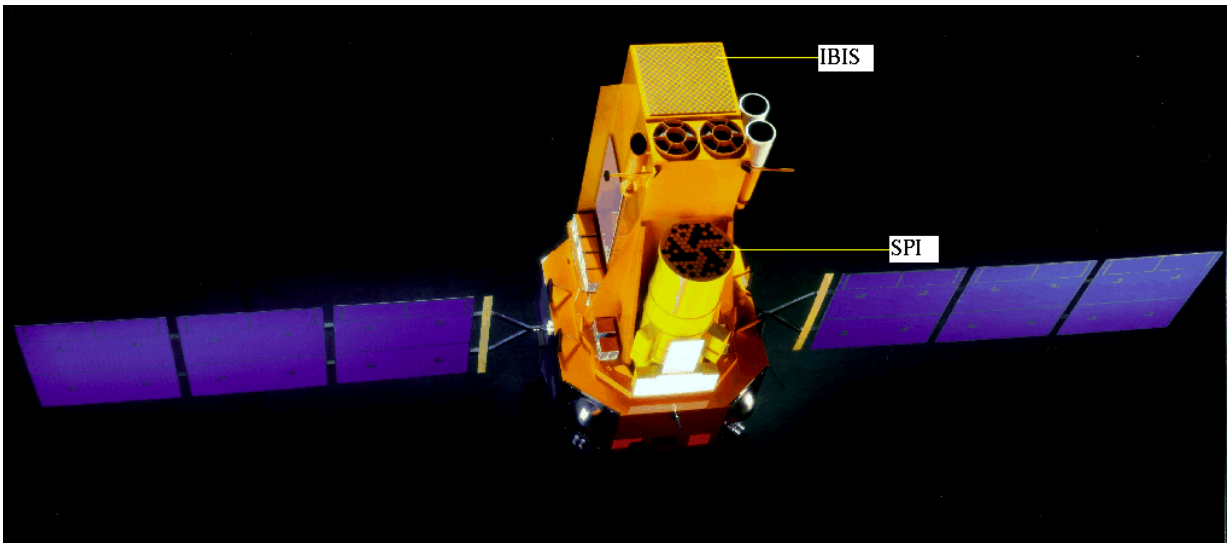

Figure 3.3: View of INTEGRAL, http://www.isdc.unige.ch/integral/outreach/integral

shows INTEGRAL spacecraft with its imager and spectrometer. Sec. 3.3 addresses the instrumental response and more mathematical details of the response form applied in IFT. Then in Sec. 3.4 the physics of the instrumental background and its related hardware will be explained. SPI datasets will be disscussed in the final section of this chapter.

3.2 SPI

SPI measures energy (20 keV- 8MeV) with high energy-resolution of 2.2 keV (FWHM) at 1.33 MeV. This allows to measure the γ-ray energy with an extraordinary resolution which is at least 20 times better than previous γ-ray instruments. Its angular resolution is 2.5° within 16° field of view (FoV).

The core of the instrument is a γ-ray camera composed of 19 detectors located in an hexogonal array. The hexagonal shape is chosen to minimize the volume occupied by the detectors and therefore minimize the weight of the system which surrounds the camera. The material used for these detectors is high-purity Germanium (Ge) with a total detection area of 500 cm^2. There is a central bore between each detectors which is 6 mm in diameter and 5.5 cm in length. Each individual detector has an hexagonal shape, 3.2 cm on side, 7 cm deep, and a center-to-center distance of 6 cm. The mean volume for these detectors is 178 cm^3 and the total weight is 18 kg. In Figure 3.4, an overall cut-away view of the instrument is represented.

The SPI detectors are actively cooled down to achieve the optimum sensitivity and resolution and to reduce the radiation damage. Cryocooler connected to a cold plate reduce the detectors' temperature (85-90 K). The detector assembly is placed inside a cold box, kept at 210 K by the passive cooling system. Temperatures are regularly monitored and information can be used for the data processing.

So far 4 of the 19 detectors have been disabled: detectors 2, 17, 5, 1 died in 2003, 2004, 2009, 2010 on revolution 140, 215, 776, 929, respectively. From revolution 929, the instrument effective area is about 85% of the optimal one.

The Ge crystal structure can be damaged by radiation. To overcome this problem, the detectors go through an annealing operation every 6 to 12 months. The detectors are heated to 105 ℃ for few days. The instrument is not available during the heating and the following cooling phases.

The Ge detectors are shielded against the background (photons and particles) by the Anti-Coincidence Shield (ACS) from sources outside the FoV. For this reason, the energy resolution of SPI is close to the best ever (Vedrenne et al., 2003).

There are two types of event which are recorded by the detectors. When single detector events (SE) hit the detectors, their energy deposit is left in only one detector, while multiple detector events (ME) interact through Compton scattering and deposit their energies in more than one detector. The shape of these events are different from each other and there is a system in SPI called pulse shape discrimination (PSD) that can recognize the events type. Each photon that is absorbed in a Ge detector will give a pulse which is sent to the electronics. The electronics register the incoming pulses and tags good photons originating from the single events with their energies, arrival times and events type.

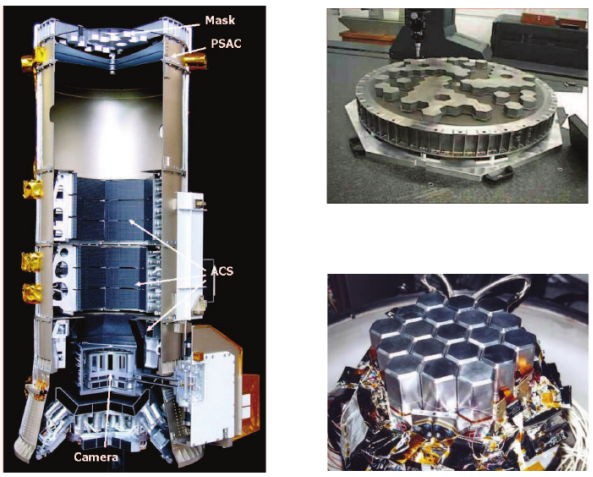

Figure 3.4: Left : A cut-away view of the SPI instrument. The mask, plastic scintillator, camera and ACS subsystems are highlighted. Right : view of the coded mask (top) and detector plane (bottom) (INTEGRAL Science Center, 2012).

When Ge detectors are damaged by radiation, their energy resolution is damaged. This radiation are either primary radiations or secondary ones. Primary radiations come directly from the sky, while secondary radiations first reach other parts of the instrument and then hit the detectors. This effect is due to the hole traps that happen in center of the detectors, and cause decrement in the energy resolution. To minimize the probability trapping of the radiation in the hole, the length of the detector traversed by the holes has to be minimized.

A coded mask is suited at 1.71 m above the detector plane and is made of tungsten blocks. It is composed of a circular passive array of 127 hexagonal pixels in which 63 pixels are opaque and 64 ones are transparent to the γ-ray. The opaque cells are made of a tungestan alloy and are 30 mm thick. This allows an opacity better than 90% over the entire energy range.

In Fig. 3.4 a general view of SPI is illustrated (left panel) including the detectors (bottom right panel) and coded mask (upper right panel).

3.3 Observational challenges

3.3.1 Dithering strategy

Our goal in this thesis is to determine the intensity of each pixel of the sky image. The final FoV in both negative and positive directions is $\sim 30^\circ$. Since the angular resolution of SPI is $\sim 2^\circ$, the observed area by the instrument in one exposure time would be roughly $(30/2)^2 = 225$. This means there are 225 unknown pixels that should be identified in this observation. But, for a single exposure with only one source, the recorded counts are not enough for 225 unknown parameters. Since for one pointing (exposure) only 19 data points are registered (which is equivalent to 19 equations). To obtain these 225 unknown pixels we have to increase the number of equations. Therefore multiple exposures are performed for a given field. These exposures are implemented by re-orientation of the instrument by a slight offset. The regular changes in the direction of the observations is called dithering strategy and it provides a sufficient large number of equations to solve the unknowns. This individual direction toward the sky is called pointing and it lasts between 30 and 60 minutes. Notice that this offset has to be small compared to FoV. Its direction varies around a target by steps of 2° within a five-by-five square pattern (Bouchet et al., 2013).

3.3.2 Coded mask

To make the sky image of the γ-ray emission, a coded mask above the detectors is required. When the instrument observes a source emitting γ-ray in the sky, the light from the source projects shadow of the mask onto the detector plane. The length of the opaque and transparent pixels in the coded mask is designed in a way which can only produce one unique shadow from a particular direction of an object in the sky. If the instrument changes its direction by a little offset, providing the change is small compared to FoV of the instrument, tracing of the two different shadows from a same source in two directions can help observers to localize the source. But if direction changes of the instrument is not small, a composite image will be created from maybe two sources or more. The different counts from these various sources are illustrated in Fig. 3.5 in red and black. There is a computer in SPI which can separate this composite image and it creates the scene of individual source in the sky.

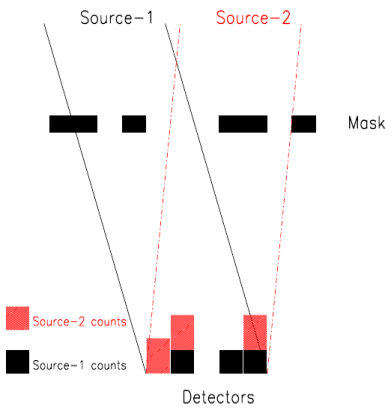

Figure 3.5: SPI image principle. The mask consists of elements transparent or opaque to the radiation. The opaque elements are shown in black. The shadow of the sources cast into the detectors plane and it depends on the source direction. Here the counts in the different detectors of source-1 and source-2 are shown in black and red (Bouchet et al., 2013).

3.3.3 Instrumental response

Consider a source emits an emission and a detector in the instrument receives it. In this case we want to study the spectrum of the counts. A γ-ray line is a flux of photons, each of them with specific energy. An ideal detector reports exactly the energy with which the photons are emitted. But in practice no detector is ideal. Every instrument has limitations for the energy resolution. For example in SPI this resolution is 2.3 keV. If the detectors can measure 500 keV, the next energy detection would be 502.3 keV. In the case of 500 keV for the energy of the flux, the received spectrum is a delta-peak function which looks like for an ideal detector. But if the flux energy would be between 500 and 502.3 keV, a spectrum for the energy will be detected, which has a specific line-width. This spectrum shape which forms the response function can be modeled as a Gaussian distribution. As

more narrow the line-width, as more ideal is the response. However the response of SPI degrades over time and the width of its Gaussian-shape broadens. For this reason, twice a year, the procedure of annealing is applied to the detectors, which was explained in Sec. 3.2.

The unit of the response is $cm^2\times$ Sr $\times$ sec. Sr (Ω) stands for steradian which is the unit of the spatial angle in SI, and for a full sphere is given by

$$\Omega = \frac{A}{r^2} = \frac{4\pi r^2}{r^2} = 4\pi. \tag{3.1}$$

From the response unit we can find that the response contains the effective area of the detectors. Since this area is not same for all the detectors, the multiplication of the area is contributed to the response elements (cm^2). The absolute determination of effective area of the detector camera was obtained from measurements performed just before the launch of SPI in 2002 October (Vedrenne et al., 2003). Furthermore, since not all the pixels in the sky image have the same size, response elements are also multiplied by the angle (Sr). As well, each pointing lasts between 30 and 60 minutes and so the time spent for all pointings are not equal. For this reason, the response elements are multiplied by the time duration (sec). Therefore, IRF (Instrumental Response Function) is made with respect to the physical characteristics of the instrument, and it is stored in a set of FITS (Flexible Image Transport System) files. The physical characteristics of SPI are different for each detector, pointing, energy and the position which the instrument looks at the sky. Therefore, the response is function of pointing (p), detector (d), energy (e) and also position in the sky (Galactic longitude: l, Galactic latitude: b). Since one energy bin is assumed to be considered in most of this thesis, the dependency of the response on energy is dropped. However, we will see in chapter 5 that the analysis in the energy range of 500-520 keV will also be investigated in the sky imaging. The modeled response for individual pixel in the sky image is contributed according to Eq. 1.12. As it was mentioned in chapter 2, in IFT the response is considered as an operator, and a matrix is associated to it. If we also consider the contribution of the SPI background

$$d_{p,d} = \sum_{l,b} R_{p,d,l,b} s_{l,b} + b_{p,d} + n_{p,d}, \tag{3.2}$$

where p and d are the pointing and detector, respectively. Data (d) a 1D vector with number of pointings $\times$ number of detectors ($n_p\times$ n_d), which is equal to 1,398,210 counts. The background (b) and noise (n) live in the same space that the data does. The indices l and b refer to a particular position in the sky: Galactic longitude and latitude, respectively. This leaves a 2D field for the signal in the form of a plane which shows the image of the sky.

Sparse matrix

As it was mentioned before in Sec. 3.3.1, dithering strategy is performed during observations by SPI to produce multiple exposures and to create more equations in which we

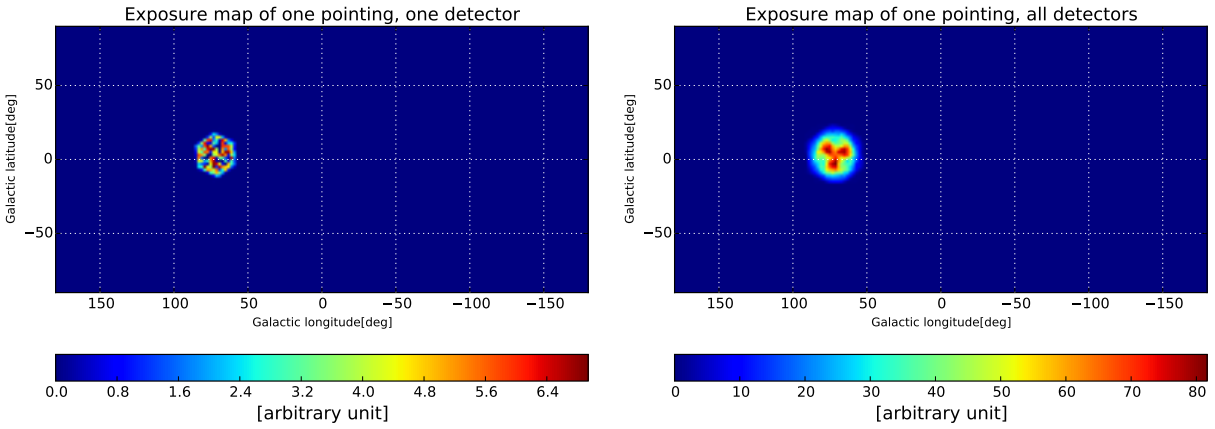

Figure 3.6: Exposure map of SPI for (a) one detector and one pointing, (b) one pointing and all detectors.

can solve unknown intensity of the pixels in the sky image. This strategy creates a large amount of data up to $n_p \times n_d$ data points. Therefore the response which is originated from this large data will also be a large linear system.

Given FITS file of the response matrix (Strong, 2003) in this thesis includes only the observed areas in the sky (colorful pixels in the exposure map of Fig. 3.6) which are different for each pointing and each detector. This initial matrix response has roughly 10^9 elements and it constitutes a part of the sky image. Notice that number of these observed pixels in the sky image for 19 detectors is not same for all pointings. To locate the observed pixels for all the detectors in the whole sky image, sparse response matrix will be replaced in which zero elements are replaced by unobserved areas of the sky image and non-zero are replaced by the initial response matrix which includes the observed pixels. For instance, suppose in one pointing the pixels number of the observed area is 400 pixels for all 19 detectors. Pixels length of the sky image is 2 degree and there are totally 16200 pixels in the sky image, 180 along the Galactic longitude and 90 along the Galactic latitude. Out of these 16200 pixels, only 400 pixels are observed by the instrument and 15600 pixels remain unobserved. This causes more than 90٪ of the elements in the sparse matrix to be zero, and only a few percent remain non-zero. If all pointings would be considered for the final sparse matrix, the sparsity of the matrix increases up to 22 M elements. This large sparse matrix for the response occupies a huge amount of memory in system, which is out of memory of even supercomputers. One way which can help to overcome this difficulty is application of the sparse matrix to the signal field, pointing by pointing. This method decreases greatly the memory usage, however for a large number of pointing it is too expensive. Therefore, it is required to utilize a way in which calculation of the response-signal (Rs) would be less time-consuming, while it does not exceed the memory of the system. Using Einstein summation in python programming lets us avoid looping in the algorithm which performs tensor contraction. Therefore the problem of time is also greatly will be solved so that applying the sparse matrix as an operator on the signal field pointing by pointing becomes

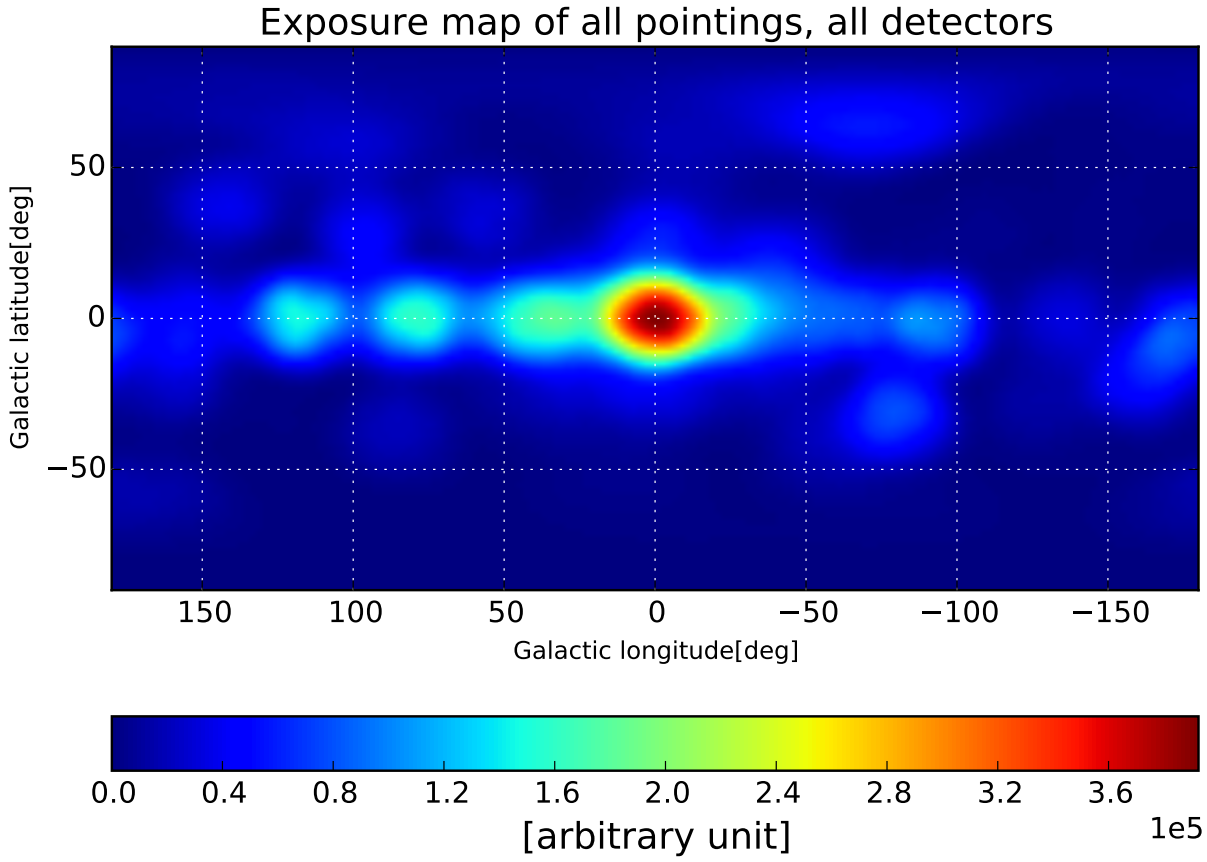

Figure 3.7: Exposure map for all pointings and all detectors.

180 times faster compared to the initial application.

The correctness of the made sparse response matrix can be proved by the exposure maps. Exposure map is back projection of the response (response transpose) on a unit simulated data into the signal space. When the adjoint of the response ($R^{\dagger}$) is applied on a unit data field, the intensity of the instrument in each pixel of the sky is summed over all the detectors and pointings in different measurements. This sky image is called exposure map that identifies which regions of the sky with which intensity are observed by the instrument. From the first data analysis of INTEGRAL/SPI (Knödlseder et al., 2005), we know that the Galactic spatial distribution of the 511 keV emission from positrons annihilation is strongly concentrated in the Galactic center (bulge) and less along the Galactic disk. Therefore it is expected that SPI observes areas which are more likely for detection of the γ-ray emission in the sky. For this reason the exposure maps for all pointings and detectors of Fig. 3.7 are stronger in the Galactic center.

It is also required to have a tool to illustrate the correctness of the response transpose, $R^{\dagger}$. Two vectors A and B are complex conjugate if

$$(AB)^{\dagger} = B^{\dagger}A^{\dagger}. \tag{3.3}$$

Similarly the following scalar product has to be fulfilled if the response transpose is made properly

$$(R(\text{test} - \text{s})R(\text{test} - \text{s}))^{\dagger} = (\text{test} - \text{s})^{\dagger}R^{\dagger}(R(\text{test} - \text{s})), \tag{3.4}$$

where the test-s is a random field.

3.4 Instrumental background

3.4.1 Physics of background

The SPI data is dominated by the instrumental background. The main source of the SPI background is the diffuse emission which are produced by cosmic ray particles.

The background consists of two components. The dominant component which is called continuum has a power-law spectra shape and it is related to those particles that are not totally emitted, but they leave some part of their energies in the detectors. If we study this continuum contribution, we will see no significant and sudden changes occur in its intensity with energy, but the changes are gradual. The other component is the line background which is superimposed to the intensity of the continuum background. This part of the background is related to those primary high energy particles which are promptly produced in the detectors. In the emission of the line background, atomic nucleus are activated by cosmic rays and nuclear de-excitation emissions are totally produced in the detectors. In general, the intensity of the received background by the instrument depends strongly on the irradiation of the background and also the particles of the cosmic rays hit the instrument. Since the sensitivity of the instrument against the sources of the background components is different, the line and continuum contributions should be determined separately.

The background happens as the interactions of the different sources with the instrument. These sources include many astrophysical events in the sky and can be listed as following:

1. cosmic rays (CRs) are dominant sources to produce the instrumental background. They interact inelastically with the instrument nucleus and create high energy secondary particles: p,n,π^0,π^+,γ,..... . These protons and neutrons produce β particles and/or de-excitation photons by nuclear interactions in materials. When β particles are emitted in the Ge detectors, they deposit their energies in the detectors and generate a continuum energy distribution. The de-excitated photons are emitted in passive material or in the Ge detectors and produce γ-ray lines. Also the de-excitated photons are partly might absorbed in the Ge detectors, contributed to the continuum emission and they leave a fraction of their energy in the detectors via Compton scattering.

Fig. 3.8 shows the line and continuum components of the background for the first pointing and different detectors. Since the index starts from 0, the 0th refers to the first detector, 2nd to the third and so on. If we consider both the line and continuum components, by summing over all energies, we will see that the dominant component of the background is the continuum one. Panels (a), (b) and (c) in Fig. 3.9 illustrate this combined background for the first pointing and different detectors: 0, 2 and 10. Also combined, line and continuum background for all pointings are indicated in panels (d), (e) and (f) of Fig. 3.9, respectively.

2. The other source that causes additional counts in the data is due to the repeated annealing. As it was explained in Sec. 3.2, the performance of the Ge detectors would be damaged over a long time (a year or more) of the observations. Measurements after radiation damage shows a 20% increment in the FWHM of a line component background at 1.3 Mev. During this time the data are not available. The detectors are held at temperature of 105° C during two days. While an annealing strategy over two days is quite enough to recover the energy resolution of the detectors, it also produces additional scattering in the data which has contribution for the background.

3. Solar flare is a sudden change in the brightness of Sun. At the moment of this phenomena, the magnetic energy of the solar atmosphere is suddenly released which is equivalent of millions of 100-megaton hydrogen bombs exploding simultaneously. Electromagnetic radiations which are produced in this flare cover all wavelengths from radio to x-ray and γ-ray. It is believed that the observed spikes in the SPI data might come from this radiation.

4. After revolution 900, a strange behavior was seen in the detector rates. Measurements showed that these count rates per detector increased monotonically until revolution ~ 900, and then followed by a gradual decrease until now. This is a large-scale variation in the counting caused by the 11 years solar cycle. Solar cycle is a periodic change in Sun activity which also affects the atmosphere of Earth.

3.4.2 Anticoincidence shield (ACS) and plastic scintillator

To recover the signal in the sky, the aim is to remove any kind of disturbance as much as possible in both computational and technical methods. In terms of technical method the

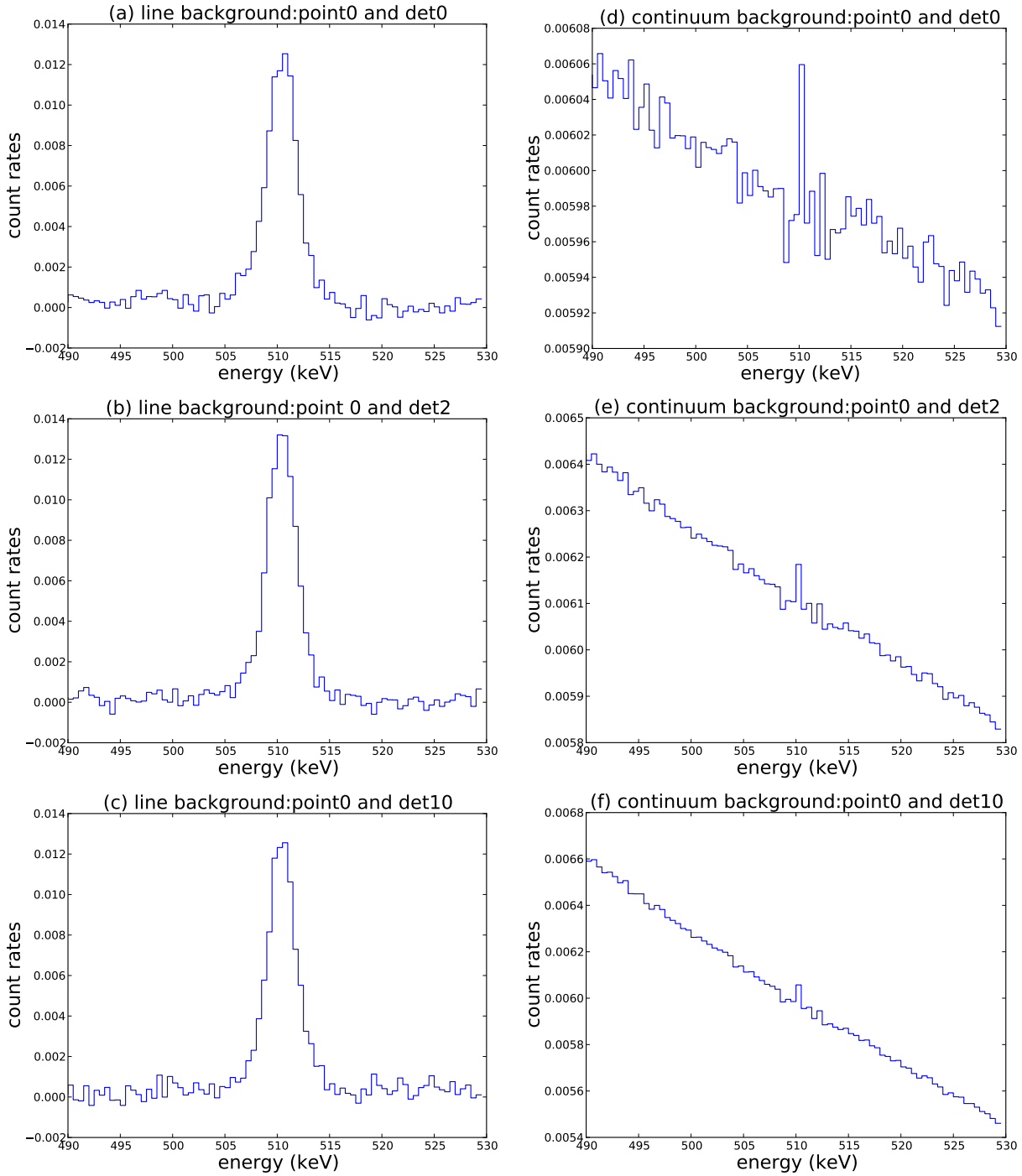

Figure 3.8: left panels show the line background in detector number 0, 2, 10 and right panels show the continuum background for the same detectors. These detectors and pointings are chosen randomly; however, the plots are similar for other pointings and detectors.

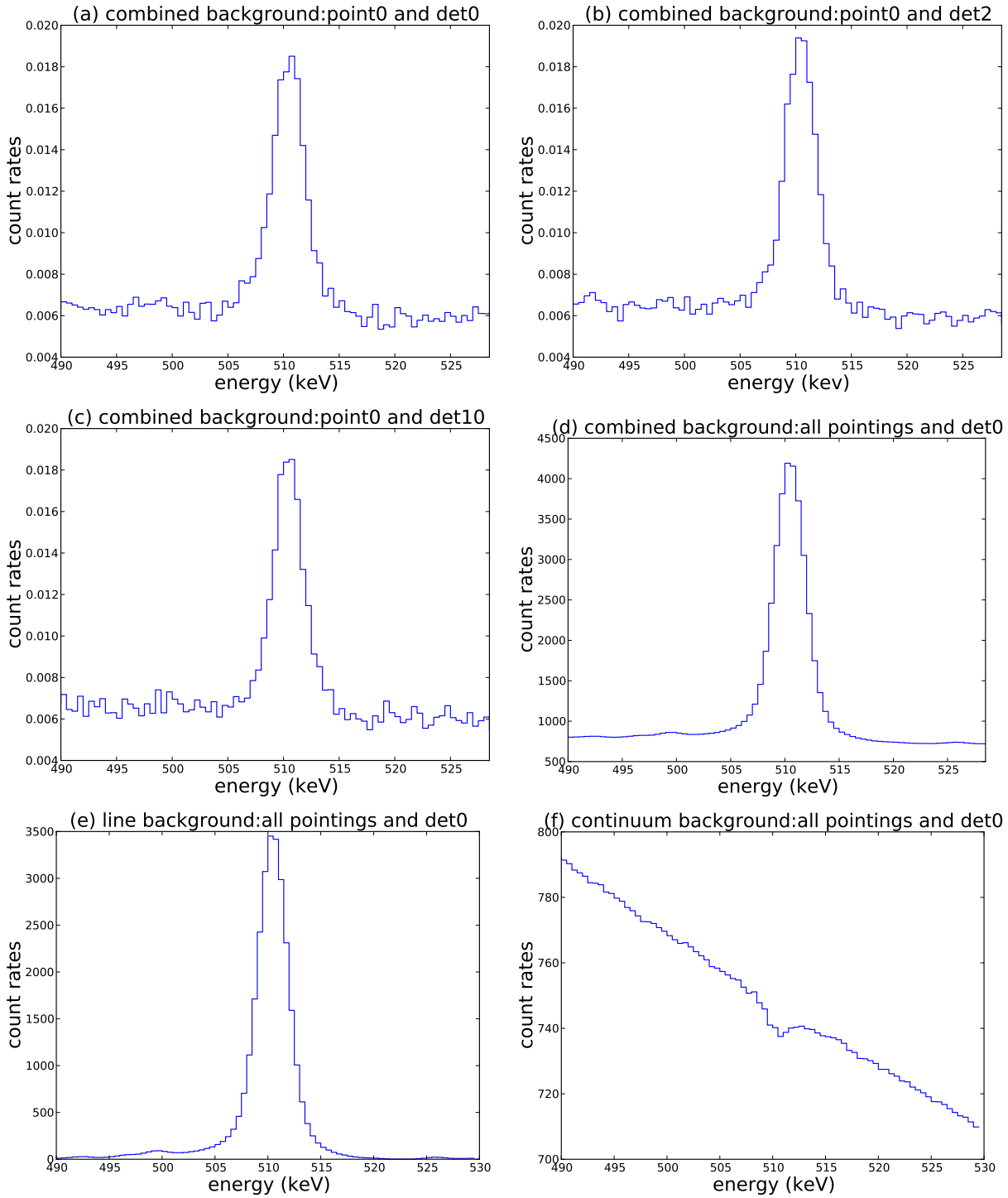

Figure 3.9: panels (a), (b) and (c) show the spectrum of the combined background in different detectors and first pointing, and panels (d),(e) and (f) are the background spectrum for all pointings and first detector in which the background is summed over all pointings.

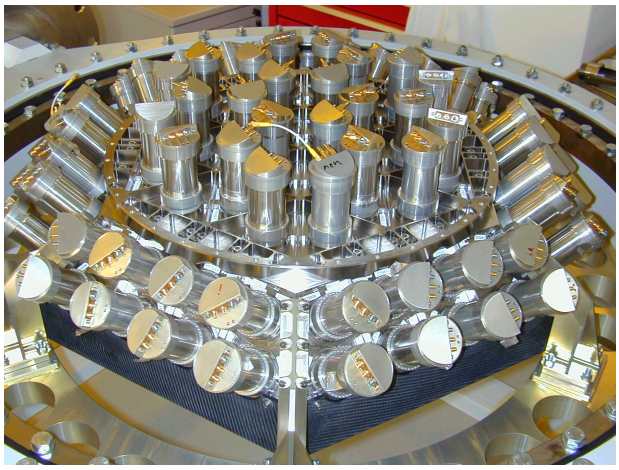

Figure 3.10: SPI Veto System (sci.esa.int/integral/39876.-th-spi-veto-system/)

instrument is equipped with some subsystems. As it was pointed out in Sec. 3.2, the Ge detectors are shielded by a large hexagonal anticoincidence shield (ACS) that consists of 91 Bismuth Germante Oxide (BGO) crystals. This subsystem extends around the bottom and side of the detectors almost completely up to the coded mask. The ACS prevents a large fraction of the line background which are produced immediately. The high energy charged particles from cosmic rays release intensive fluxes toward the instrument. The ACS avoids the entrance of these particles and reduce significantly the rate of the line background events. However this subsystem is not as efficient as for the continuum component of the background, since the lifetime of the radioactive nuclei is more than ACS veto duration, the emitted charged particles that produce signal in the Ge detectors can not be blocked by the ACS. Therefore in addition to ACS, a plastic scintillator (PSAC) is also provided below the mask. Monte-Carlo simulation indicates that a significant background reduction can be achieved by use of this veto system (Jean et al., 2000). A view from SPI veto system has been shown in Fig. 3.10.

3.4.3 Background modeling

Background is the systematic noise which has to be recognized and then removed from data to infer the signal in the sky. Besides hardware efforts which was described in previous section, various computational techniques are developed to predict the background events in the detectors. The approach of predicting the background in order to be substracted from the data is called background modeling, although the sources and physical processes which provides this kind of noise are complex.

One strategy which is used to estimate the background is called INTEGRAL/SPI empty field observations. Since the 511 keV emission is strongly concentrated along the Galactic disk and more in the bulge, the other parts like the Galactic halo are less likely to include the sources. These regions which are called empty field can be used to detect the background. For example exposures in the areas in which $|b| > 30$ satisfies this issue (Bouchet, Roques

& Jourdain, 2010).

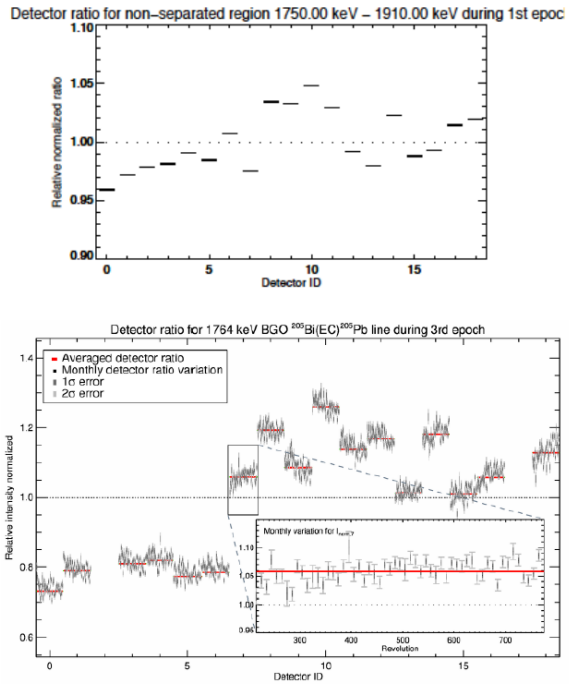

Figure 3.11: Ratio of intensities in different detectors (from Siegert, 2013).

Background modeling which is used in this thesis (Siegert et al., 2016) is based on a pattern called detector ratios. Detector ratios are the ratio of the recorded counts for individual detector per average received counts for all 19 detectors. By use of the coded mask and studying the detector ratios, it is possible to distinguish celestial sources in the sky from the sources of the background. When the instrument is exposed to an object in the sky, the celestial source imprints a particular pattern in the detector so that some of the detectors might record the events and some are in shadow of the opaque pixels of the coded mask. When the instrument reorients to the next pointing, the pattern of the detectors will be changed, while pattern of events for the background sources do not vary with changing of the pointing. Because the coded mask is not relevant here and the background hits the instrument from all directions. This pattern can be studies by the detector ratios illustrated in Fig. 3.11. Relative count rates of detectors with respect to each other remains roughly stable, unless time scales of large enough will be treated. Scaling the pattern in a large time interval is treated as well in the model. However in a limited time up to even hundred pointings, there is no expectation of variations in the detector ratio. This description is the method which is used in this background modeling (Siegert et al., 2016). This model is determined in such a way that the detector ratios are set to 1.0, and can be seen from Fig. 3.11. This is equivalent to assume that the background is isotropic and it hits the instrument with the same intensity.

The scaling of the detector ratios should be treated for both components of the back-

ground in different ways from each others, since the component sources are different.

3.5 SPI datasets

The data is noisy because of the Poissonian nature of photon count rates. In principle noise is any deviation of the real data from the expected data given the correct emissivities. When the raw data are received on the ground, they have to be prepared for users. This process is called calibration. In this procedure, the corrupted data is removed and this means losing information. Corrupted data is for example identified as data in which the photon rate per pointing is varying too much for a single detector. The recorded data during some periods such as solar storm or detector annealing (discussed in section 3.2) are not usable and must be excluded.

Every detector records counts in the total energy range of SPI. The smallest energy bin in which the counts are recorded is 0.5 keV. In most of the mock and real reconstructions performed in this thesis, the employed data are in one energy bin within 510.5-511 keV. Therefore, all the data, response, and background are built for this energy bin. However, the two final reconstructions performed in this work will cover the energy range of 500 to 520 keV, including 41 bins. The data contain 73590 pointings for 19 detectors. The data are arranged in 3D in terms of the three factors: energy, pointing and detector. Therefore, it is possible to keep two of these parameters constant and to check for variations of the data with respect to the other factors. In Fig. 3.12 the count rates in three different cases are investigated. In panel (a) one detector (number 10) and one pointing (middle pointing: 36000) are considered and variations of the count rates over different energies are shown. Here, the energy range of 490 to 530 keV is chosen to illustrate the count rate variations with respect to the energy. Also here any random detector and pointing can be chosen. There is a peak in the count rates around 511 keV which is also repeated for other detectors and pointings. In principle, this increase in the counts at 511 keV motivates us to look for astrophysical sources of it.

Panel (b) in Fig. 3.12 shows how the count rates changes over different detectors, when the energy is constant and the instrument looks at a particular region in the sky (constant pointing). In contrast to panel (a), it can be seen here that the count rates of the 19 detectors are roughly the same.

Panel (c) shows the variations of the count rates over the entire set of pointings for one energy bin (55th) and one detector (10th).

In addition to pointing, detector and and energy, the data also contain information on live and dead times of the detector. In the following section, these information will be discussed.

3.5.1 Live time

SPI performance time in which the detectors record counts is not totally useful. Sometimes during a normal exposure time, for technical reasons the detectors stop working and there

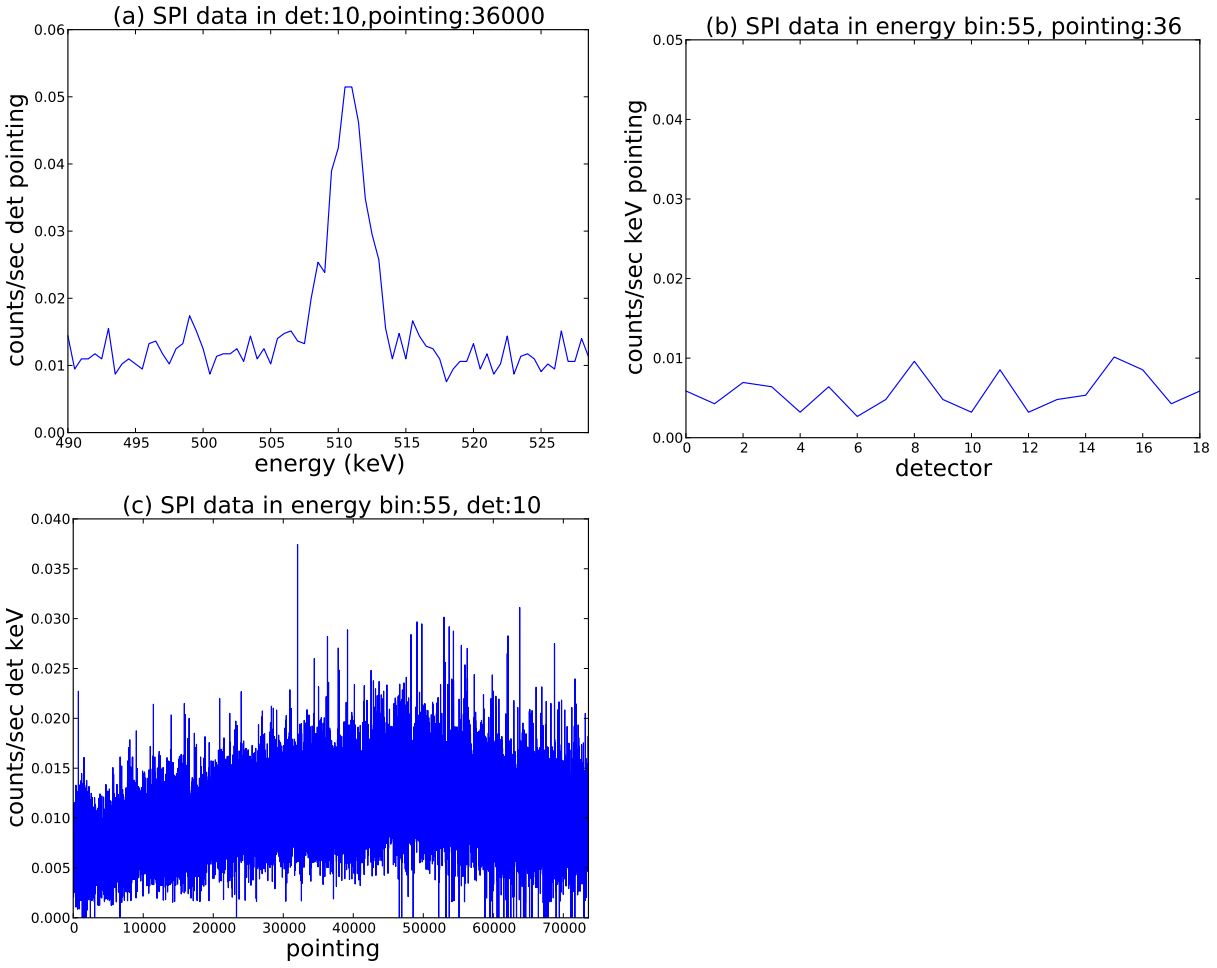

Figure 3.12: SPI count rates variations with respect to (a) energy, when detector number 10 and pointing number 36000 are chosen (b) detector, when energy bin number 55 and pointing 36 are picked up and (c) pointing, when energy bin number 55 and detector 10 are selected. These results are same for any other random detector, pointing and energy bin.

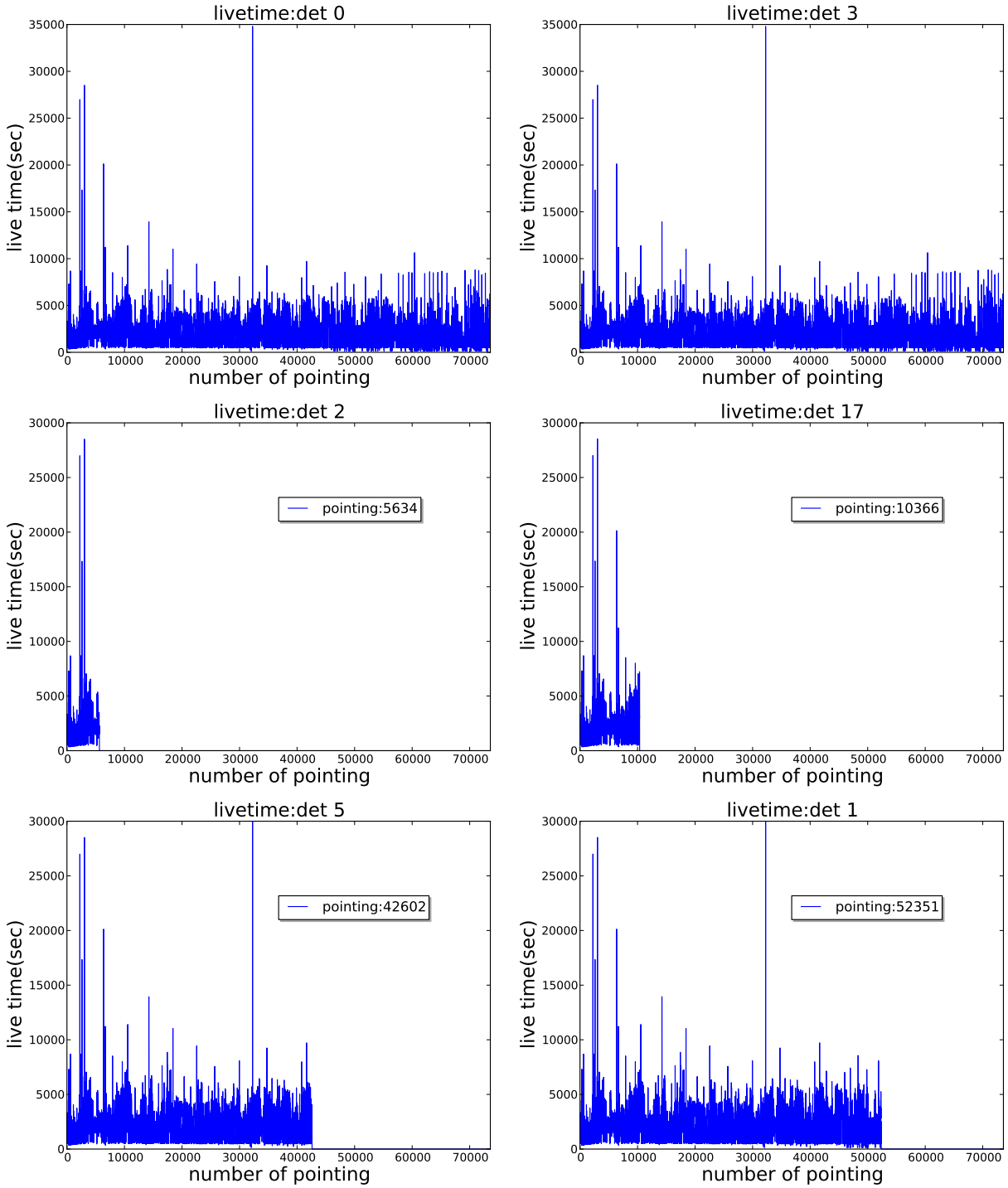

Figure 3.13: live time in terms of different pointings. For the first two plots, the live time over all pointing is non zero; while for det 1, 2, 5 and 17 becomes zero in a particular pointing, which are identified.

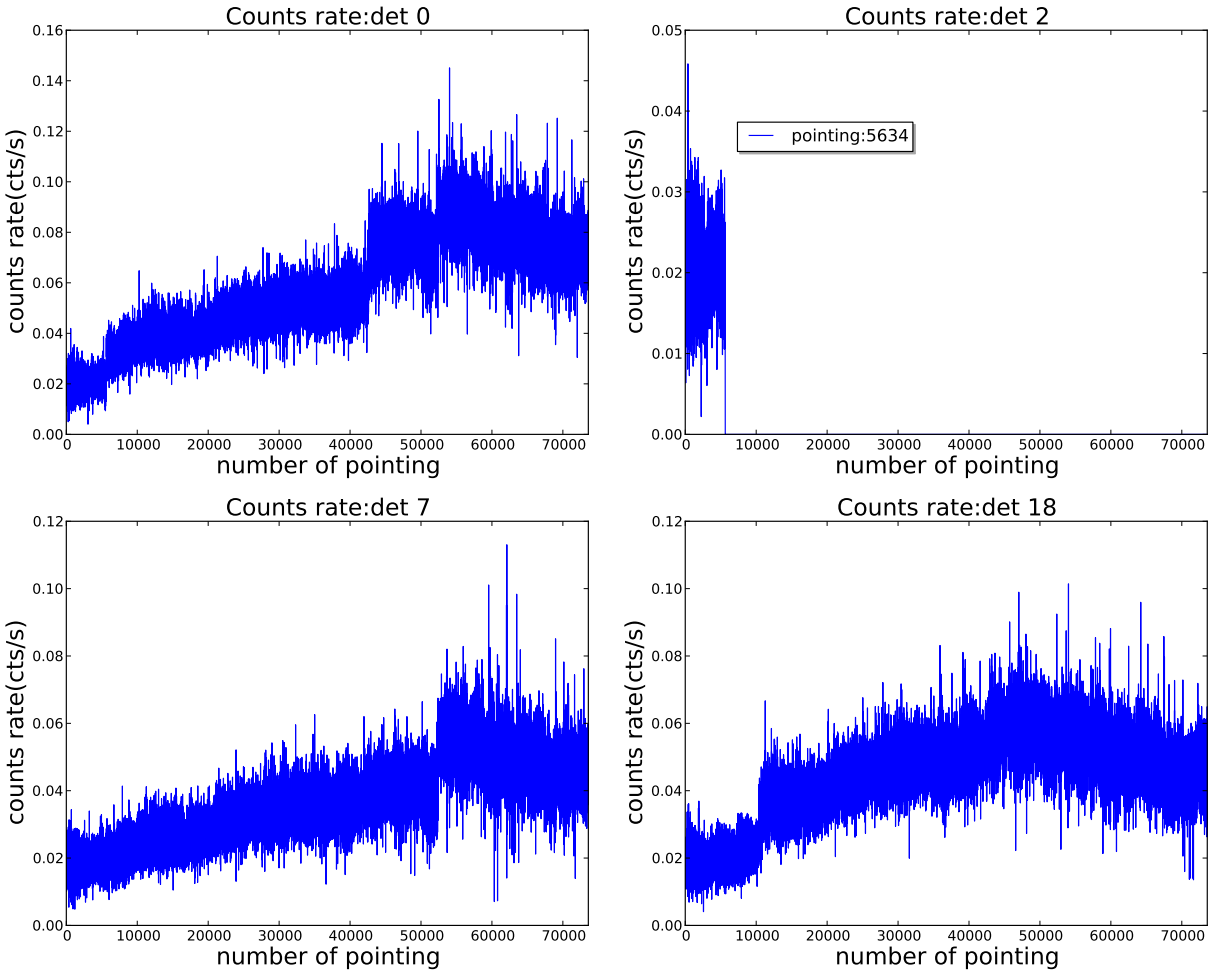

Figure 3.14: The counts rate (cts/s) over 73590 pointings and some detectors.

is no useful scientific data. This useless time is called dead time for individual detector during a particular pointing. On average about 15 % of the observation time is spent as the dead time. The rest time in the observation in which the detectors record counts is called live time, and the plots in Fig. 3.13, show this live time over different pointings for different detectors. For the first two plots, the live time is shown over 73590 pointings, and it fluctuates mainly between 0 to 10000 seconds, except some spikes that can be seen in a few pointings. From different and particular times and for unknown reasons, the detectors numbered 1, 2, 5 and 17 have ceased to function properly and unfortunately they are now disabled. The time for these stop workings are illustrated in the last four plots of Fig. 3.13.

A revolution on average consists of 85 pointings, and this number might be changed for other revolutions. Close to the end of each revolution, the dead time of individual detector increases and careful observers should not take into account their measurements close to these moments. For this reason, the interesting quantity to illustrate the number

of photons will be count rates in which contribution of individual detector for their live time is considered. The count rate can be derived by dividing the counts (the number of photons that is recorded for each pointing, and each detector) by the live time. Following plots in Fig. 3.14 shows these count rates against different pointings. Since detector 5 did not work anymore from Sep 5 in 2009, the count rate will be zero.

Chapter 4

Simulation with INTEGRAL Response

As it was mentioned in Sec. 3.4.1, the background contributes largely to the SPI data. The standard deviation of the data with respect to this background is

$$\sigma_{b,d} = \sqrt{\frac{1}{N}\sum_i (d_i - b_i)^2} \simeq 8.936, \tag{4.1}$$

where N is the number of data bins. The standard deviation is a positive quantity, which shows the amount of variations in the residual after subtracting the expected background from the SPI data. These variations consist of contributions from the sky (our signal) and from the statistical noise n. Since in γ-ray range the photons are detected individually, the noise is Possonian shot noise. Its variance is given by

$$\sigma_n = \sqrt{\lambda} = \sqrt{Rs + b}. \tag{4.2}$$

From SPI data and our background model we infer that the noise variance is $\sigma_n \approx 8.061$. If this amount for the noise variance is subtracted from the residual variations given by Eq. 4.1, one finds that only a tiny sky contribution remains. In order to give an impression how small the sky contribution to the SPI data is, we present normalized back projections of the data, the background model, and the difference of the data and background in the three panel of Fig. 4.1. In the first and second panels of this figure the quantities $R^\dagger(d)/R^\dagger(1)$ and $R^\dagger(b)/R^\dagger(1)$ are respectively shown in the signal space. Here d and b are the data and background and 1 represents a constant unit field in data space. In the lower panel of Fig. 4.1, the difference $R^\dagger(d-b)/R^\dagger(1)$ is displayed. This provides an approximate sky image. The strong structures at $l = 100^\circ$ and $b = 75^\circ$ (bottom left corner) is most likely due to a problem with the background model. The central emission region, with fluxes of about 1 with arbitrary unit is potentially a real sky signal. Note, that its intensity is about 100 times weaker than that of the back projected data or background model. Note further that these back projections are only a very approximate way of imaging and accurate sky signal images will be obtained by application of IFT algorithms on SPI data. However,

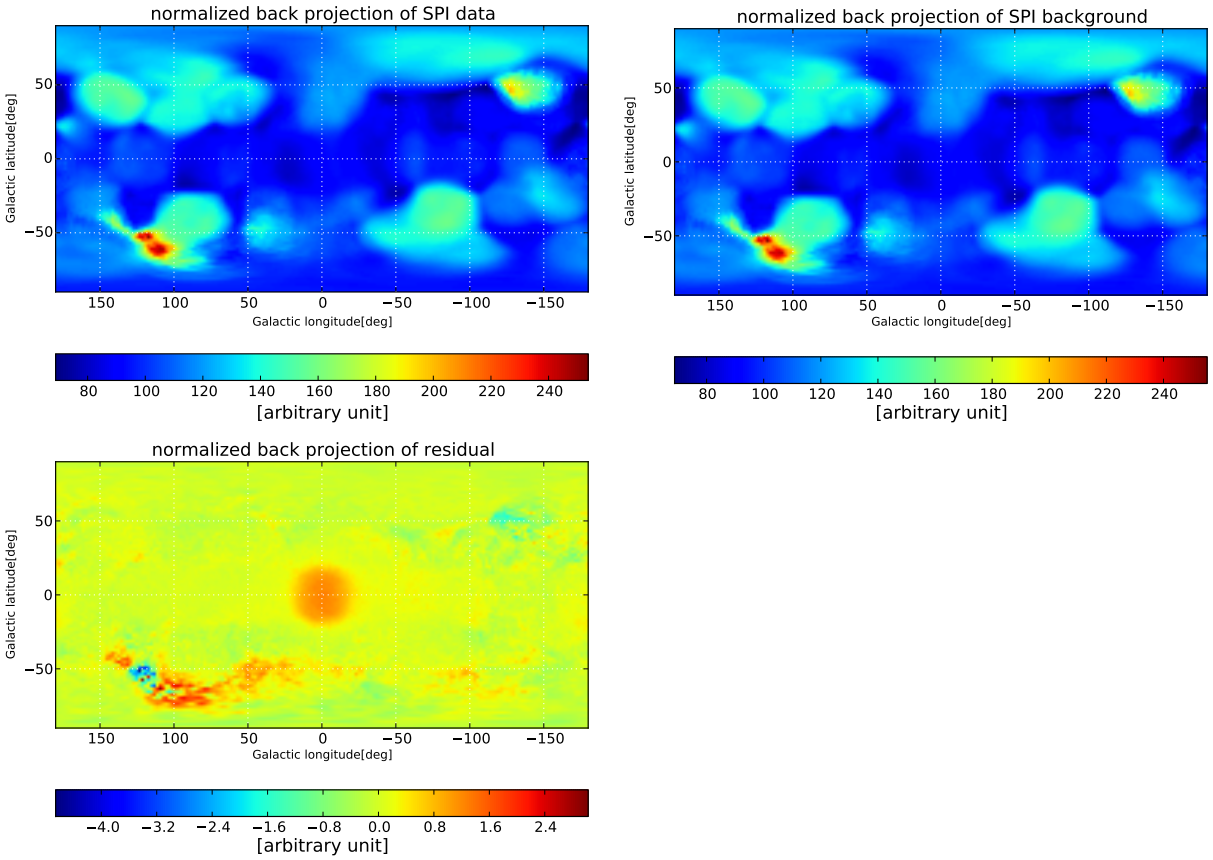

Figure 4.1: Normalized back projection of the data (top left), the background (top right) and their differences (bottom left) onto the sky. The latter can be regarded as a poor man's imaging algorithm, the first two illustrate the challenge image, a sky signal that contributes only 1% to the background dominated data.

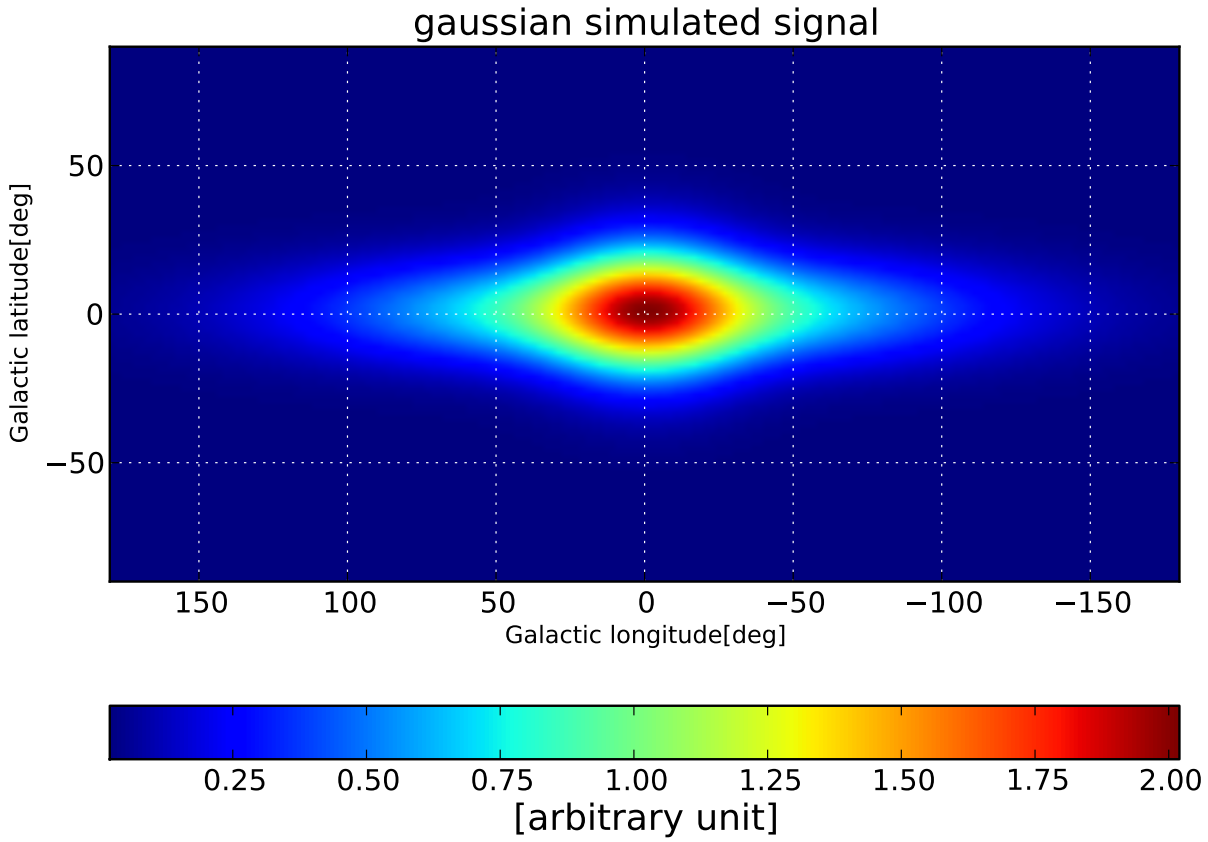

Figure 4.2: Gaussian zero-mean model used for the simulated data

these simplistic back projections illustrate how demanding is the signal inference from INTEGRAL/SPI data. To study this we perform a number of simulations with several signal models to test the strength of the available IFT filters.

The way simulations are performed in this thesis is described in the following:

1) A signal model for the γ-ray emission in the sky image is considered. This model can be chosen with an arbitrary morphology and intensity. But in our case, the Galactic γ-ray emission at 511 keV is known to be strongly concentrated in the center of the Galaxy, therefore a Gaussian morphology is used.

Fig 4.2 shows this emission model in the sky image, a blub in the Galactic center which extends along longitude to indicate the Galactic disk. To have more structure in the morphology of the simulated signal, a random log-normal field is multiplied to this Gaussian shape. A log-normal field has field values of which the logarithm is distributed as a Gaussian. To adopt the intensity of the simulated signal we multiply the whole field by an appropriate factor. The resulting structured simulated signal is depicted in Fig 4.3. Thus, our simulated signal is given by a Gaussian shape field times a log normal field.

2) Simulated data are then produced according to the data model. Since the background has to be included in this we have

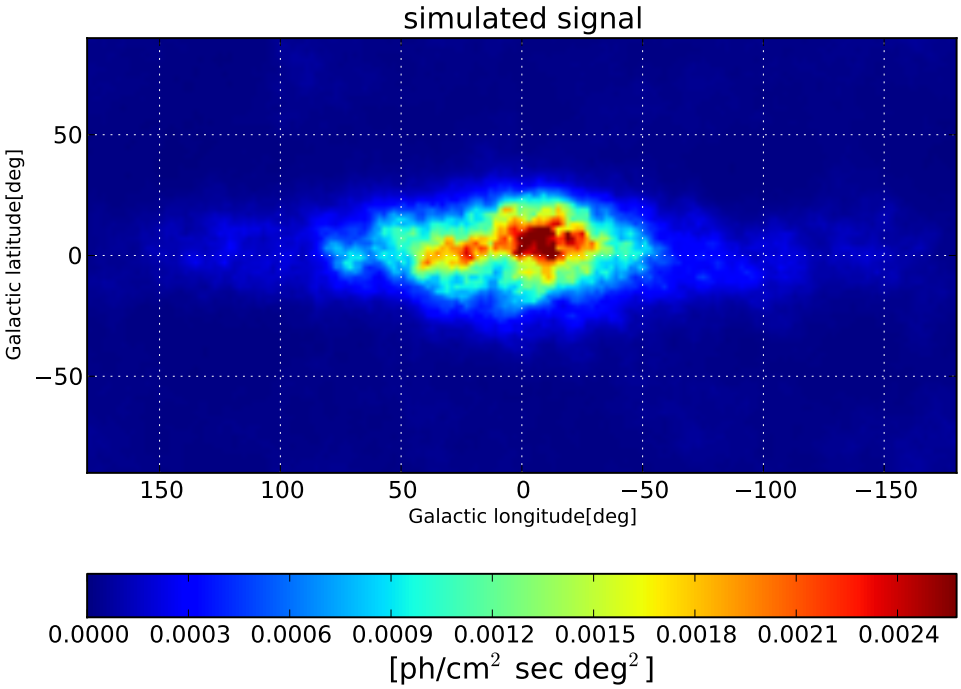

Figure 4.3: Final simulated signal with units of ph/cm^2 sec deg^2

$$d = Rs + b + n. \tag{4.3}$$

To generate such simulated data, the response is first applied to the simulated signal and the signal-response Rs is produced. Afterward a SPI background model by work of Siegert et al. (2016) is added. Signal response plus background are the expected number of counts per data bin. According to the related noise level (Eq. 4.2), Poissonian noise is drawn and added into the data model (Eq. 4.3).

3) Then the considered reconstruction methods, the Wiener Filter (WF) and D^{3}PO, are applied to the simulated data to reconstruct the simulated signal. As it was described in chapter 2, D^{3}PO can also reconstruct the power spectrum of the signal.

4) The reconstructed signal can then be compared to the simulated signal. The residual error is given by

$$r = s - m, \tag{4.4}$$

where s is the original signal (simulated signal) and m is the reconstruction. Considering that the expectation of the signal given the data is the reconstructed map

$$m = \langle s \rangle_{(s|d)}, \tag{4.5}$$

then a straightforward calculation shows that (Enßlin, Frommert & Kitaura, 2009) the variance of the residual error is provided by the propagator itself

$$\langle rr^\dagger \rangle_{(s|d)} = \langle ss^\dagger \rangle_{(s|d)} - \langle s \rangle_{(s|d)} \langle s^\dagger \rangle_{(s|d)} = \langle ss^\dagger \rangle^c_{(s|d)} = D, \tag{4.6}$$

in which D is the propagator operator. Thus, the posterior should be modeled as a Gaussian distribution

$$P(s|d) = \mathcal{G}(s - m, D). \tag{4.7}$$

The standard deviation of this distribution is given by $(\mathrm{diag}D)^{\frac{1}{2}}$ which is used to determine the uncertainty map. This map quantifies the comparison between the reconstructed and simulated signal and it indicates how much and in which sky regions the used filter provides us with an accurate signal inference. Also the residual map is provided to show the regions in which the reconstructed map is different from the simulated signal.

The results of the simulations for both, the WF and D^{3}PO, will be present in the next sections.

4.1 Simulation with Wiener Filter

Here the employed structure of the WF remains similar to what explained before in Sec. 2.3.1, only with a difference which comes from the data model in Eq. 4.3. In this case the likelihood is written as

$$P(d|s) = \mathcal{G}(n, N) = \mathcal{G}(d - Rs - b, N) \tag{4.8}$$

Then Eq. 2.20 becomes

$$P(s|d) \propto P(d|s)\, P(s) = \mathcal{G}(d - Rs - n, N)\, \mathcal{G}(s, S). \tag{4.9}$$

If we calculate the Hamiltonian (Eq. 1.15) and consider just the s-dependent terms (the sign $\widehat{=}$ indicates equality up to irrelevant constants)

$$H(d, s) \widehat{=} - s^\dagger R^\dagger N^{-1} d + \frac{1}{2} s^\dagger R^\dagger N^{-1} R s + b^\dagger N^{-1} R s + \frac{1}{2} s^\dagger S^{-1} s. \tag{4.10}$$

Derivative of the Hamiltonian with respect to s and setting this to zero to identify the minimum yields

$$- R^\dagger N^{-1} d + R^\dagger N^{-1} R s + R^\dagger N^{-1} b + S^{-1} s = 0. \tag{4.11}$$

Solving this for s gives the WF equation

$$m = (S^{-1} + R^\dagger N^{-1} R)^{-1} (R^\dagger N^{-1} (d - b)) = Dj \tag{4.12}$$

As can be seen from Eq. 4.12, the information source defined in the WF context is informed about the background presence, $j = R^\dagger N^{-1}(d - b)$.

Based on previous studies of SPI data on the Galactic γ-ray emission at the 511 keV, the whole sky intensity amounts to $\sim 10^{-3}$ ph/cm^2 sec (Siegert et al., 2016). Although one

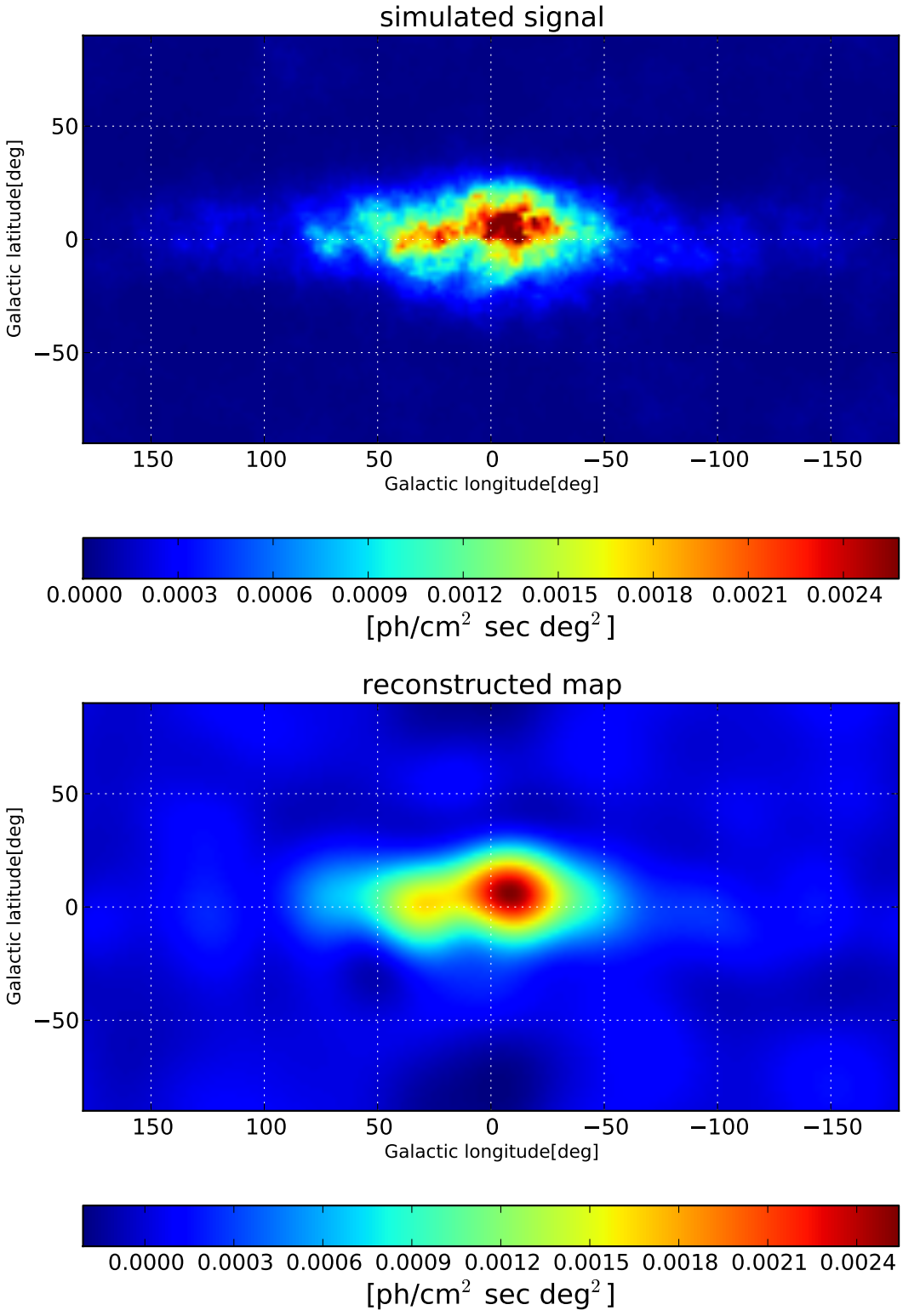

Figure 4.4: Results of the mock reconstruction using the WF.

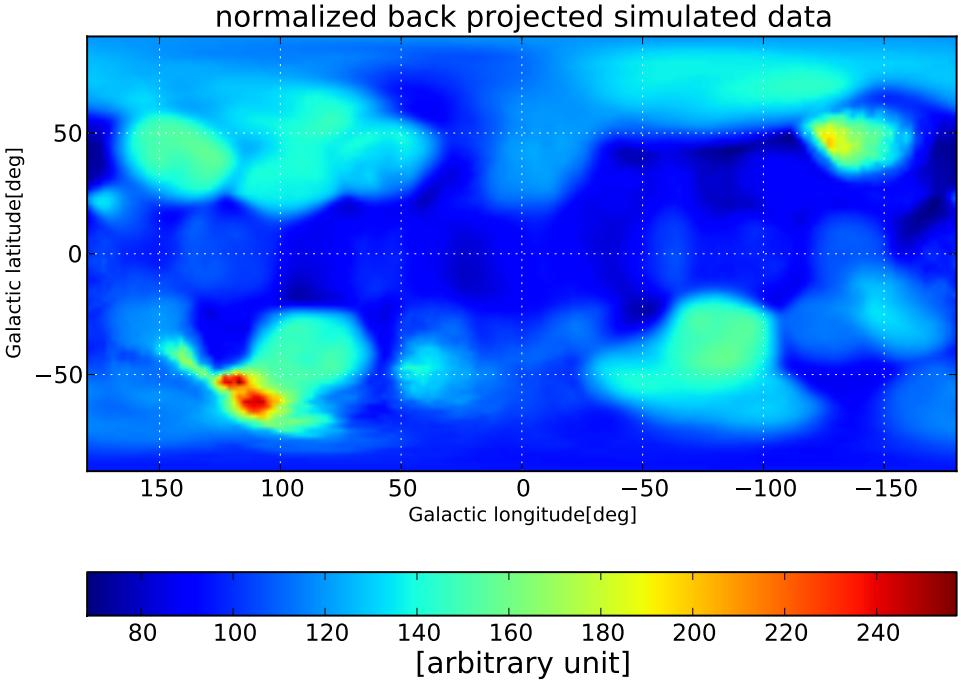

Figure 4.5: Normalized back projected simulated data using the WF.

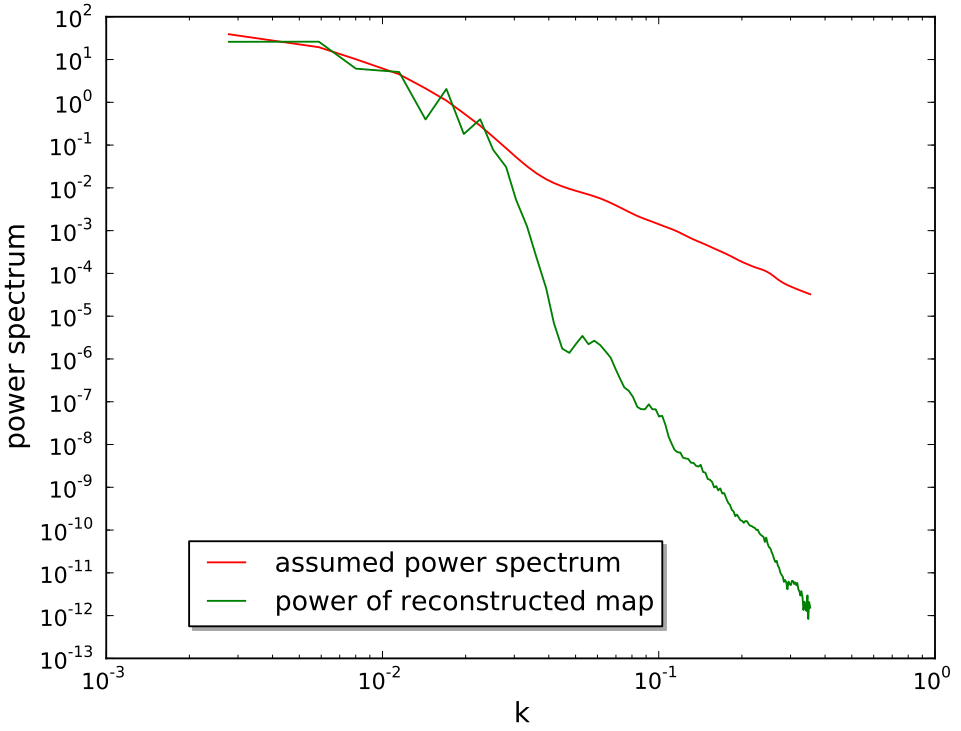

Figure 4.6: Corresponding power spectrum of the mock reconstruction using the WF.

of our goals is an independent determination of the total flux of Galactic γ-ray emission at 511 keV and we should therefore avoid any default value in this determination, the referred flux is used to set up our simulated signal in order to have a realistic mock scenario.

We assume to know the correct signal power spectrum for the WF based data analysis and for this we choose the power of the simulated signal. The WF reconstructed map of the simulation as given by Eq. 4.12, is presented in the bottom panel of Fig. 4.4. Also the simulated data which are projected back onto the sky and normalized to the exposure map is presented in Fig. 4.5.

In bottom panel of Fig. 4.4, the extension of the emission toward the positive Galactic longitude is reconstructed well, however small-scales features of the simulated signal are not inferred from the simulated data. The smooth appearance of the reconstructed map in comparison to the original signal is a consequence of the fact that the WF suppresses structures with low signal-response to noise ratio, due to the high noise in small scales. As the assumed power spectrum of the WF was the actual signal power spectrum (red line of the bottom panel in Fig.4.6), the WF reconstruction is actually a bit too optimistic (green line of the bottom panel in Fig. 4.6 shows the power of the reconstructed map).

The total flux of the simulated signal is $\sim 2.8 \times 10^{-3}$ ph/cm^2 sec and is $\sim 2.7 \times 10^{-3}$ ph/cm^2 sec for the reconstructed map, which is an accurate match. The match can also be shown by the central flux which is measured in a frame from l=100° to -100° and b=-50° to 50°, since the frame resides in the region where the exposure of INTEGRAL to sky is strong and in these areas more accuracy is expected compared to higher latitudes. The central flux for the simulated signal and reconstructed map is 2.2 $\times$ 10^{-3} and 2.27 $\times$ 10^{-3} ph/cm^2 sec, respectively. Also, to investigate the accuracy of the derived map the WF uncertainty and residual map are shown in Fig. 4.7. The residual map is the difference between reconstructed map and the original one. We expect the residual map to scatter around zero. The strong residual structures in areas of the sky signal is due to the fact that the WF is not resolving the small scales of the simulated signal in the reconstruction. However with residuals being on the order of 10^{-4} gives us the impression that the performance of the algorithm in recovering the signal from the simulated data is satisfying. As well, the uncertainty map is higher in those areas in which the instrument observed less in comparison to the Galactic center.

4.2 Simulation with D^{3}PO

Now we investigate the performance of the D^{3}PO algorithm using the same simulated signal as it was employed in simulation with WF. We expect the D^{3}PO to perform better than the WF, as it is based on a more realistic signal model, which is assuming a log-normal signal statistic, and the correct Poisson noise statistics. Also in D^{3}PO algorithm the Hamiltonian (Eq. 2.44) has a tiny change, only in the definition of λ in Eq. 2.45 and regarding that our goal is only the reconstruction of the diffuse emission

$$\lambda = R(e^s) + b, \tag{4.13}$$

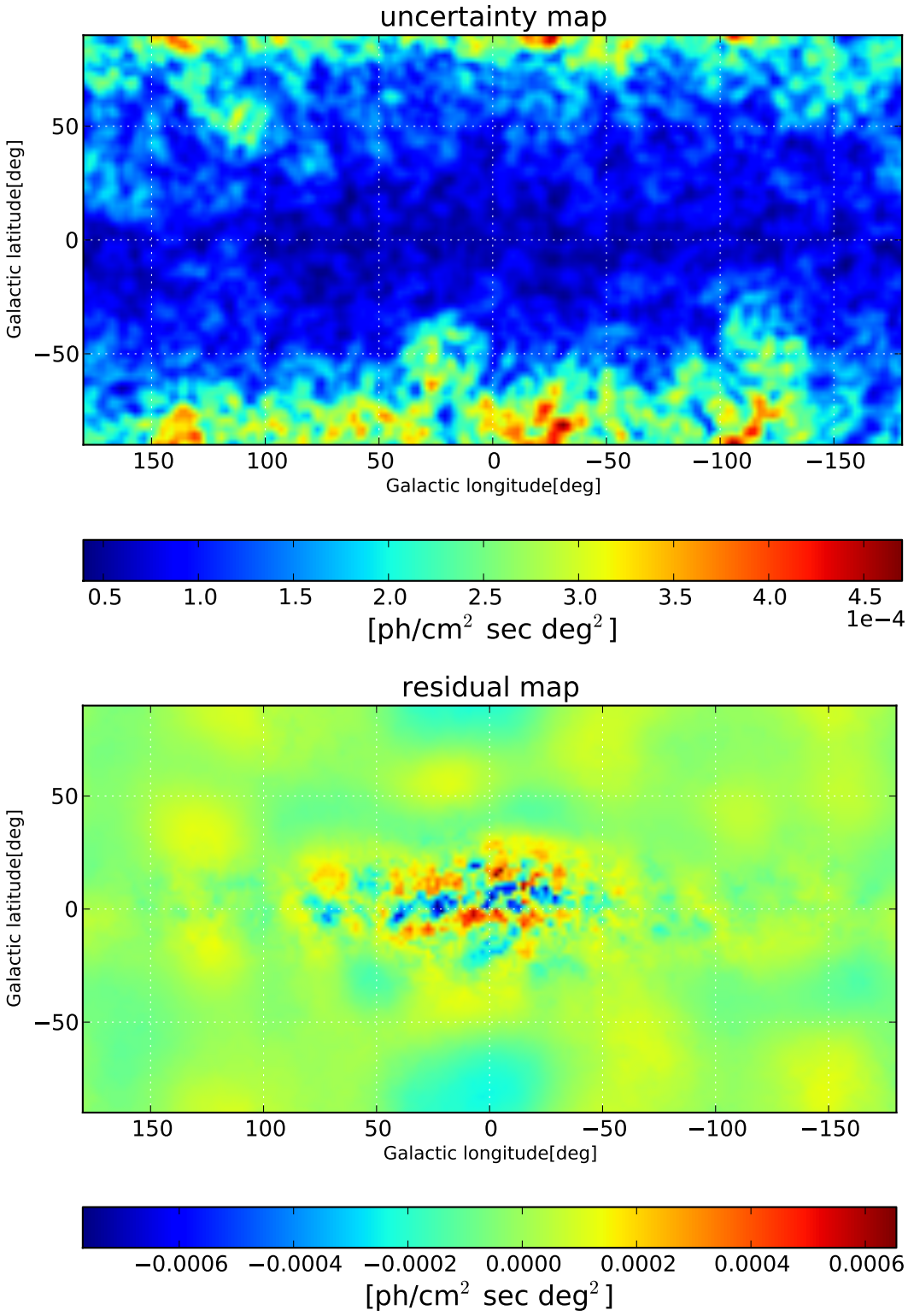

Figure 4.7: Uncertainty and residual maps derived from mock reconstruction with WF.

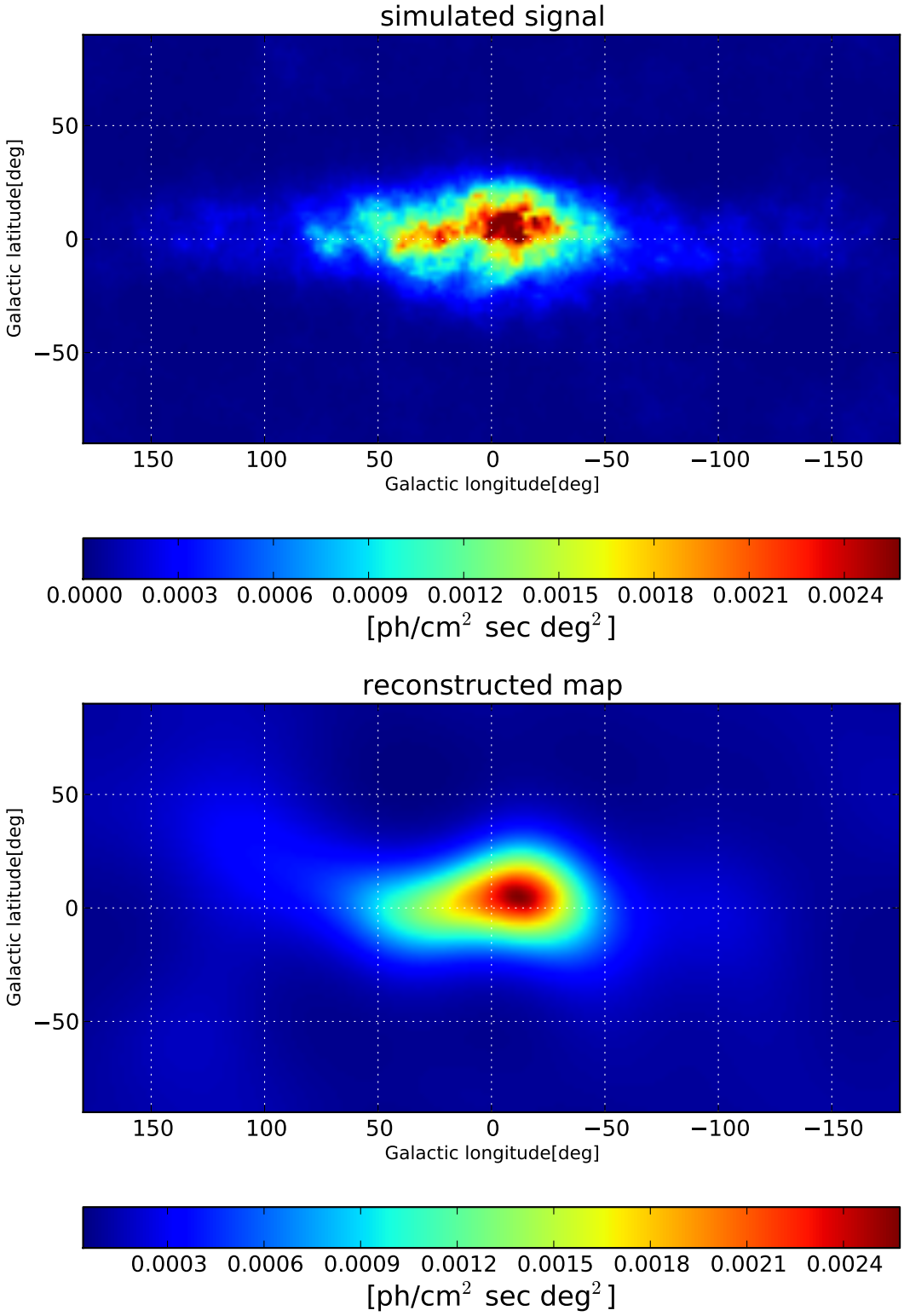

Figure 4.8: Results of the mock reconstruction with realistic signal-to-noise ratio using the D^{3}PO.

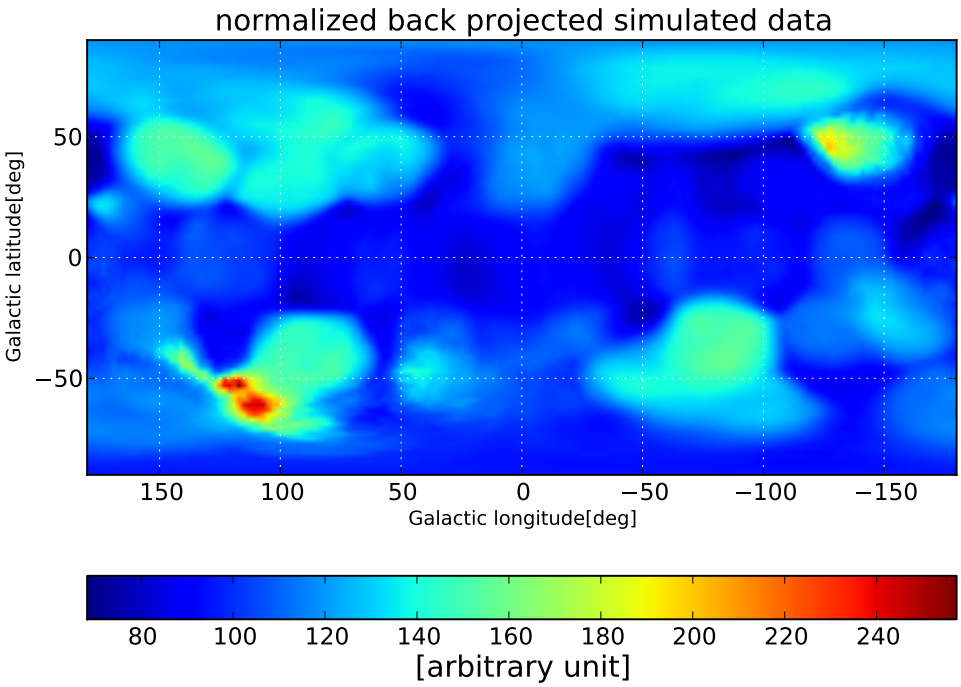

Figure 4.9: Back projected simulated data with realistic signal-to-noise ratio using the D^{3}PO.

in which the SPI background is added to the expected data, λ. Therefore, the data model is written as

$$d = R(e^s) + b + n, \tag{4.14}$$

As in the WF simulation, we provide D^{3}PO with the correct signal power spectrum. The map reconstructed by D^{3}PO is shown in Fig. 4.8 which is in good agreement with employed simulated signal in terms of morphology (on the large scales) and intensity. The corresponding uncertainty and residual maps are also presented in Fig. 4.11. The total flux of the simulated signal and reconstructed map is $\sim 2.7 \times 10^{-3}$ and 4.4×10^{-3} ph/cm^2 sec, respectively. The better match is in the case when we compare the central flux of the simulated signal and reconstructed map, which is 2.2×10^{-3} and 2.8×10^{-3} ph/cm^2 sec, respectively. The reason for getting a better match between the total flux of the original signal and that of the reconstructed map in case of WF compared to D^{3}PO reconstruction is due to a Guassian prior, used in WF, allows for less dynamical range compared to a log-normal one, which is used as prior in D^{3}PO. These assumptions cause the total flux in WF to be lower than the one in D^{3}PO. However, considering the central flux for the comparison is more accurate, since the central flux resides in the Galactic bulge and somehow in the Galactic disk, where the sky is observed by the instrument better than the other regions in the sky.

Enßlin (2014) showed that the here used method to estimate the power spectrum (Eq.

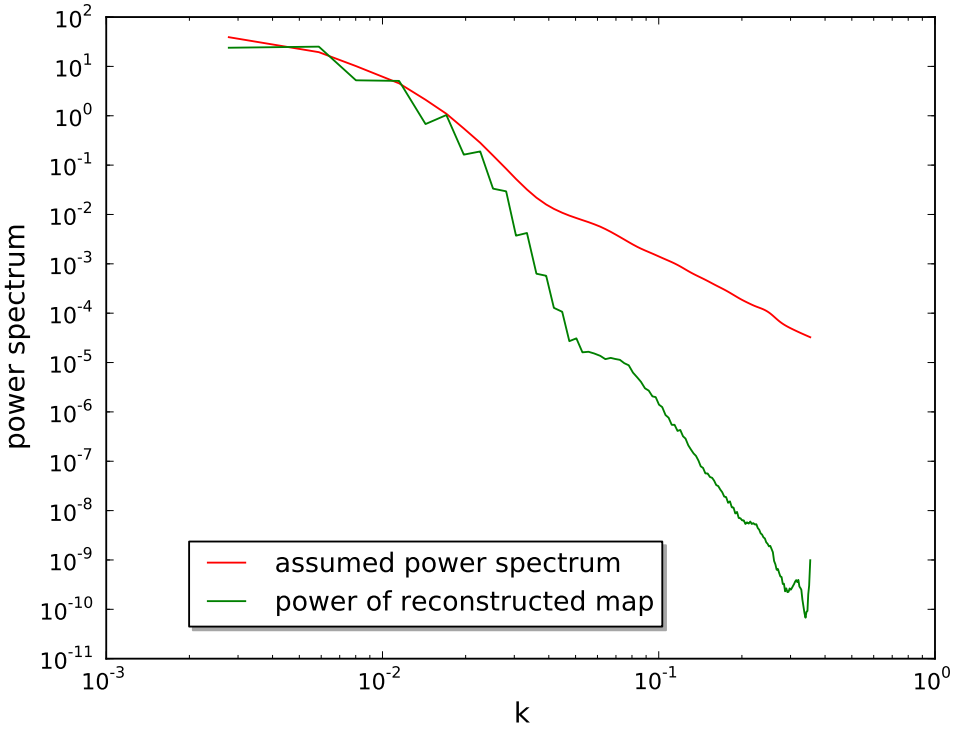

Figure 4.10: Corresponding power spectrum of the mock reconstruction with realistic signal-to-noise ratio using the D^{3}PO.

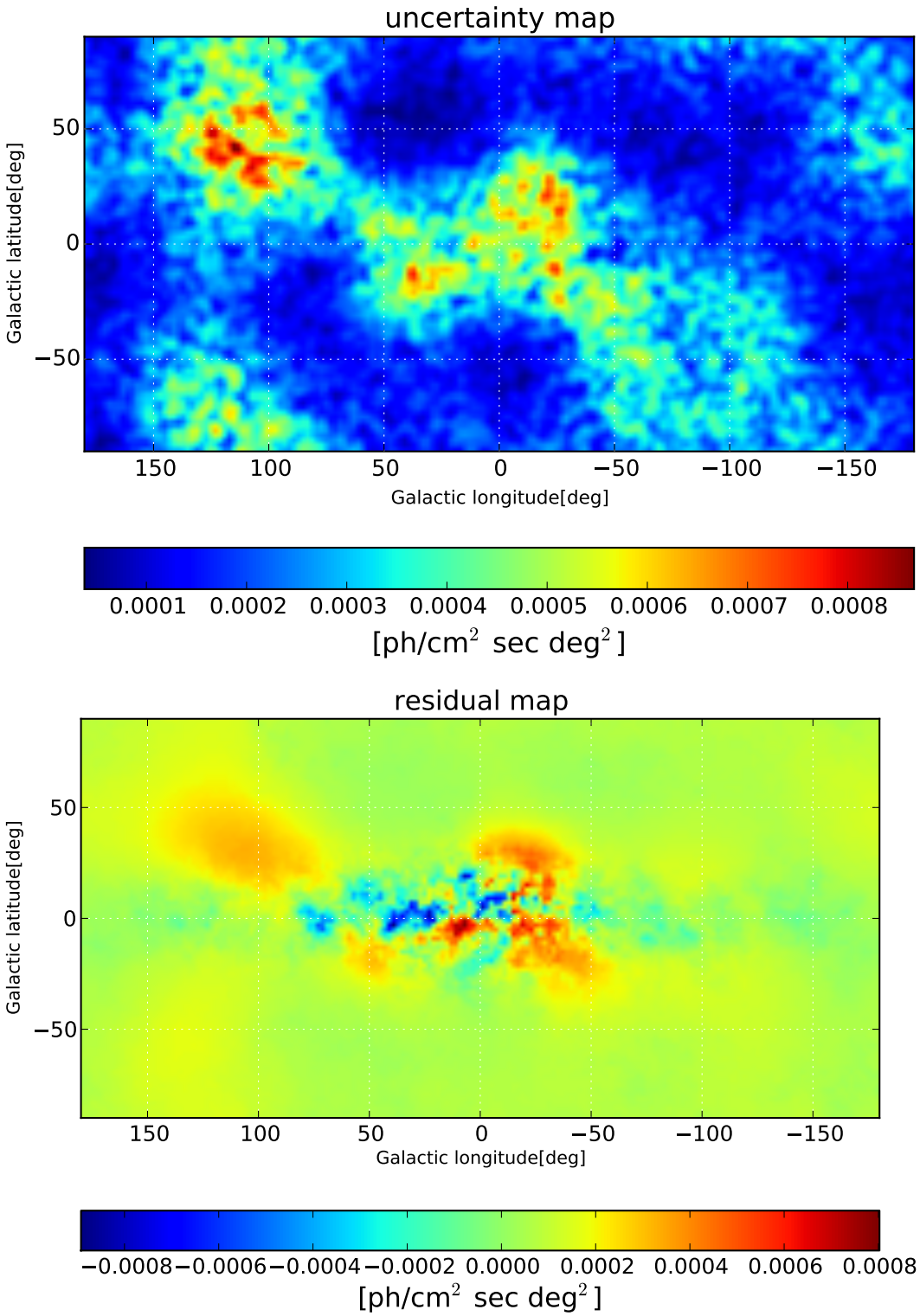

Figure 4.11: Uncertainty and residual map for the simulation with realistic signal-to-noise ratio using the D^{3}PO.

2.53) has a perception threshold. For many Fourier modes the ratio of the signal-response to noise level is too low for D^{3}PO to recover the signal power spectrum. Therefore, despite of D^{3}PO's ability to reconstruct power spectra, here we assume that the power of the signal is known, to be for which power of the simulated signal. However, if the power of the signal would be stronger, there is the possibility to recover the power of the signal as well. Figure 4.12 shows a simulation in which the sky brightness was increased by a factor of more than 100. This increase in the intensity of the simulated signal leaves a well observable signature in the back projection of the simulated data onto the sky in the Galactic center, which is depicted in Fig. 4.13. Although the Poisson noise level thereby largely increases (by a factor more than 10), the improved signal-response-to-noise ratio permits D^{3}PO not only to reconstruct accurately a detailed sky map, the thereby used power spectrum was also determined by the data. This reconstruction in the power spectrum of the reconstructed map is observable in Fig. 4.14, where power of the simulated signal and reconstructed map are very similar on many different scales. As it could be seen from the reconstructed map shown in Fig. 4.12, the small scale structures of this stronger simulated signal have good chances to be inferred from the simulated data. The total flux of the simulated signal and reconstructed map is 4.1×10^{-1} and 6.1×10^{-1} ph/cm^2 sec, respectively. Also the central flux amounts to 3.4×10^{-1} ph/cm^2 sec and 3.6×10^{-1} ph/cm^2 sec, respectively for the simulated signal and reconstructed map, which shows a good match.

The uncertainty map was calculated as the pixel-wise standard deviation as approximated by the inverse Hessian of the log-normal Poisson Hamiltonian, which is used in D^{3}PO. This uncertainty and also the residual map are depicted in Fig. 4.15.

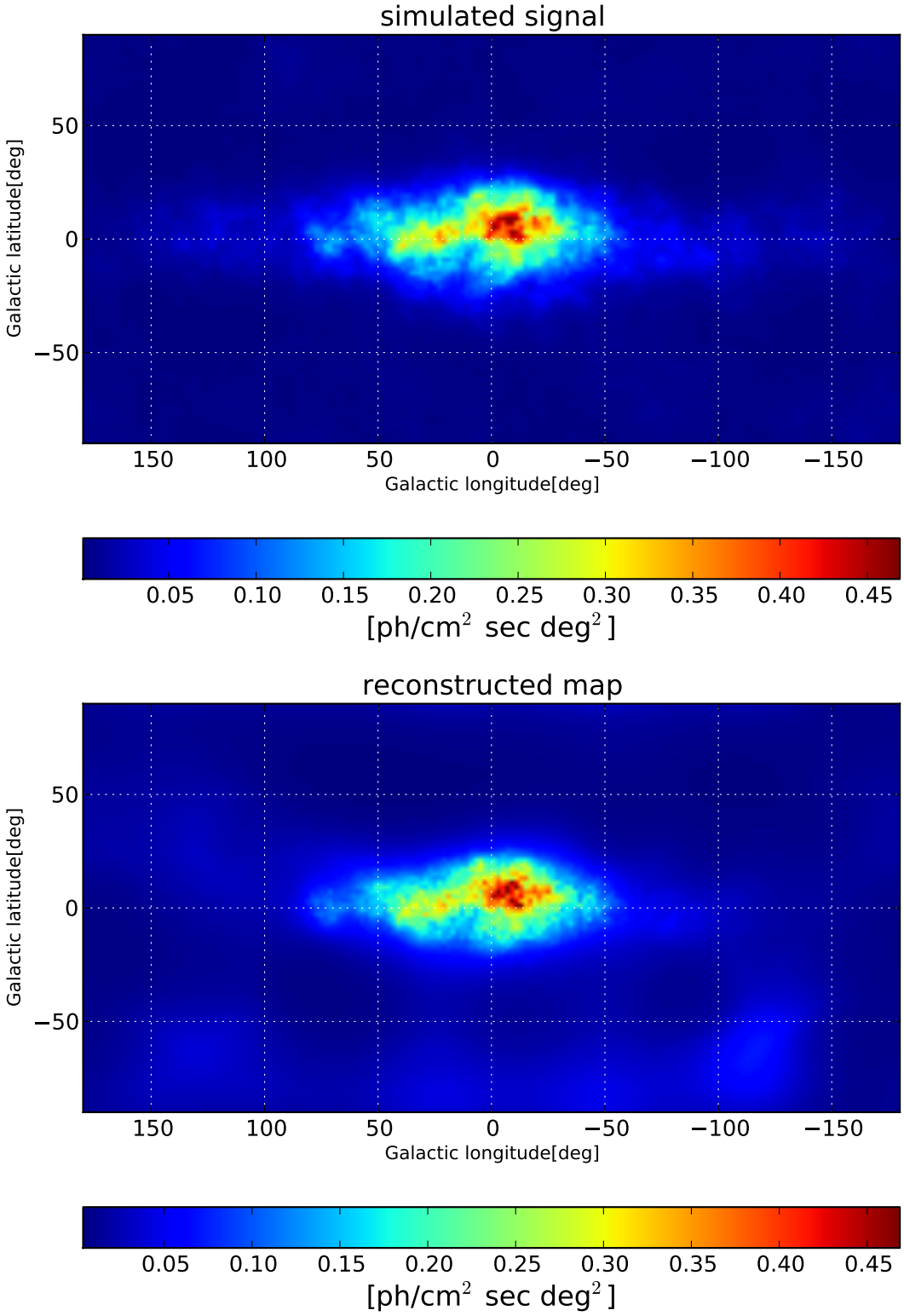

Figure 4.12: Results of the mock reconstruction with 100 times enhanced signal-to-noise ratio using the D^{3}PO.

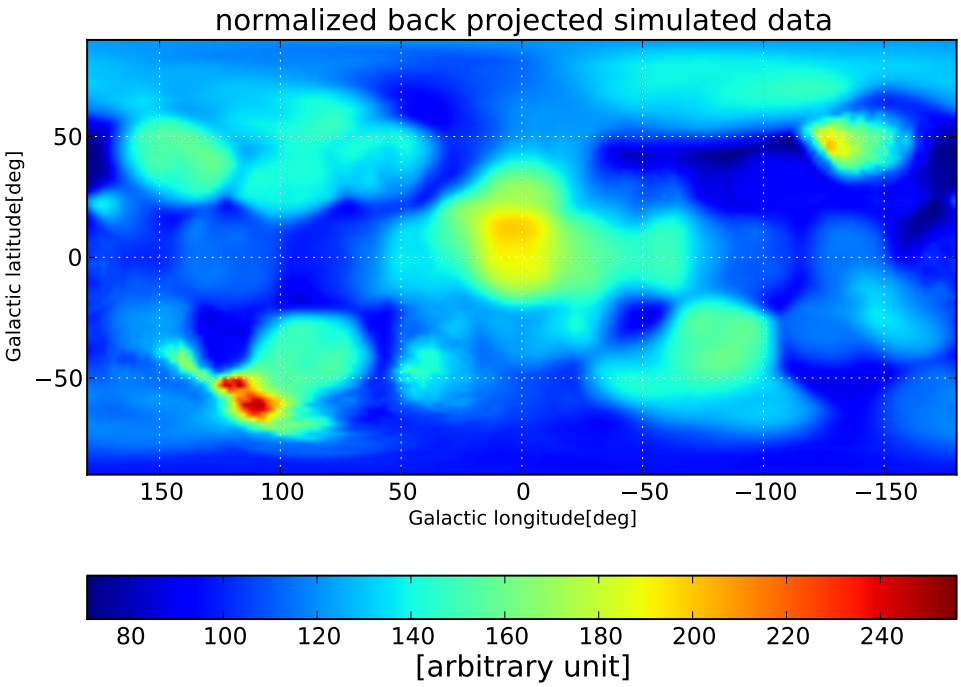

Figure 4.13: Back projected simulated data of mock reconstruction with 100 times enhanced signal-to-noise ratio using the D^{3}PO.

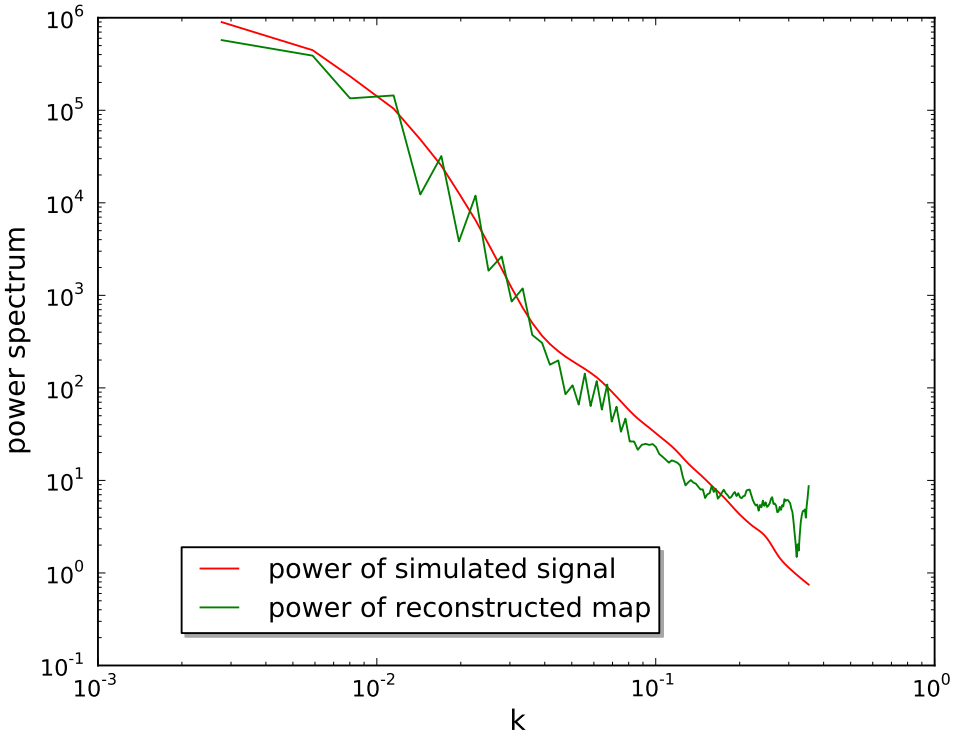

Figure 4.14: Corresponding power spectrum of the mock reconstruction with 100 times enhanced signal-to-noise ratio using the D^{3}PO.

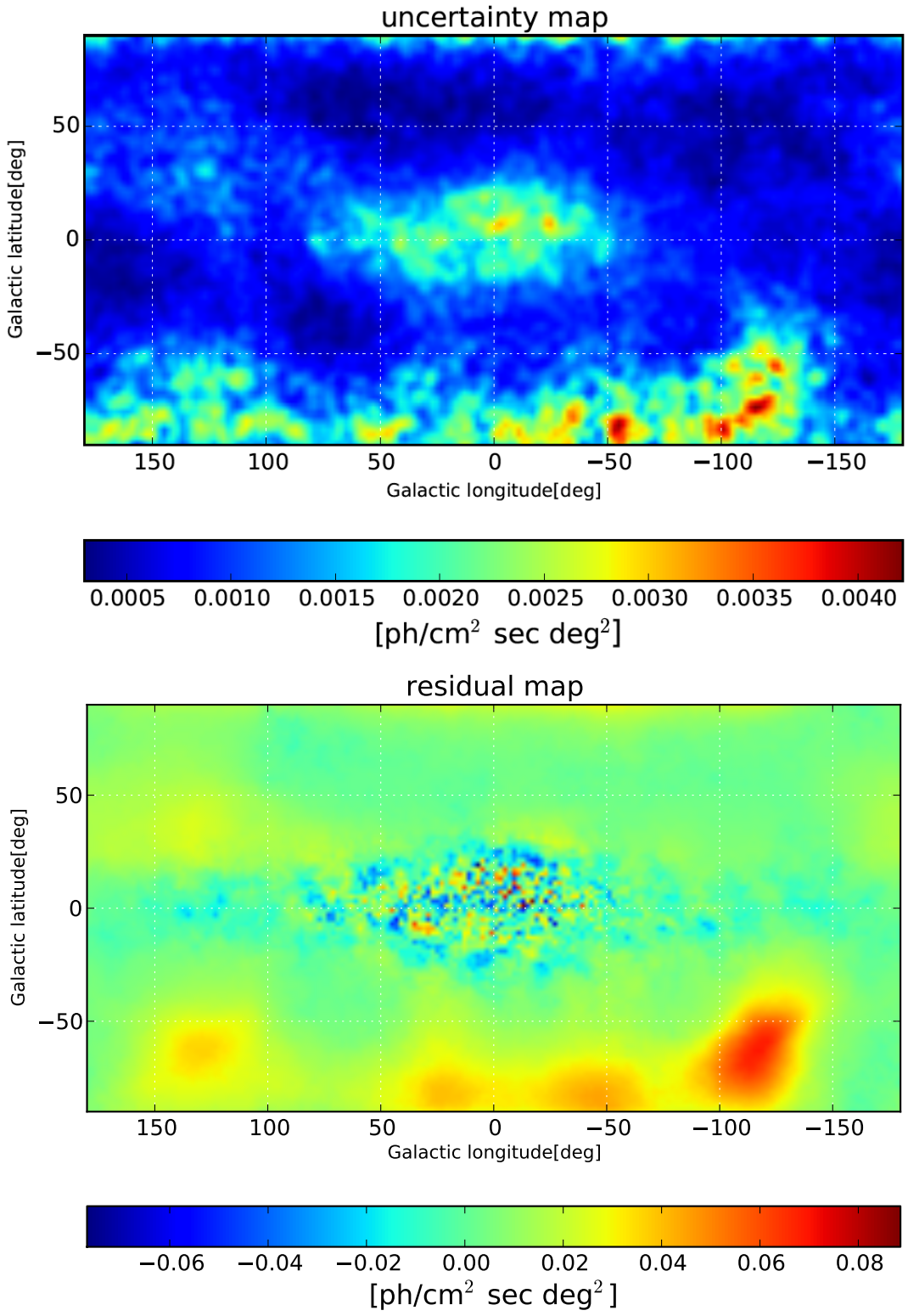

Figure 4.15: Uncertainty and residual maps for the mock reconstruction with 100 times enhanced signal-to-noise ratio using the D^{3}PO.

Chapter 5

Application of D^{3}PO to observational data

The investigation of the spatial distribution of the 511 keV γ-ray emission originating from positron annihilation permits us to determine the origin of the emission, if we assume the annihilation site to be close to the positron production region. This would be the case if positrons do not travel too much across the Galaxy after production. In the first section of this chapter, we present the first result of the reconstruction of the positron annihilation emission on the sky through an application of D^{3}PO to SPI data at 511 keV. In the second section the sky emissivity of the positronium line is reconstructed by using information from an energy range covering 500-520 keV. In both sky mappings, the background model of Siegert et al. (2016) is used. In contrast to this, in Sec. 5.3, the background level is reconstructed as a function of the exposure time simultaneously with the sky signal. The reconstruction of the background requires the description of the associated response that translates background flux variations in enhanced count rates as a function of detector, energy, and pointing. As a consequence, the resulting sky map of the γ-ray emission at the 511 keV will becomes largely independent of the background model by Siegert et al. (2016) used in the previous reconstructions, which had been constructed by using a parametric sky model including a Galactic center 511 keV source. Only by the usage of an independent background model, the existence of the Galactic center signal can be confirmed.

For this analysis we again use observations from the SPI spectrometer, which were accumulated over eleven years INTEGRAL mission, from orbit number 21 to 1279, with gaps due to enhanced solar activity, calibration, or annealing periods (Siegert et al., 2016).

5.1 Sky map of the positron annihilation at 511 keV

As it was discussed in the previous chapter, for those Fourier scales for which the signal-response-to-noise ratio drops below a certain value, the reconstruction of the signal power spectrum becomes impossible. This happens in our case for most of the Fourier modes since the contribution of the signal to the SPI counts is rather small compared to that of

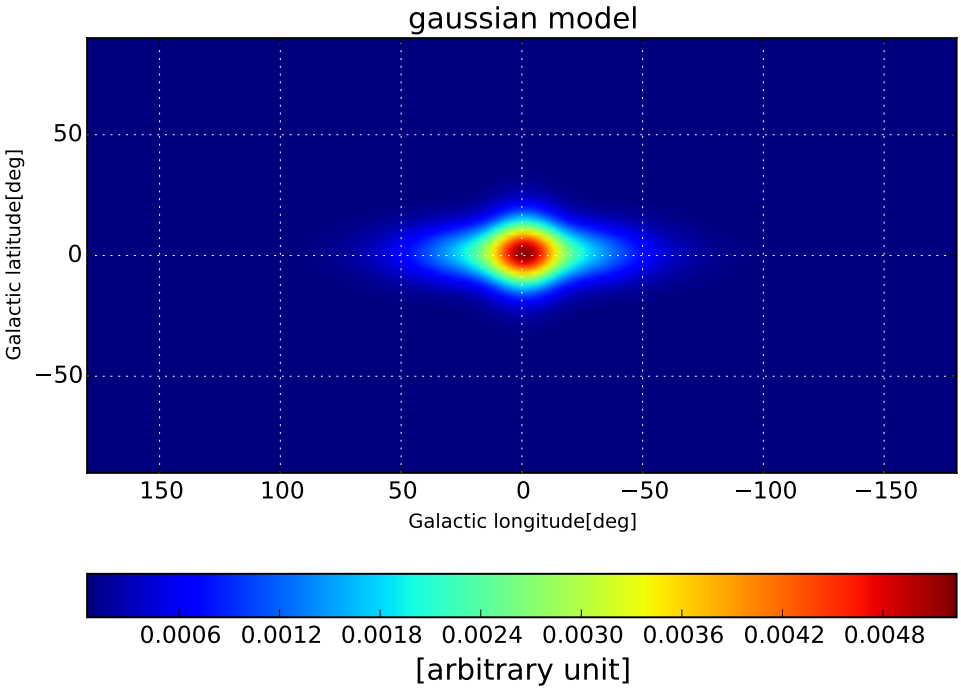

Figure 5.1: Gaussian model for the 511 keV sky emission.

the background. We therefore have to assume the power spectrum. We assume it follows a simple power-law,

$$P_s(k) \propto (1 + \frac{k}{k_0})^{-2}. \tag{5.1}$$

Having a slope of k^{-2} means that gradients are disconverged, and that we basically adapt a Gaussian Markov random field model as it is frequently adapted for non-parametric reconstructions. To specify the amplitude of this power law, our simulation model is used, which is depicted in Fig. 5.1. The blub-like distribution shown there is intensive in the Galactic center and it is extended along the Galactic longitude. Also the total flux of this model is adapted to that of the reported positron annihilation line emission of $\sim 2.7 \times 10^{-3}$ ph/cm^2 sec (Siegert et al., 2016). In Fig. 5.2 the green line shows the power spectrum of the simulation model.

Our first application of D^{3}PO to the SPI data focuses only on the energy of 511 keV, which is at the peak of the Galactic positron annihilation emission. D^{3}PO was modified to handle the SPI response and the presence of an instrumental background (Eq. 4.14), as explained in chapter 4. The result is illustrated in the top panel of Fig. 5.3. This sky map shows spatial distribution of the γ-ray emission at 511 keV in the Galaxy with total flux of $\sim 4.6 \times 10^{-3}$ ph/cm^2 sec within energy range of 510.5-511 keV. This flux is consistent with a total flux of the sky map of the 511 keV emission derived by analyzing 5 years of data by Weidenspointner et al. (2008c). However the central flux of the map from l=30° to -30°

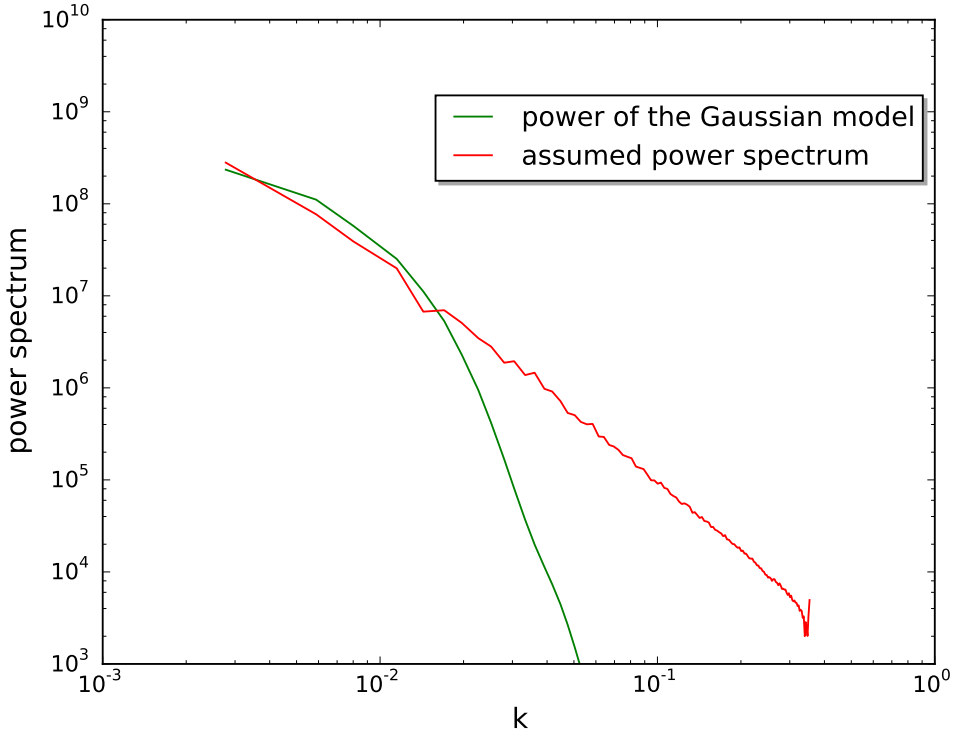

Figure 5.2: Power spectrum of our simulation model (in green) and the power law spectrum adapted for the reconstruction so that it has the same monopole as our model has (in red).

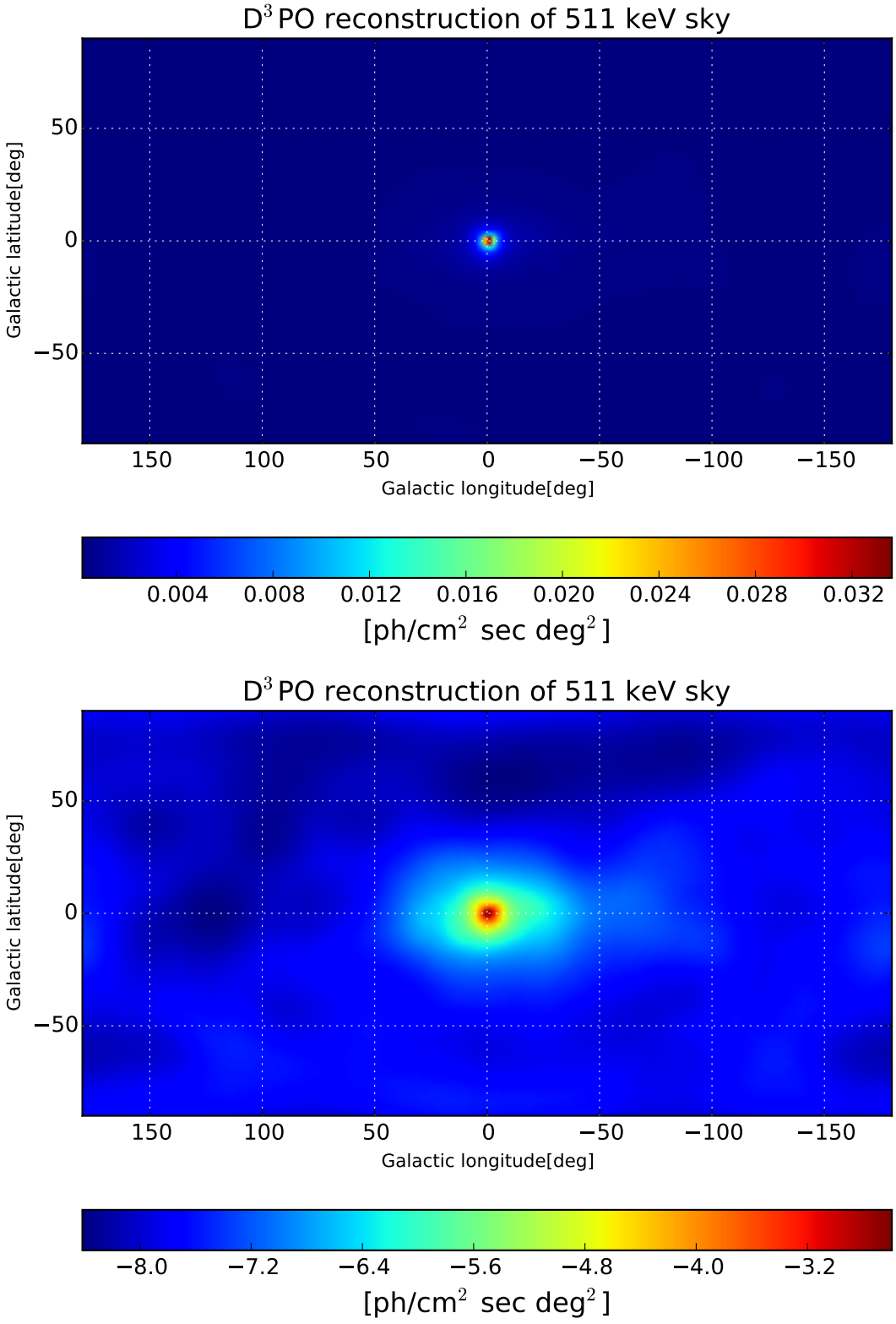

Figure 5.3: Reconstructed sky map in the energy range of 510.5-511 keV using the D^{3}PO and the background model by Siegert et al. (2016). The top panel shows the original map in linear scale, while the bottom one is plotted in logarithmic scales ($\log_{10}$).

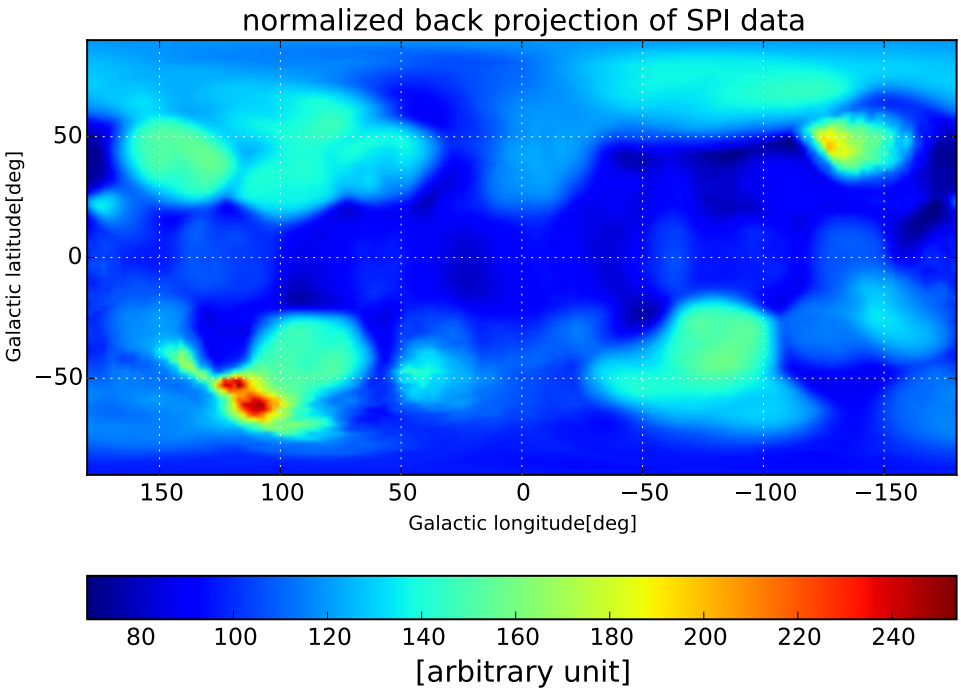

Figure 5.4: Back projected of SPI data in energy width of 510.5-511 keV.

and b=-20° to 20° shows the amount of 1.5×10^{-3} ph/cm^2 sec, exactly consistent with the result by Siegert (2016, priv. comm). Visual inspection of the map shows that the emission in the Galactic center is strong and intense, and that it extends only faintly for a few degree from the center. However the emission plotted on logarithmic scale (bottom panel of Fig. 5.3) shows the faint contribution of the Galactic disk to the positron annihilation emission. The uncertainty map of our reconstruction is depicted in Fig. 5.6. As can be seen from it, the accuracy of the map in the Galactic center, where the main emission structure resides, is of the order of 10^{-3} ph/cm^2 sec deg^2, which is acceptable low compared to the field value of 0.032 ph/cm^2 sec deg^2 in the sky map (top panel of Fig. 5.3). Also the significance of the Galactic center and disk seen in the bottom panel of Fig. 5.6 are $\sim 12\sigma$ and 4σ, respectively, which shows a bright emission in the center of the Galaxy. However the 4σ is not enough for the Galactic disk detection. At this moment, we assumed that we know the background level and power spectrum of the signal in the uncertainty calculation. But in reality we are not sure about them. Thus the real uncertainty should be larger than that is now calculated, and consequently, the significance for the Galactic disk would be estimated less than it is shown in the bottom panel of Fig. 5.6.

In order to illustrate how strongly background dominated the data is, the normalized back projected SPI data onto the sky is shown in Fig. 5.4. In principle the sky emission should show up in this image, if the background variations would not imprint much stronger structures onto it.

The corresponding power spectrum is also shown in Fig. 5.5 .The green line of the

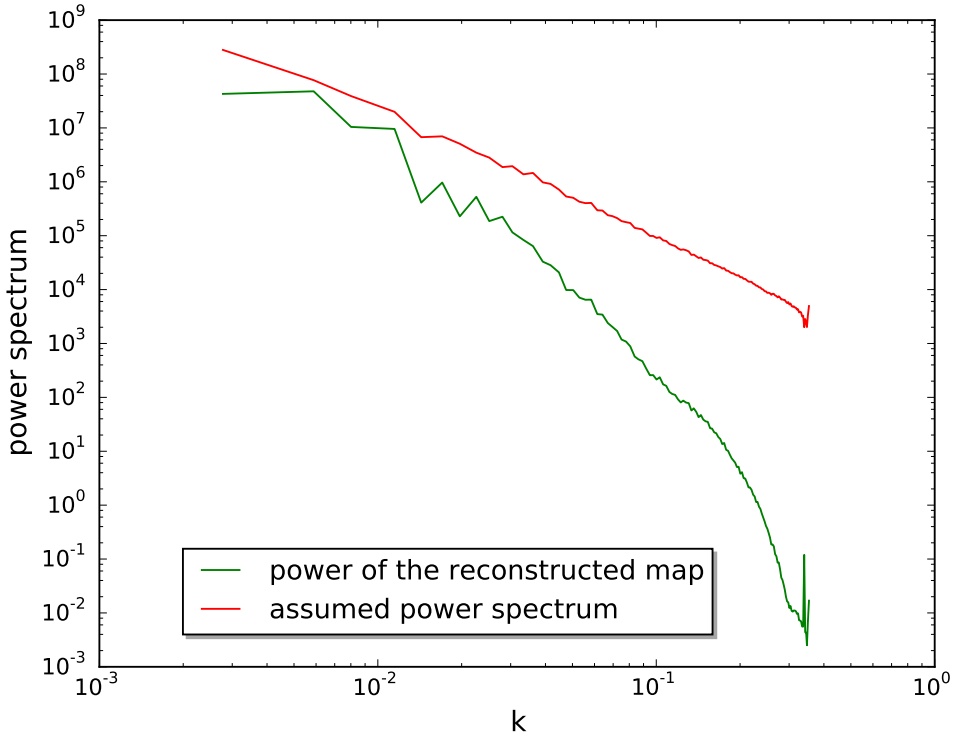

Figure 5.5: Corresponding power spectrum of the reconstructed map from the data in the energy range of 510.5-511 keV (Fig. 5.3) in which the red line is the power that was assumed by D^{3}PO and the green one shows the power of the reconstructed map.

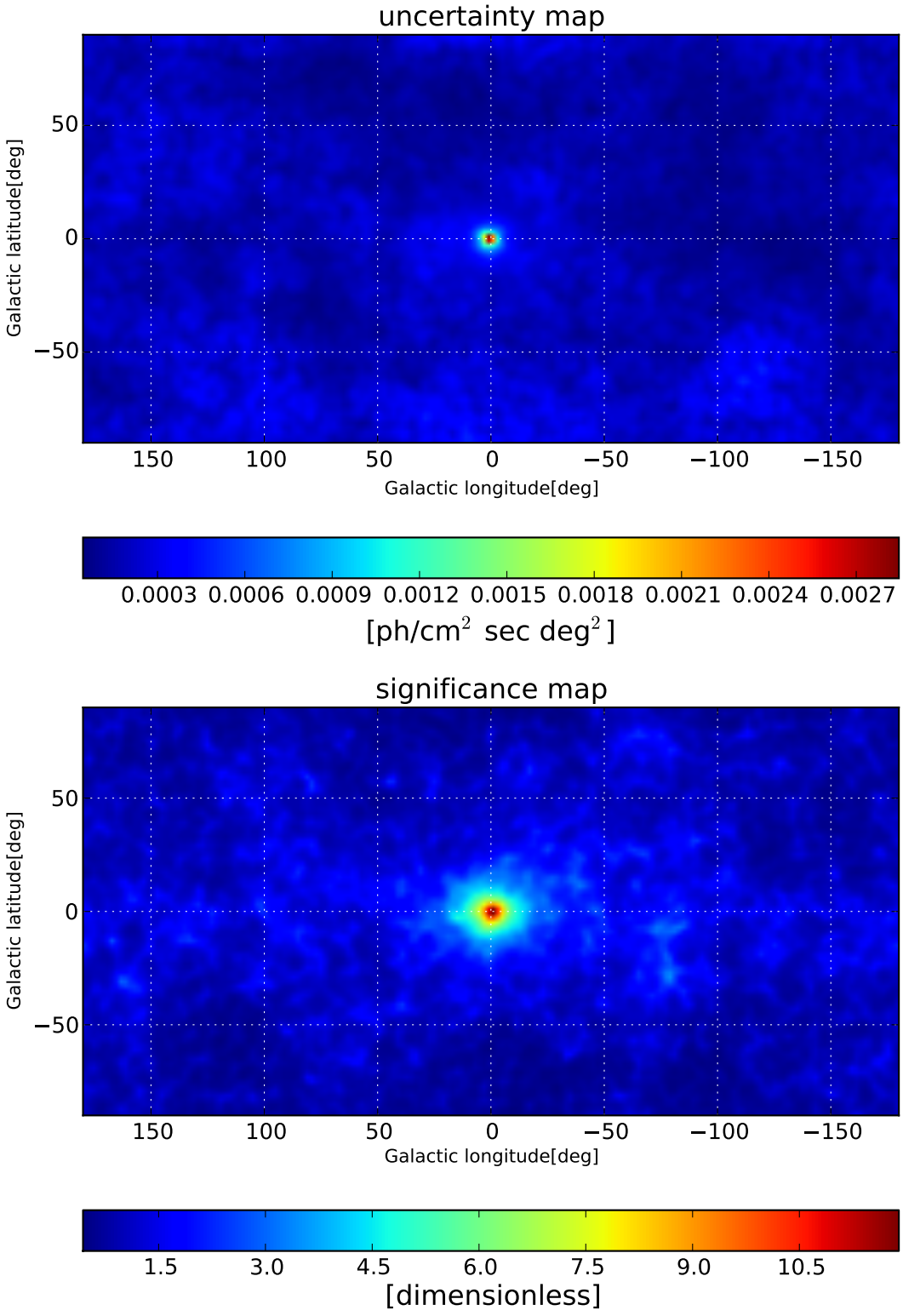

Figure 5.6: Uncertainty and significance maps of the reconstruction using D^{3}PO and background model by Siegert et al. (2016) in the energy range of 510.5-511 keV.

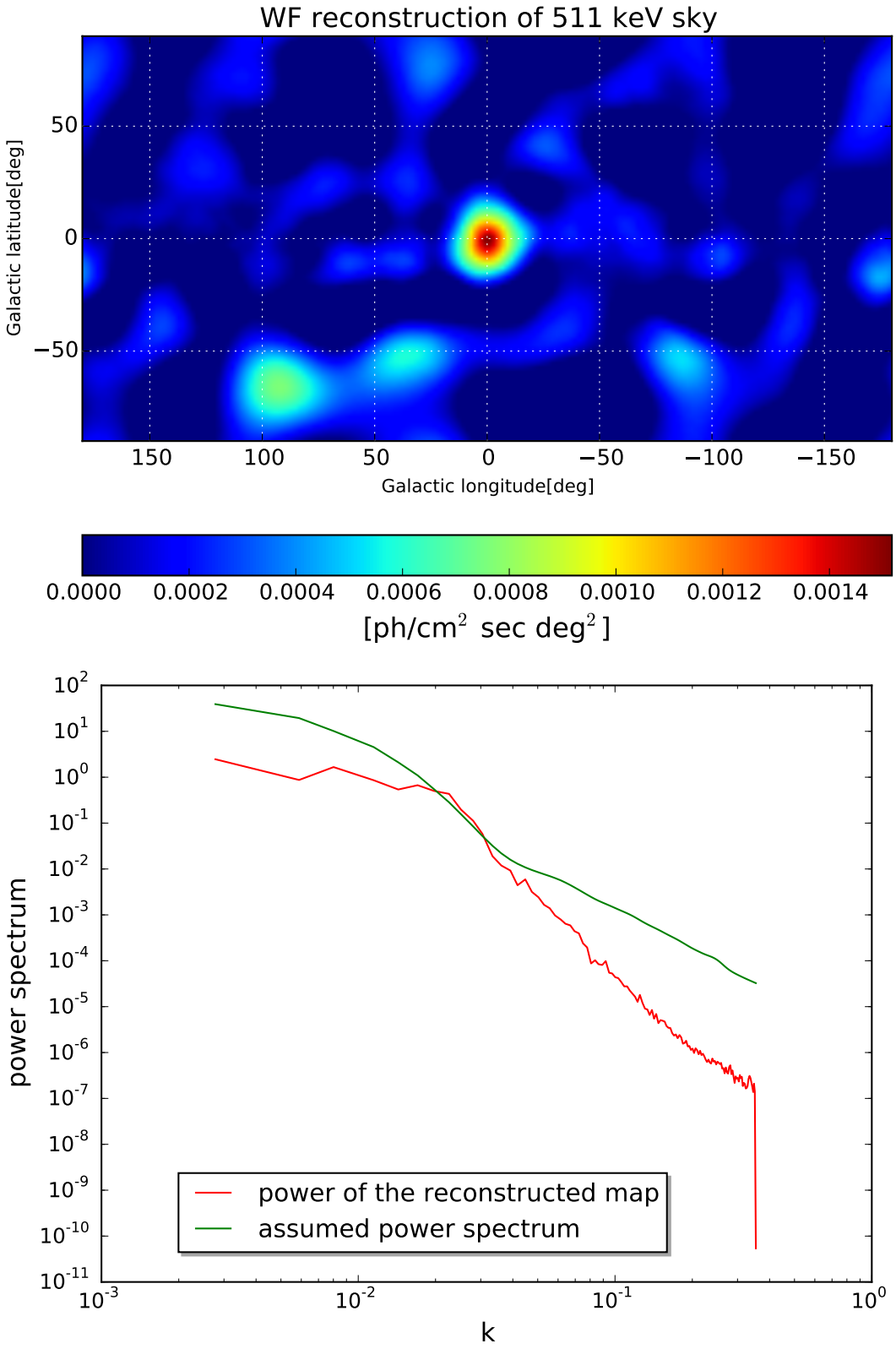

Figure 5.7: Reconstructed sky map of positron annihilation in the energy range of 510.5-511 keV using the WF (top panel), and corresponding power spectrum (bottom panel).

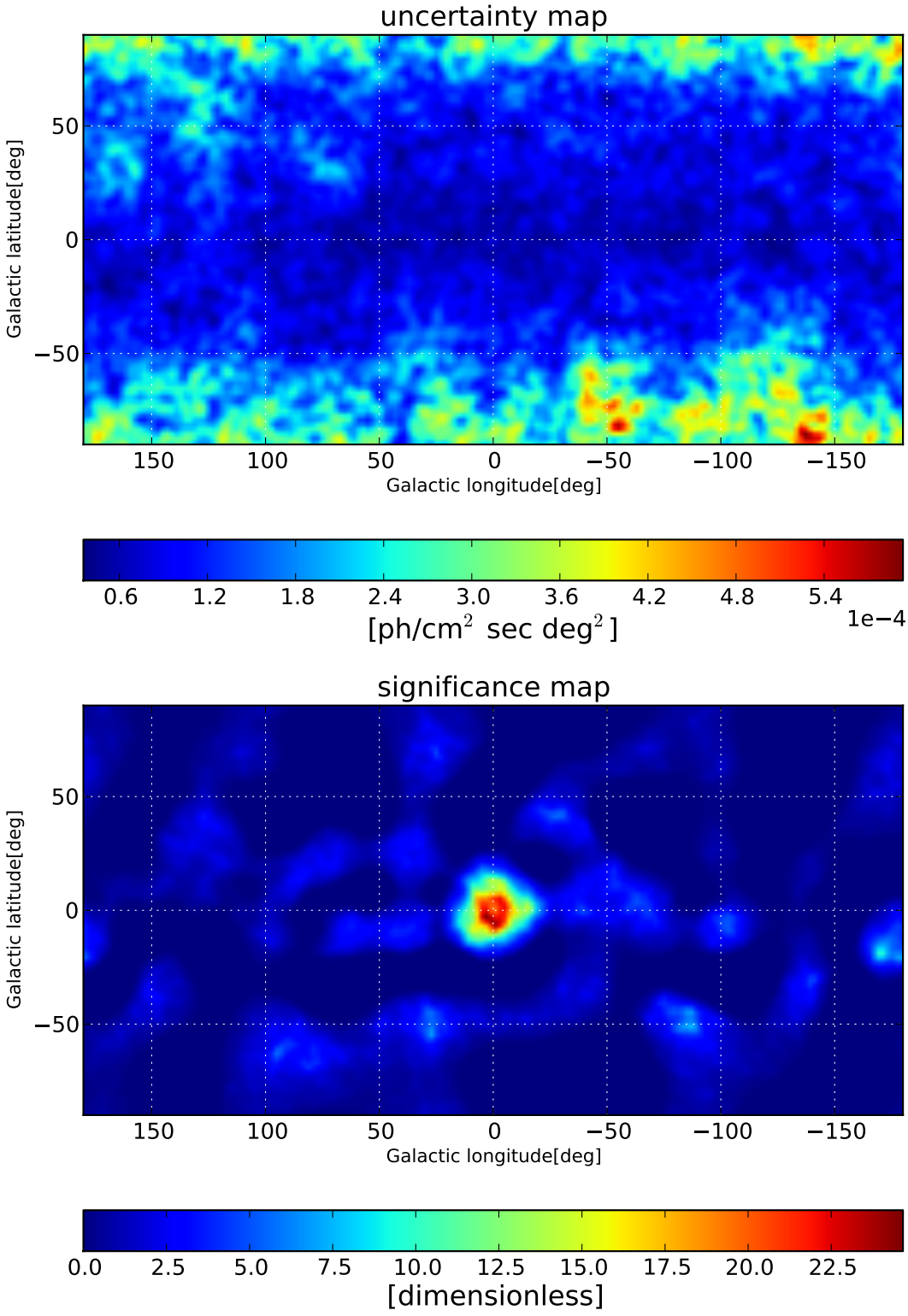

Figure 5.8: Uncertainty and significance map of the reconstruction using WF and the background model by Siegert et al. (2016) in the energy range of 510.5-511 keV.

bottom panel is the power of the reconstructed map, while the red line is the employed power in the algorithm as it was discussed earlier (Fig. 5.2).

The D^{3}PO reconstruction can be compared to the result of the WF algorithm on the SPI data. The all-sky map of γ-ray emission at 511 keV is in the top panel of Fig. 5.7 and like the D^{3}PO map, this map shows a strong concentration of the emission in the Galactic center, however it is less than the D^{3}PO reconstruction shows. The central flux of the map measured in a frame from l=30° to -30° and b = -20° to 20° amounts to 3.92×10^{-4} ph/cm^2 sec, which is 3.8 times less intense than the central flux of the map reconstructed by D^{3}PO. In addition, the total flux of the WF map amounts to 1.9×10^{-3} ph/cm^2 sec in energy range of 510.5-511 keV, which is 2.4 times less than the total flux of the sky map obtained by D^{3}PO. Two bright emission regions at $l = 90°$, $b = -25°$and $l = 35°$, $b = -40°$ are probably due to SPI background model errors. The uncertainty map showing the accuracy of the obtained map is represented in Fig. 5.8. The significance map in the bottom panel of Fig. 5.8 shows the significance of the emission in the Galactic center is up to $\sim 22\sigma$.

5.2 Sky map of the positron annihilation emission from data in the energy range of 500-520 keV

Up to now, we concentrated on the counts in each detector and pointing which were recorded exactly the energy of 511 keV. However, the 511 keV photons are recorded in a wider range of energies due to imperfection of the instrument. This is described by a spectral response function that specifies with which probability a photon of 511 keV is reported to have a different energy. This function is determined by the physical properties of the instrument, which was explained in detail in chapter 3.

In the ideal case, the response function is time independent and narrow. However, for the SPI instrument the narrowness degrades over time because of two effects: cosmic rays and the Earth's radiation belt. These radiations can defect the lattice structure of the Ge detectors, disturbing its regular structure. For this we need a model function which is time and energy-dependent, and here it is chosen as a truncated exponential tail-function T(E). This function has to be incorporated to the ideal case of a response function, which is modeled as a Gaussian line shape G(E) (Siegert et al., 2016). Thus for each line i and detector j the response is given by

$$R(E) = \sqrt{\frac{\pi}{2}} \frac{A_{0,ij}\sigma_{ij}}{\tau_{ij}} \exp\left(\frac{2\tau_{ij}(E - E_{0,ij}) + \sigma_{ij}^2}{2\tau_{ij}^2}\right) \operatorname{erfc}\left(\frac{\tau_{ij}(E - E_{0,ij}) + \sigma_{ij}^2}{\sqrt{2}\sigma_{ij}\tau_{ij}}\right), \qquad (5.2)$$

where $A_{0,ij}$ is the normalization factor, $E_{0,ij}$ is the energy peak of line i and detector j, σ_{ij} is the line width in a particular detector and τ_{ij} is the degredation parameter which shows the changes of the line shape with respect to time.

The energy response function ($R(E)$) is normalized to 1. So it shows the contribution of the 511 keV to the nearby energies (from 500 to 520 keV) at the 511 keV line. Since the

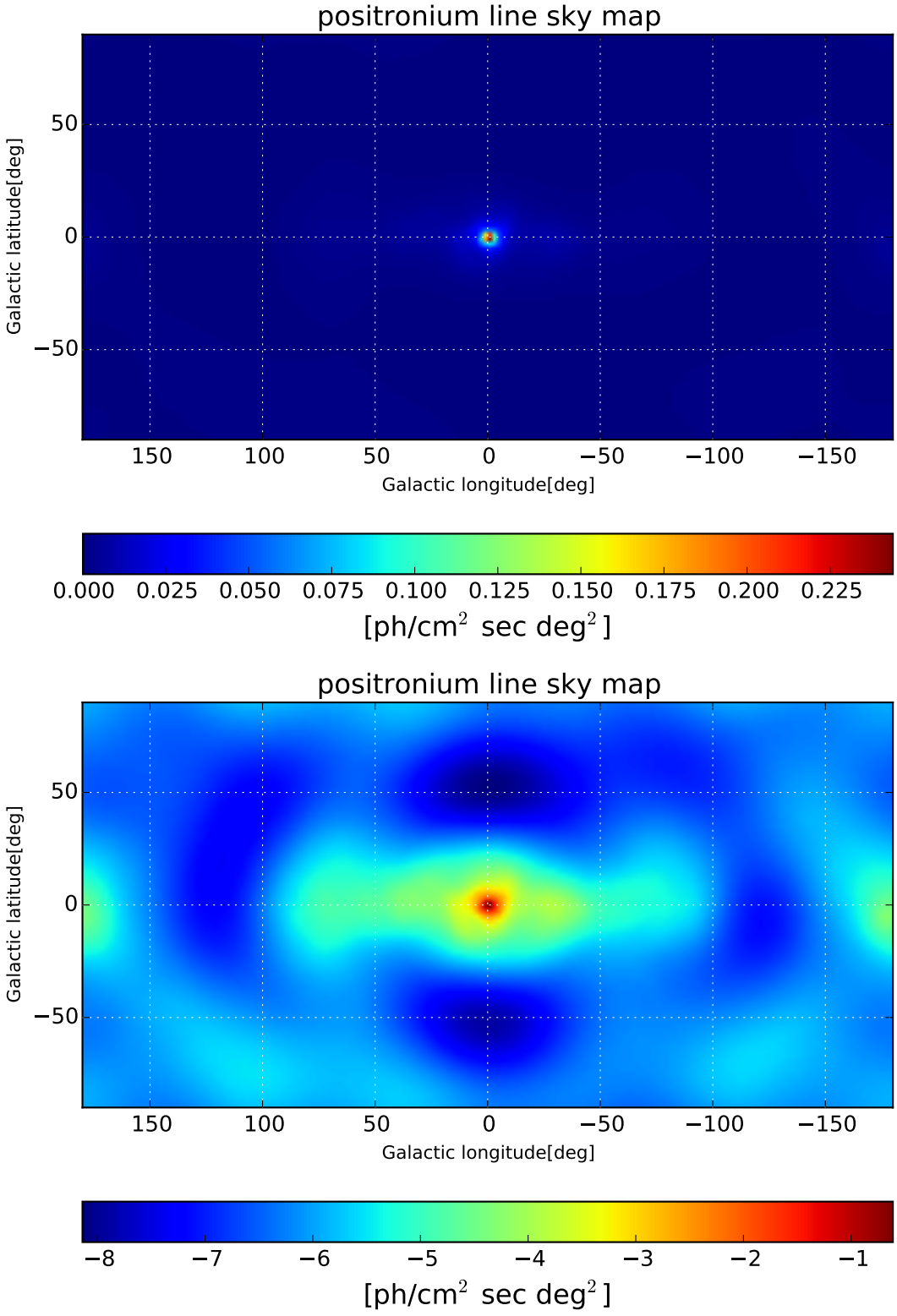

Figure 5.9: Reconstructed sky map from data in the energy range of 500-520 keV using D^{3}PO and the background model by Siegert et al. (2016) (top panel). The map on a logarithmic scale ($\log_{10}$) is depicted in the bottom panel.

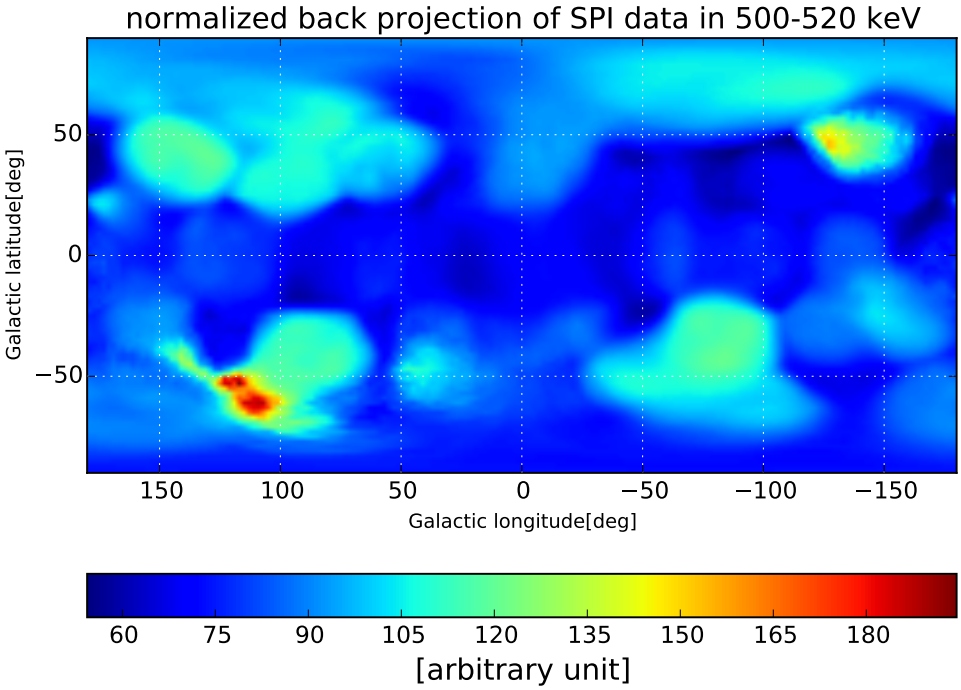

Figure 5.10: Normalized back projection of SPI data in the energy range of 500-520 keV.

energy bin width in our analysis is 0.5 keV, the total number of used bins is 41. Here the data are combined across a broader range of energies from 500 to 520 keV, all in all $n_d \times n_e$ numbers for each pointing

$$R_e^{(\text{spectral})} \otimes \sum_{l,b} R_{p,d,l,b}^{(\text{spatial})} s_{l,b} = \lambda_{p,d,e}, \tag{5.3}$$

in which p, d, e and l, b are pointing, detector, energy and position, respectively. λ is the expectation value of the data.

The noise is Poissonian shot noise, which varies according to Eq. 4.2. Similar to the signal contribution within the energy range of 510.5-511 keV, the sky signal in the energy range of 500-520 keV contributes only 1% to the background dominated data. Thus, to cope with this low signal-to-noise ratio, we assume to know the correct signal power spectrum during the analysis with D^{3}PO. The algorithm is modified to the SPI response, which includes the sparse matrix representation of the spatial response (Eq. 3.2) and the energy response (Eq. 5.2). The reconstructed map is shown in the upper panel of Fig. 5.9. According to it the Galactic positron annihilation emission from data in the energy range of 500-520 keV is concentrated in the Galactic center, confirming our previous finding, with significance of $\sim 25\sigma$ (the bottom panel in Fig. 5.12). However, it is more intense than that in the sky map of reconstructed from only one energy bin. The central flux from l=30° to -30° and b=-20° to 20° is 1.06 $\times$ 10^{-2} ph/cm^2 sec. The reconstructed map in logarithmic units (the bottom panel of Fig. 5.9) also reveals that the disk contribution to

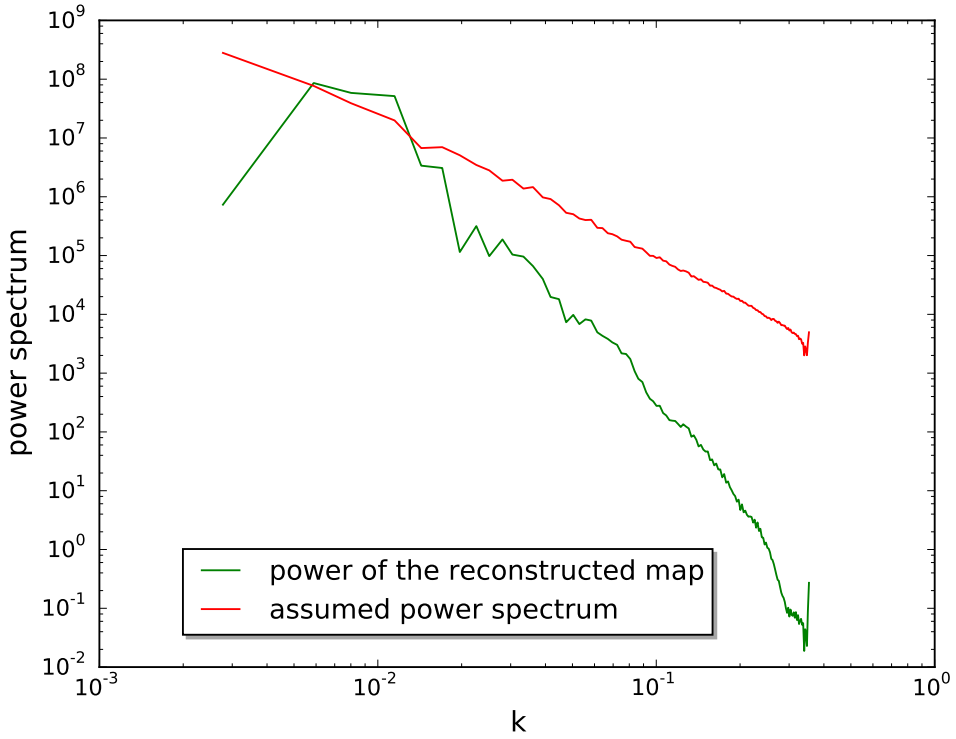

Figure 5.11: Corresponding power spectrum of the reconstruction from the data in the energy range of 500-520 keV.

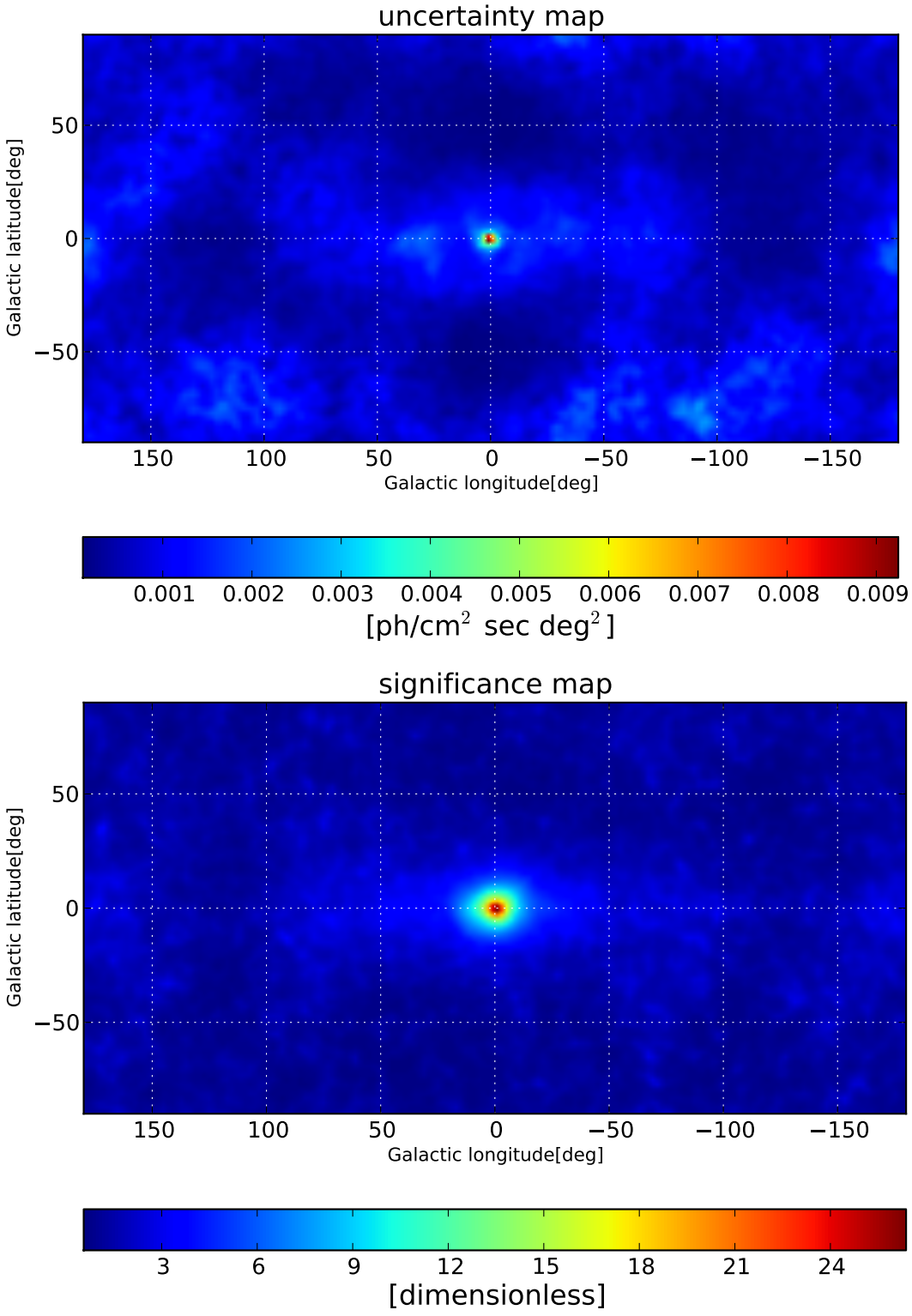

Figure 5.12: Uncertainty and significance maps of the reconstruction using D^{3}PO in the energy range of 500-520 keV and the background model by Siegert et al. (2016).

the positron annihilation emission is larger than that of only one bin of 510.5-511 keV. The total flux in the reconstructed emission from the data in the energy range of 500-520 keV is $\sim 2.7 \times 10^{-2}$ ph/cm^2 sec, that is about 5.8 times more than that of the sky map from the data in the energy range of 510.5-511 keV. Here also the power spectrum, discussed in Sec. 5.2, is used. The uncertainty map of the reconstruction map is shown in the upper panel of Fig. 5.12. If we compare the field value of the map in the Galactic center (top panel in Fig. 5.9), which is ~ 0.2 ph/cm^2 sec deg^2, to that of the uncertainty map, amounting to ~ 0.009 ph/cm^2 sec deg^2, we can see the obtained map has the required accuracy.

5.3 Simultaneous reconstruction of the SPI background and sky

Up to now, for all mock and real reconstructions performed in this thesis, a model for the SPI background by Siegert et al. (2016) was used. This background model is based on the detector pattern as noted in Sec. 3.4. The background which hits the instrument comes from several sources, and the detectors have different reactions with respect to these sources due to the different detectors location. For example, central detectors with respect to the camera, receive more intense radiations from Ge isotopes, contributing to the line component of the background, while side detectors in the outskirt of the camera see less of this background. Also for the continuum background, a particular signature is expected. We know from Sec. 3.2 that there is a system in SPI, the so-called pulse shape discrimination (PSD), which distinguishes between single and multiple events, removing multiple nearby detector events from the data, as these are mostly caused by cosmic rays hits. It was also mentioned that some of the detectors stopped working at various points in time. From these moments on, the cosmic ray event rejection of the adjacent detectors becomes less effective, and their background counts increase significantly. This effect can be seen in Fig. 5.13. For these reasons, the individual detectors have their individual expected background count rates, which have to be determined from the data. Furthermore, the intensity of the detector pattern changes with time when the instrument reorients to different pointings.

The variations in cosmic rays intensity are due to changes in the solar activity, or due to orbital changes as radiation belts variations. The intensity changes in the background model scale the detector pattern for each revolution. This temporal variation was reconstructed through a MLE method by Siegert et al. (2016). In this section of the thesis, we identify this normalization through D^{3}PO, instead. The background changes with respect to the pointings are investigated in order to image the sky of Galactic positrons annihilation independently of the time variation of Siegert et al. (2016)' background model (but not of the spectral structure of it). In this process we retain the detector ratios from Siegert's model, but reconstruct the background time variation while using a prior in the form of a log normal distribution. To reconstruct the background variations, it is required to determine a normalized response function for the background. Thus in this case, the

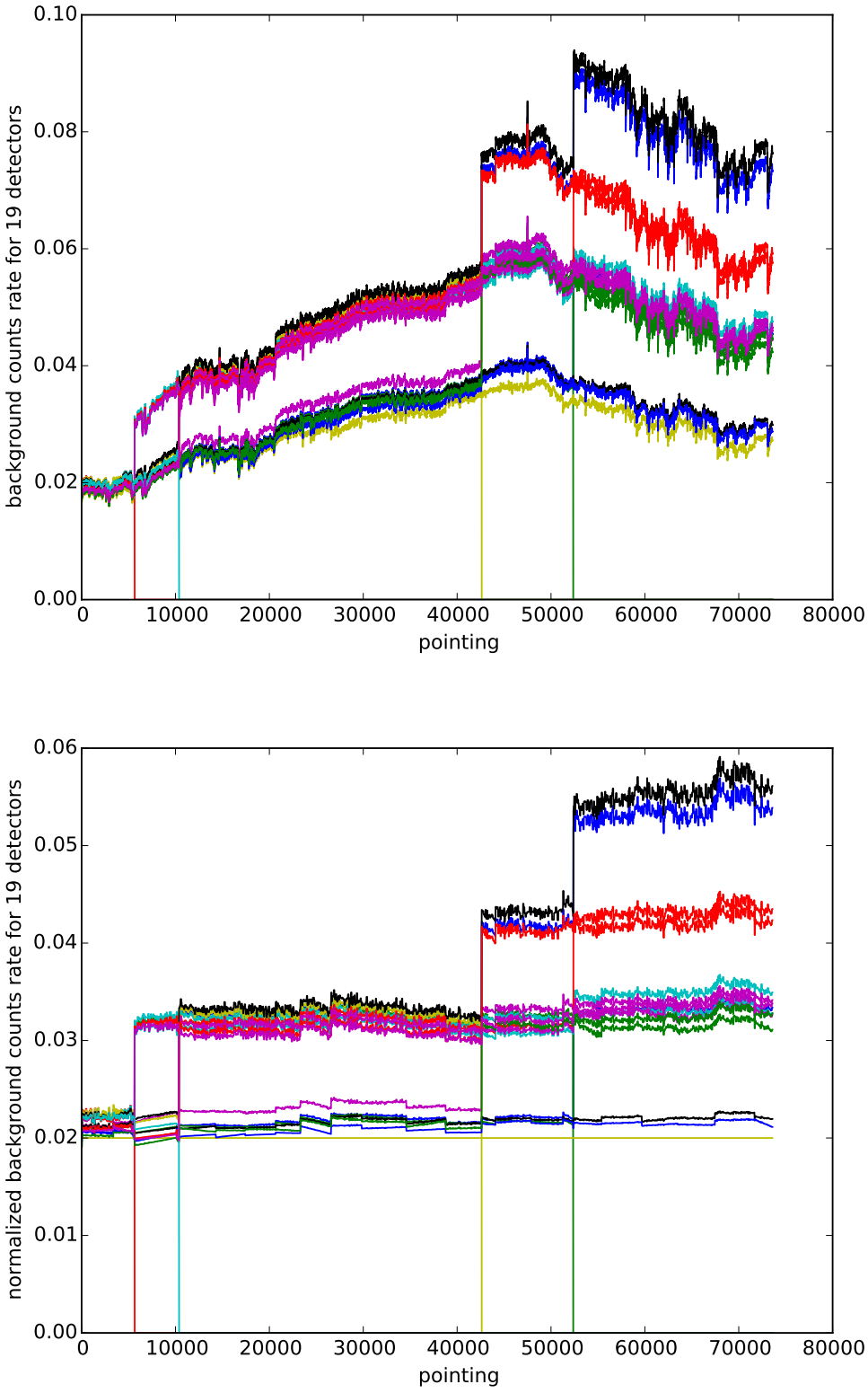

Figure 5.13: Unnormalized (top panel) and normalized (bottom panel) form of the SPI background count rates.

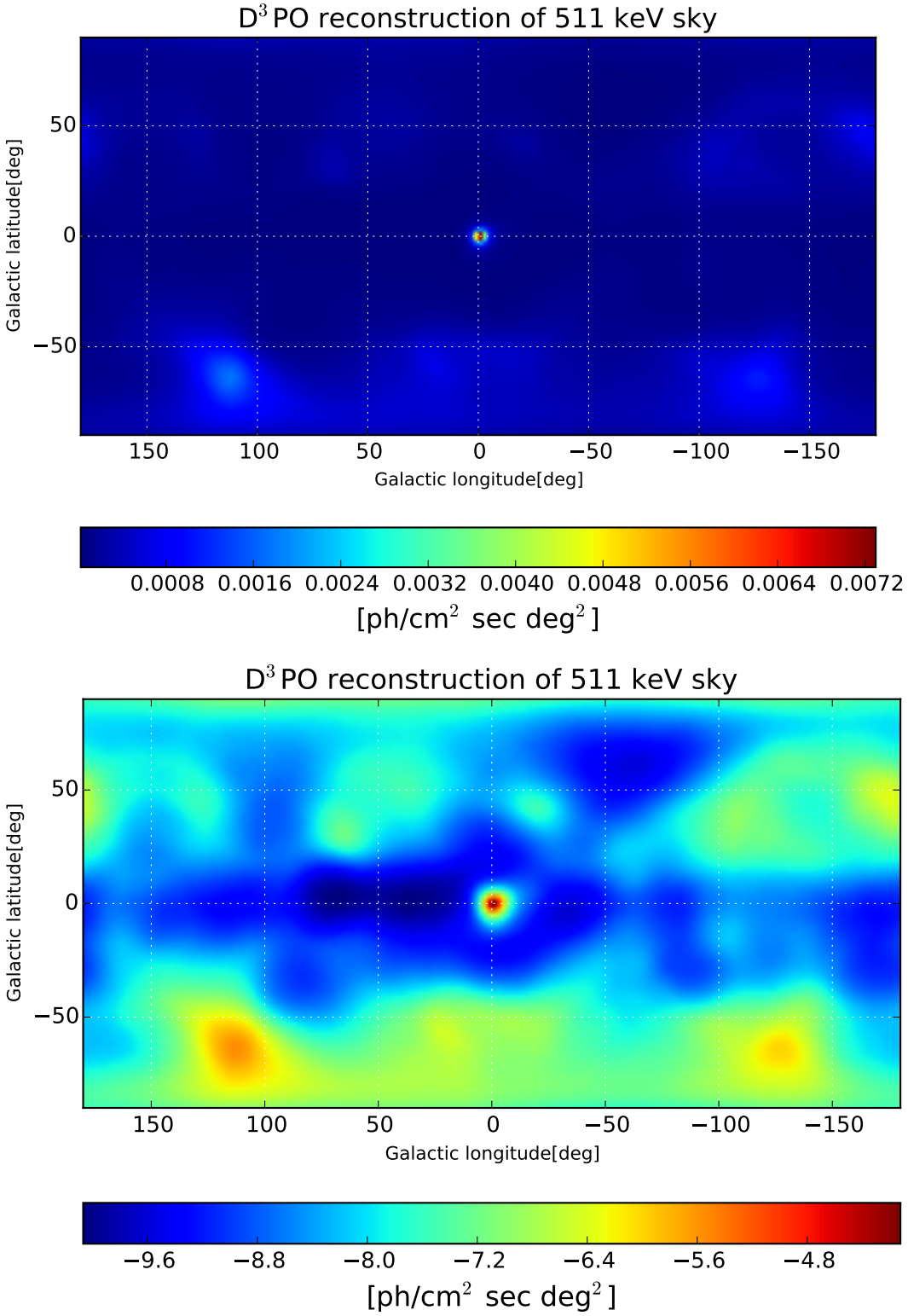

Figure 5.14: Reconstructed sky map from data emission in the energy range of 510.5-511 keV by D^{3}PO, while simultaneously inferred the time component of the background inferring (top panel). The same in logarithmic unit ($\log_{10}$) (bottom panel).

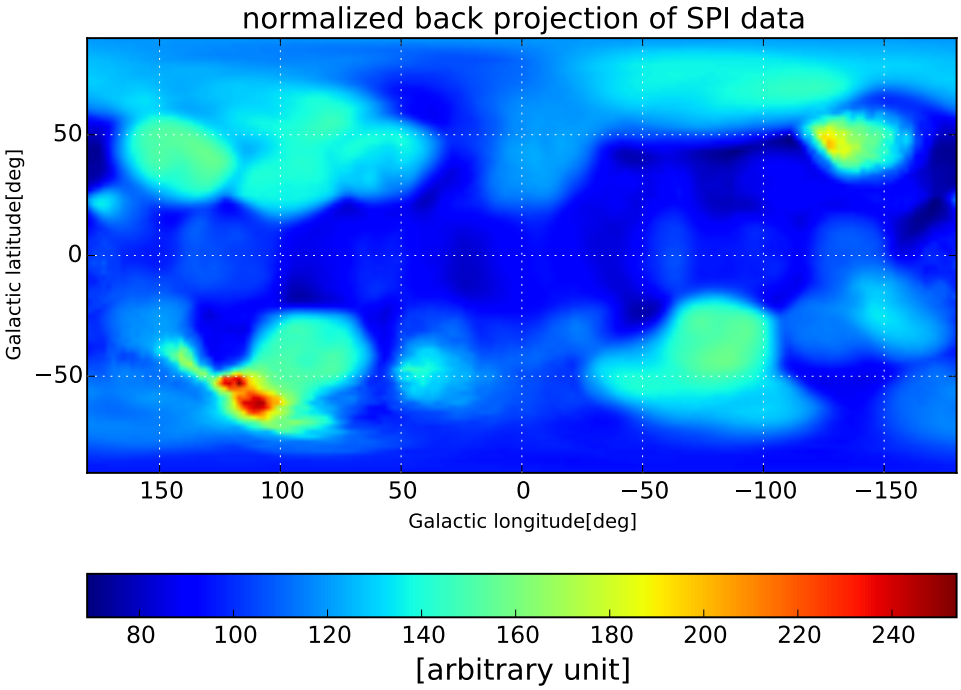

Figure 5.15: Normalized back projected SPI data in th energy range of 510.5-511 keV

data model in D^{3}PO is changed to

$$d = R^{(s)}(e^s) + R^{(bc)}(e^b) + n, \tag{5.4}$$

in which

$$\lambda = R^{(s)}(e^s) + R^{(bc)}(e^b). \tag{5.5}$$

As it can be seen from Eq. 5.5, two signals (sky and background) are to be inferred from the SPI data by the D^{3}PO algorithm. $R^{(s)}$ and $R^{(bc)}$ are the sky and background response, respectively. $R^{(s)}$ in Eq. 5.4 is equivalent to the spatial R in Eq. 4.3.

The background response is determined using the background model by Siegert et al. (2016), by removing the time-dependency of the background. Since the active times differ between detectors, we calculate the count rates in the background model by division of the counts by the live time for each detector over all pointings. Note that in this background reconstruction the energy bin of 511 keV is chosen. The top panel of Fig. 5.13 shows the count rates of the background model of Siegert et al. (2016), if the counts are divided by the live time for each detector over all pointings. The top panel of Fig. 5.13 shows how the count rates change for each detector across all pointings. In the bottom panel of Fig. 5.13 all count rates are normalized by the count rate of the detector number 13, which is indicated by the yellow line in the top panel of Fig. 5.13. In this step, a good fraction of the time-dependency of the count rates of the mentioned background model is removed. The normalized count rates in the form of a matrix in dimension of $\mathrm{n_{pointing}} \times \mathrm{n_{detector}}$ can then be used as the response pattern of the background

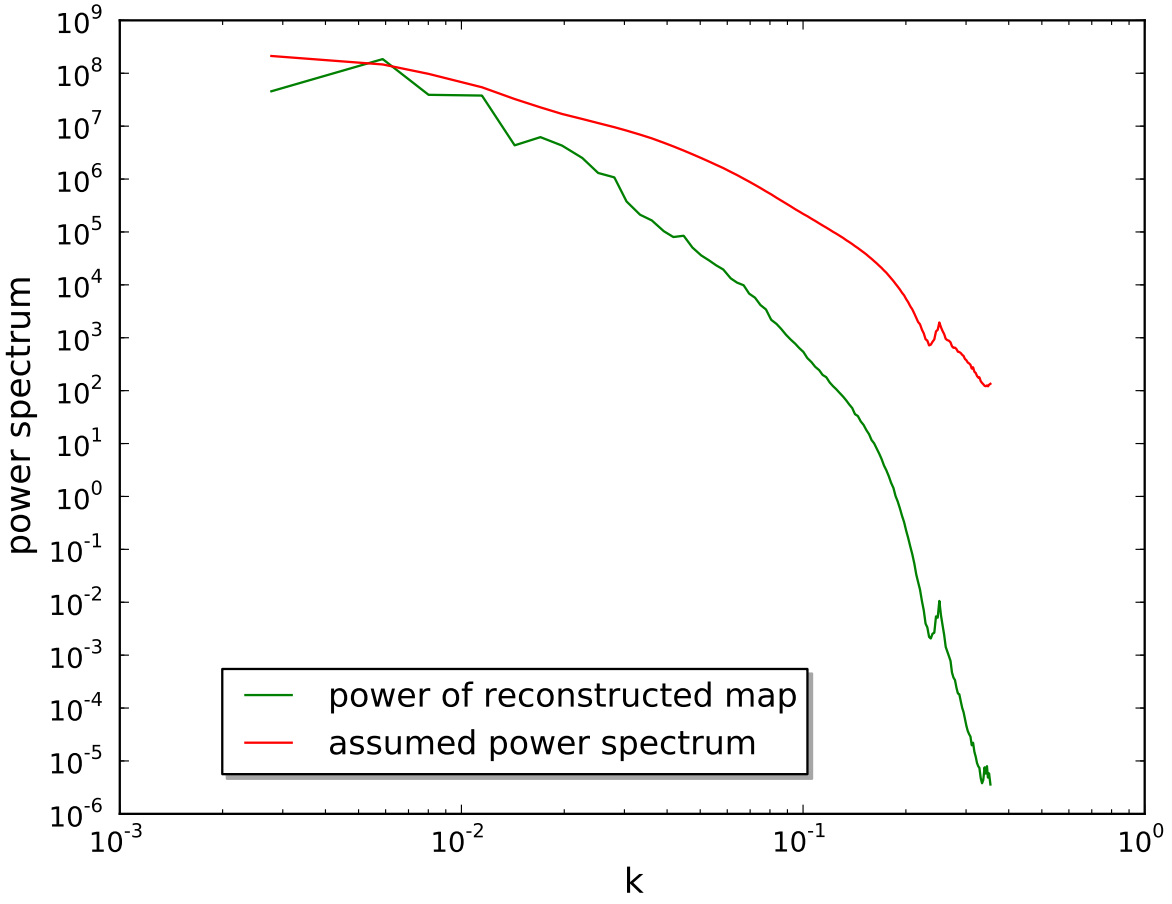

Figure 5.16: Corresponding power spectrum of the reconstruction shown in Fig. 5.14.

$$R^{(bc)}_{p,d}\, s^{(bc)}_{p} = \lambda^{(bc)}_{p,d}\,, \tag{5.6}$$

where $\lambda^{(bc)}$ is the expectation value of the data contributing to the background. $s^{(bc)}$ is going to be inferred by the D^{3}PO. In D^{3}PO, the signal for the background timeline ($s^{(bc)}$) is modeled as a log-normal field. Thus, the notation for $s^{(bc)}$ is replaced by e^b, as it is also implemented for the sky signal (e^s).

Now we investigate the performance of the D^{3}PO algorithm using the data model in Eq. 5.4. According to the D^{3}PO, we assume a log-normal signal statistics both for the sky and background. Furthermore, the Poisson noise statistics is supposed. The Hamiltonian in D^{3}PO (Eq. 2.45) changes, due to the change of λ in Eq. 5.4

The way how we perform the process of the signals inference is described as following: as it was shown in Sec. 5.1, the sky signal was inferred from the background model by Siegert et al. (2016), and it was shown as the reconstructed map in Fig. 5.3. Suppose that this map is the correct sky map of the Galactic positron annihilation emission at 511 keV. Then the first term in Eq. 5.5 is known as $\lambda^{(s)}$ and it is given by

$$\lambda = \lambda^{(s)} + R^{(bc)}(e^b). \tag{5.7}$$

Therefore, we could let the D^{3}PO algorithm reconstruct the signal associated to the background (e^b). In the next step, the derived reconstructed timeline, of the background signal (e^b) is assumed to be known, $\lambda_{(bc)}$

$$\lambda = R^{(s)}(e^s) + \lambda^{(bc)}, \tag{5.8}$$

where the $R^{(s)}$ is the response for the sky, as it was used before in all mock and real reconstructions. Now the signal sky (e^s) will be inferred by the algorithm. We do this iteration over and over until the fields of the reconstructed maps are converged. The convergence is assumed for the step in which the two last sky maps (i.e., map$^{(n)}$ and map$^{(n-1)}$) differ only as $\sim 10^{-7}$ ph/cm^2 sec deg^2 in $\sqrt{\frac{1}{N}\sum_i (\mathrm{map}_i^{(n)} - \mathrm{map}_i^{(n-1)})^2}$. This criteria is also used for the background reconstruction. As in the previous sky inference by D^{3}PO, we also assume to know the sky power spectrum, while the power of the background signal is inferred from the algorithm, which is well possible as the background dominants the SPI data.

The map reconstructed of the sky signal (e^s) by D^{3}PO is shown in the top panel of Fig. 5.14, which exhibits a strong contribution in the Galactic center. The total flux of the reconstructed map amounts to $\sim 4.16 \times 10^{-3}$ ph/cm^2 sec which is in a good match with the previous results of the sky image within energy range of 510.5-511 keV, when the background model by Siegert et al. (2016) is considered (Fig. 5.3). However the sky map, which assumes the Siegert's background model (Fig. 5.3), is much more intense in the Galactic center amounting the central flux to $\sim 1.5 \times 10^{-3}$ ph/cm^2 sec, compared to the central flux of $\sim 1.9 \times 10^{-4}$ ph/cm^{-2} sec of the sky map in case of the reconstructed timeline of the background (Fig. 5.14). The reason why the total fluxes in both mentioned reconstructed maps are roughly the same, despite of the intensity difference in the Galactic

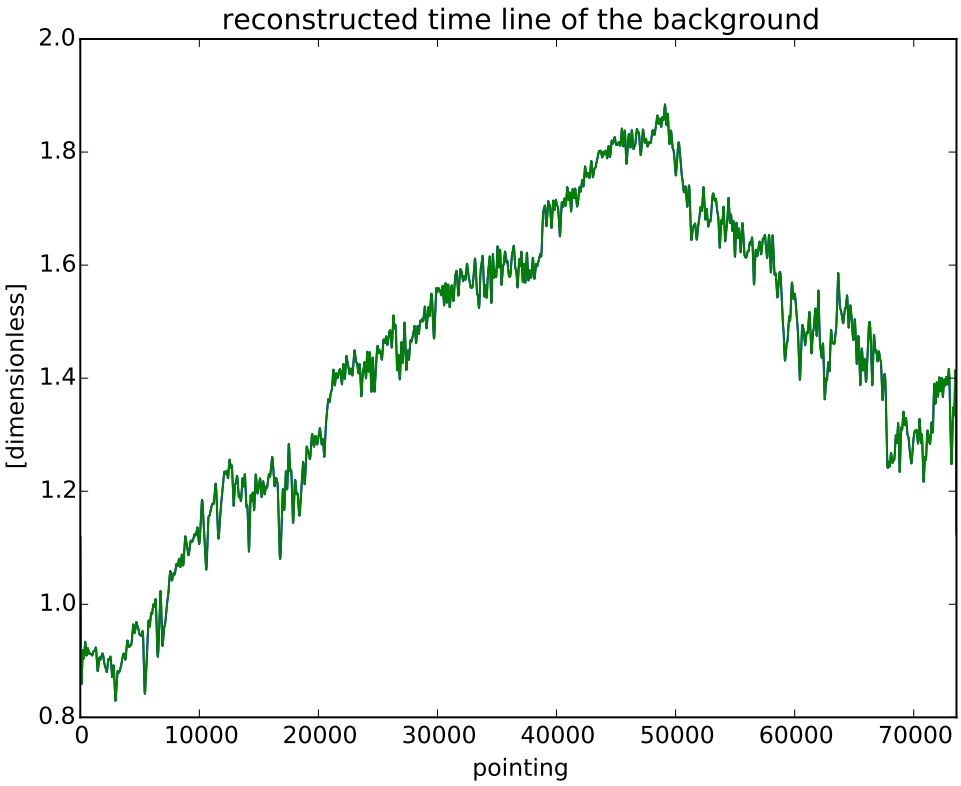

Figure 5.17: Result of the background intensity reconstructed by D^{3}PO, as a function the pointing.

center is due to the appearance of emission in other regions of the sky except the Galactic center. In fact, if we consider the map on the logarithmic scale, which is shown in the bottom panel of Fig. 5.14. Two bright regions are seen at $l = 90°$, $b = -30°$ and $l = -130°$, $b = -30°$, which are most likely artifacts, since the exposure of INTEGRAL to sky those areas is poor and not much flux is expected at those large latitudes.

The reconstructed background variations with respect to the pointings are shown in the top panel of Fig. 5.17. We remind that the background power was also inferred from the data.

In order to give an impression how good this reconstruction is, the top panel of Fig. 5.19 compares the time component of the background reconstructed by the D^{3}PO to the time component of the background model of Siegert et al. (2016). It can be seen from this comparison the good match between them. Also the corresponding power spectra are presented in the bottom panel, which shows the smooth structure of the reconstructed background, as its power falling down at small scales. Furthermore, to compare better the result of the reconstructed time component of the background to the time component of the background model, Fig. 5.20 shows the differences. The uncertainty map of the background intensity reconstruction is represented in Fig. 5.18.

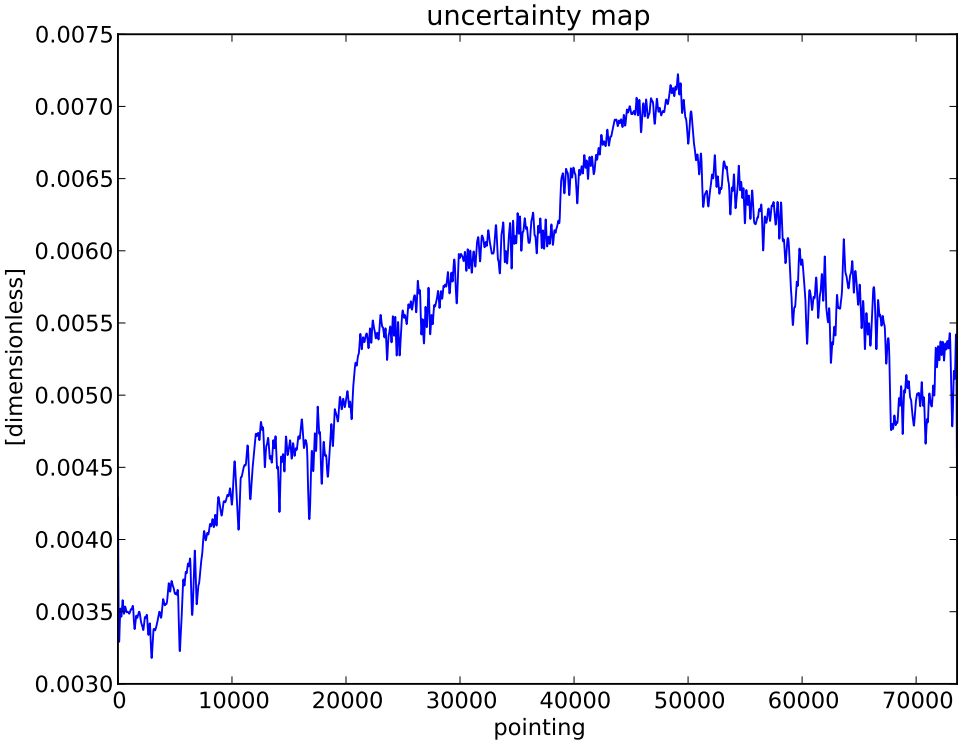

Figure 5.18: Uncertainty map of the background intensity reconstructed by D^{3}PO, as a function of pointing.

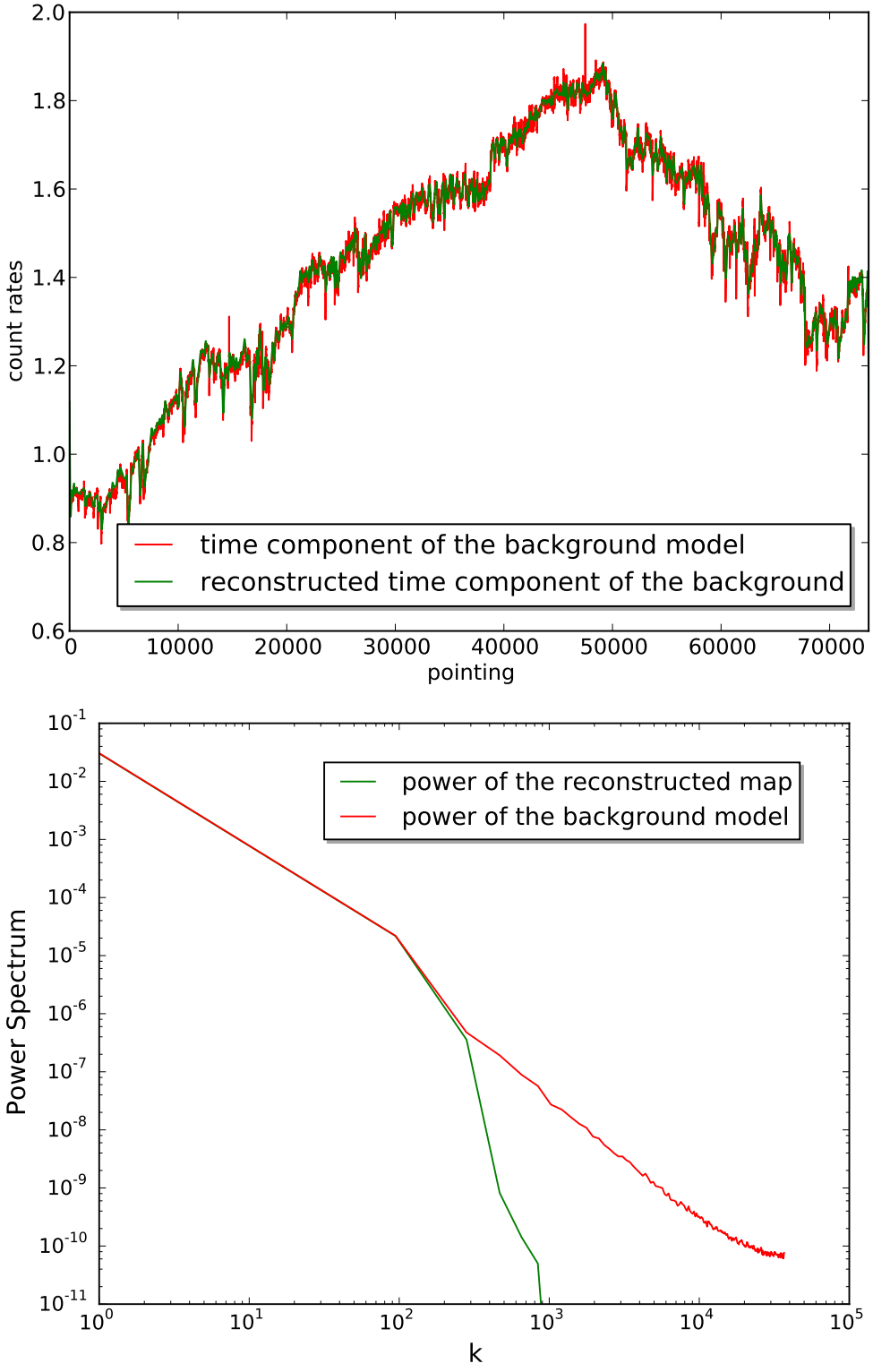

Figure 5.19: The reconstructed time component of the background using the D^{3}PO (green line of the top pannel), which is compared to the time component of the background model Siegert et al. (2016) (green line of the top pannel), and their corresponding power spectrum (bottom panel).

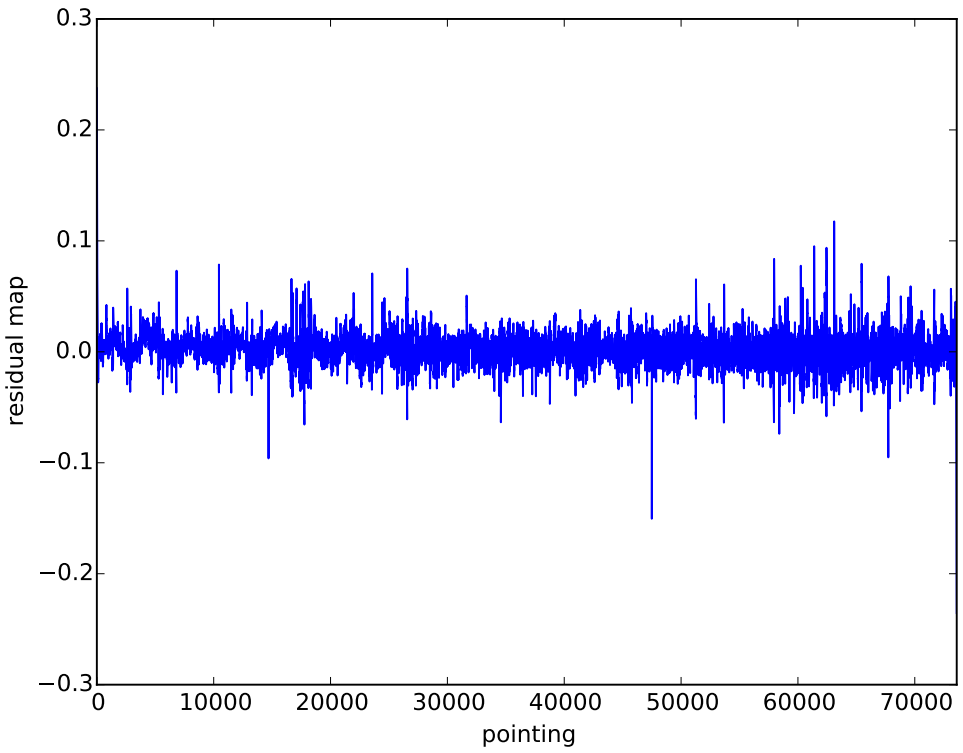

Figure 5.20: Differences of the reconstructed time component of the background using D^{3}PO minus the time component of the background model of Siegert et al. (2016).

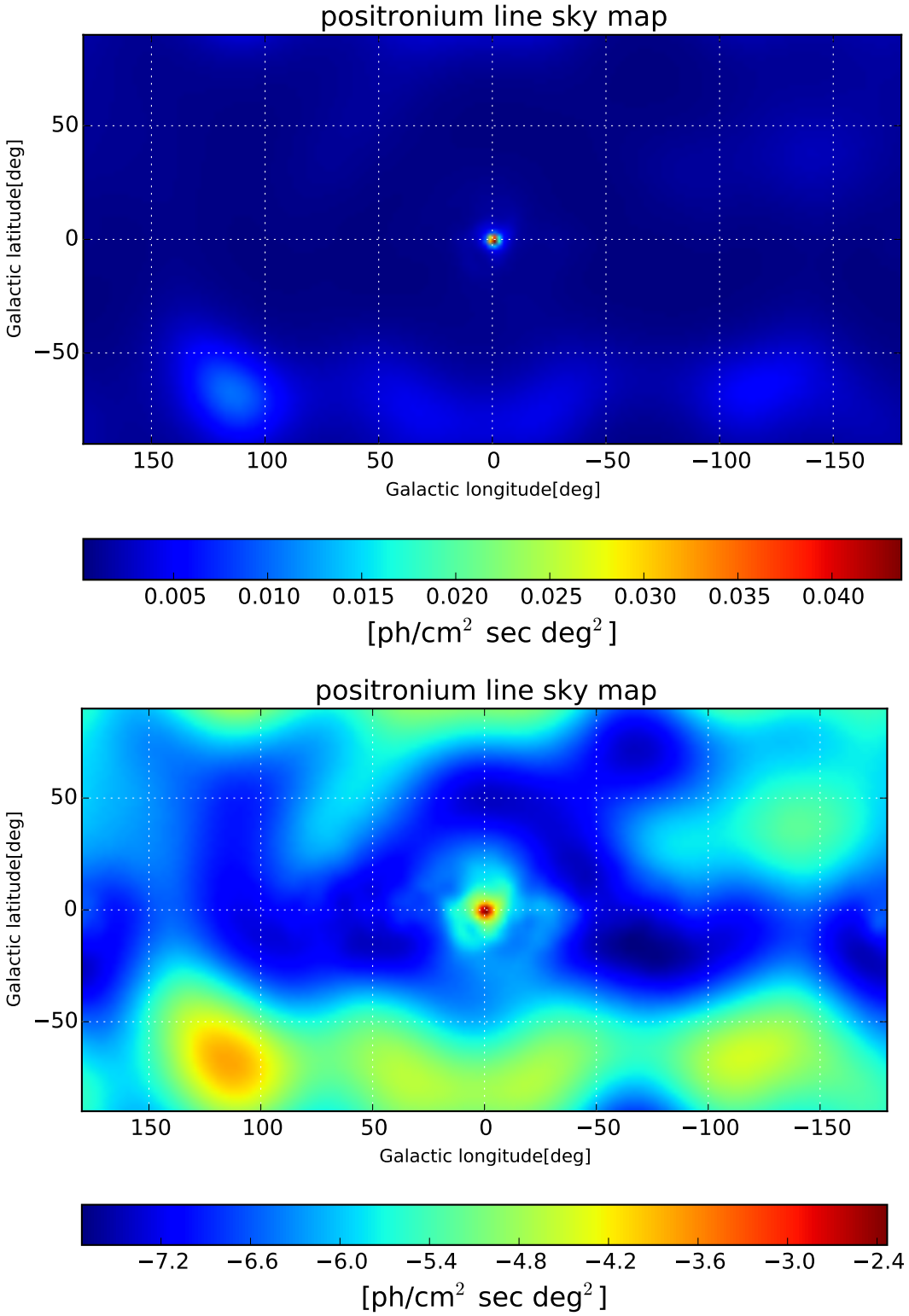

Figure 5.21: Reconstructed sky map from data emission in the energy range of 500-520 keV by D^{3}PO, while simultaneously inferred the time component of the background inferring (top panel). The same in logarithmic unit ($\log_{10}$) (bottom panel).

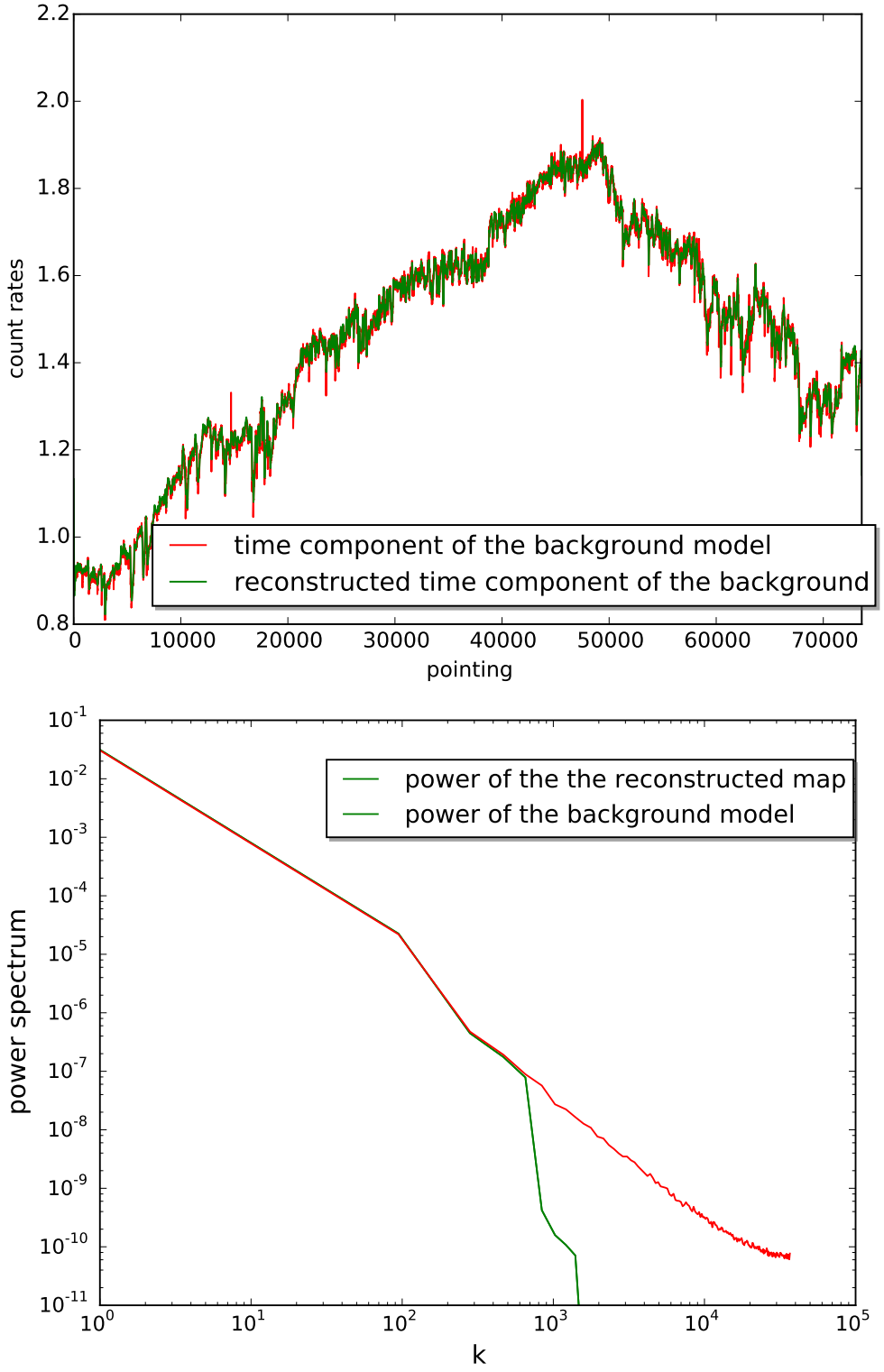

Figure 5.22: The reconstructed time component of the background using the D^{3}PO (green line of the top pannel), which is compared to the time component of the background model Siegert et al. (2016) (green line of the top pannel) in energy range of 500-520 keV, and their corresponding power spectrum (bottom panel).

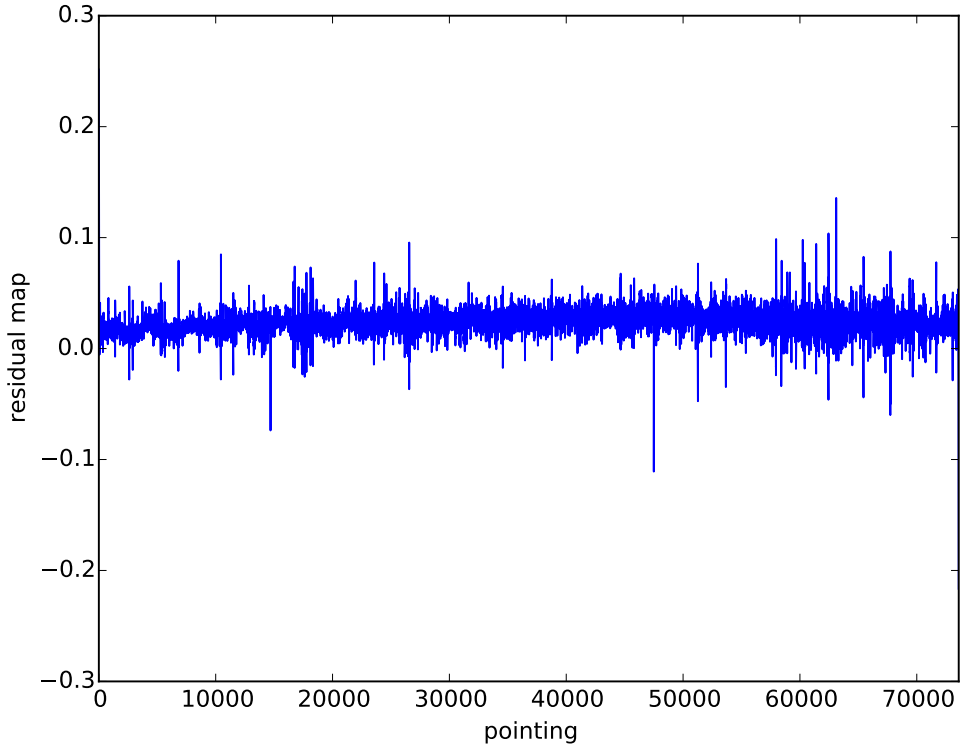

Figure 5.23: Differences of the reconstructed time component of the background using D^{3}PO minus the time component of the background model of Siegert et al. (2016) for energy range of 500-520 keV.

Chapter 6

Conclusions and outlooks

We explored Bayesian imaging techniques to study the γ-ray emission of the Galaxy. In this investigation, we used Information Field Theory (IFT) to derive imaging algorithms. These were applied onto the data observed by the SPI spectrometer, on board the high-energy space telescope INTEGRAL. Two algorithms that derived in the framework of IFT, were applied to the data: the Wiener Filter (WF) and a modified version of D^{3}PO. Since an instrumental background contributes largely to the SPI data, we performed series of simulations to test and verify these two algorithms using mock data. To simulate a realistic physical situation, we model the sky in such a way that it only contributes 1٪ to the mock data. The rest contributions include a model for the SPI background (Siegert et al., 2016) and suffers from Poisson shot noise. We presented satisfying results of the simulated data analysis using both the WF and D^{3}PO. We present the accuracy of the reconstructions by providing their uncertainty and residual maps. We further showed that it is not possible to infer the power spectrum of such a poorly produced sky signal. We repeat the simulation by modeling a stronger signal, with 100 times more total sky flux compared to that one in the previous simulation. In this case D^{3}PO is well able to reconstruct small-scales image details and the signal power spectrum.

We applied the modified D^{3}PO to the all-sky 11 years of data provided by SPI, and investigated the morphology of the γ-ray emission within the energy range of 510.5-511 keV. In this application the mentioned background model of Siegert et al. (2016) was assumed. The results of the reconstructed map, which is interpreted as the Galactic positron annihilation emission in the energy range of 510.5-511 keV, showed a strong and intense distribution of the emission in the Galactic center. This confirms that the members of the old stellar population, specifically type Ia supernova (SNIa) is the most promising source for positron production (Prantzos et al., 2011), if their annihilation site is close to the production one. The morphology of the map also reveals a weak contribution of the Galactic disk. A total flux of $\sim 4.6 \times 10^{-3}$ ph/cm^2 sec is found by the reconstruction.

In the next application of the modified D^{3}PO on the SPI data, a wider energy range around 511 keV is considered, from 500 to 520 keV. The reconstructed map of the Galactic positronium line emission in this energy range confirms our previous findings about the morphology of the emission. The emission was again strongly concentrated in the Galactic

center, and a weak contribution from the Galactic disk was observed. The measured total flux of this reconstructed map was $\sim 2.7 \times 10^{-2}$ ph/cm^2 sec, which is 5.8 times more than the total flux of the sky map in one energy bin at 511 keV.

In the final study of the thesis, we planned to image the sky of Galactic positron annihilation independently of the time variation of the background model by Siegert et al. (2016), but not of the spectral structure of it. In this investigation also the detector pattern of the background signal is kept to the one obtained by Siegert et al. (2016). The obtained sky map from this approach showed the intensive contribution of the emission in the Galactic center, however we could not see any signature of the Galactic disk at 511 keV radiation. In this map two bright regions of the emission were observable at l=90°, b=-30° and l=-130°, b=-30°, which are most likely artifacts. The measured total flux in this reconstructed map was $\sim 4.16 \times 10^{-3}$ ph/cm^2 sec.

Our future research will be the simultanous reconstruction of the sky and time component of the background in energy range of 500 to 520 keV. Due to limitations in the available computer and human time, this could not achieved during the thesis research period. If we compare the obtained result of the sky map, when the Siegert's background model was used, the sky map, when the time component of the background was reconstructed, the significance of the background determination becomes visible. Both mentioned sky maps differ in terms of the morphology and the central fluxes in the Galactic center. In the former map a faint contribution of the Galactic disk was seen, while in the latter one there is no detection of the Galactic disk emission. It is true that both sky maps show a strong and intense emission in the center, but the central flux in the former is large than in the latter. Therefore, we need to precisely determine the background contribution to the SPI data to significantly improve the imaging of the Galactic positron annihilation emission.

Bibliography

Aharonian F., Bergström L., Dermer C., 2013, Astrophysics at Very High Energies: Saas-Fee Advanced Course 40. Swiss Society for Astrophysics and Astronomy, Saas-Fee Advanced Course, Volume 40. ISBN 978-3-642-36133-3. Springer-Verlag Berlin Heidelberg, 2013, 40

Bohm, G. and Zech, G., 2010, Introduction to statistics and data analysis for physicists

Bouchet L., Amestoy P., Buttari A., Rouet F.-H., Chauvin M., 2013, Astronomy and Computing, 1, 59

Bouchet L., Jourdain E., Roques J.-P., 2015, Astrophysical, 801, 142

Bouchet L., Jourdain E., Roques J.-P., Strong A., Diehl R., Lebrun F., Terrier R., 2008, Astrophysical journal, 679, 1315

Bouchet L., Roques J. P., Jourdain E., 2010, Astrophysical journal, 720, 1772

Cheng L. X. et al., 1997, apjl, 481, L43

Diehl R., 1994, in NATO Advanced Science Institutes (ASI) Series C, Vol. 445, NATO Advanced Science Institutes (ASI) Series C, Genzel R., Harris A. I., eds., p. 3

Diehl,R., 2001, Gamma-Ray Production and Absorption Processes, illustrated, reprint edn. Springer Berlin Heidelberg

Dirac P., 1931, The Royal Society, 133

Enßlin T., 2013, 1553, 184

Enßlin T., 2014, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 1636, 49

Enßlin T. A., Frommert M., Kitaura F. S., 2009, Physical Review, 80, 105005

Ferrière K., Gillard W., Jean P., 2007, astronomy and astrophysics, 467, 611

Gregory P., 2005, Bayesian Logical Data Analysis for the Physical Sciences, illustrated, reprint edn. Cambridge University Press, 2005

Higdon J. C., Lingenfelter R. E., Rothschild R. E., 2009, astrophysical journal, 698, 350

Jaynes E. T., 1957, Phys. Rev., 106, 620

Jaynes, E. T. and Bretthorst, G. L, 2003, Probability Theory: The logic of Science, illustrated, reprint edn. Cambridge University Press, 2003

Jean P., Knödlseder J., Gillard W., Guessoum N., Ferrière K., Marcowith A., Lonjou V., Roques J. P., 2006, astronomy and astrophysics, 445, 579

Jean P. et al., 2000, 510, 708

Johnson, III W. N., Harnden, Jr. F. R., Haymes R. C., 1972, Astrophysical Journal, 172, L1

Kinzer R. L., Milne P. A., Kurfess J. D., Strickman M. S., Johnson W. N., Purcell W. R., 2001, apj, 559, 282

Kinzer R. L., Purcell W. R., Johnson W. N., Kurfess J. D., Jung G., Skibo J., 1996, aaps, 120, 317

Knödlseder J. et al., 2005, Astronomy and Astrophysics, 441, 513

Leventhal M., MacCallum C. J., Huters A. F., Stang P. D., 1982, Astrophysical Journal, 260, L1

Lucy L. B., 1974, Architecht, 79, 745

Milne P. A., Kurfess J. D., Kinzer R. L., Leising M. D., Dixon D. D., 2001, in ESA Special Publication, Vol. 459, Exploring the Gamma-Ray Universe, Gimenez A., Reglero V., Winkler C., eds., pp. 145–148

Oppermann N., Selig M., Bell M. R., Enßlin T. A., 2013, physical review, 87, 032136

Prantzos N., 1993, aaps, 97, 119

Prantzos N. et al., 2011, Reviews of Modern Physics, 83, 1001

Purcell W. R. et al., 1997, apj, 491, 725

Richardson W. H., 1972, Journal of the Optical Society of America (1917-1983), 62, 55

Selig M., Enßlin T. A., 2015, Astrophysical journal, 574, A74

Shannon C. E., 1948, The Bell System Technical Journal, 27, 379

Shannon, C. T, 1949, Communication in the Presence of Noise, Vol. 37. pp. 10–21

Siegert T., Diehl R., Khachatryan G., Krause M. G. H., Guglielmetti F., Greiner J., Strong A. W., Zhang X., 2016, astronomy and Astrophysics, 586, A84

Skilling, J. and Bryan, R. K., 1984, Monthly Notices of the Royal Astronomical Society, 211, 111

Strong A. W., 2003, aap, 411, L127

Tomlinson J. R. et al., 2015, Physical Review Letters, 115, 052702

Vedrenne G. et al., 2003, astronomy and astrophysics, 411, L63

Weidenspointner G. et al., 2008b, nar, 52, 454

Weidenspointner G. et al., 2008c, New Astronomy Reviews, 52, 454

Acknowledgments

I would like to thank all people who support me scientifically to do research at MPA and also non-scientifically to make Germany a real second home for me. Their contributions helped me a lot, so that without having them, I would not be where I am now.

First of all, I would like to acknowledge the support of my first supervisor, Torsten Ensslin, for his support in overcoming numerous problems I have been facing during my research. I would like to thank my second supervisor, Roland Diehl, who taught me astrophysics in his special and patient way, which is deserved for a real teacher.

I am really grateful of Fabrizia Guglielmetti for her special support. After her presence in our group, the procedure of my thesis really went in a much smoother way to get result of better quality.

Torsten's support did not only come from him, but also from his great group at MPA. I am so grateful of Marco Selig, one of the exceptional person I have been faced in my life to support and teach me the typical issues in IFT and NIFTY. Thanks a lot to Maksim Greiner, for his great ideas in solving some problems in my thesis, Sebastian Dorn, Henrik Junklewitz and Valentina Vacca for their effective and positive presence. I would like to thank Daniel Pumpe, my office mate in our group for his support and also contribution to translate the summary into German.

I would like to thank the people who helped me a lot in computational support, Andrew Weiss and Martin Reinecke.

Thanks a lot also to Gabriele Kratschmann, an exceptional secretary at MPA, who did her support beyond of her charge.

I am so happy for having great friends, for their positivity and intimacy. My dear Ghazaleh, Maryam and Narges, thank you so much.

My deep gratitude goes to my family, my parents to encourage and support me in this critical way. Very special thanks to my best friend in my life, Roozbeh, for his great moral support and interestingly for his great ideas in solving the numerical problems.