

UNIVERSITY OF CALGARY

Measuring Energetic Electron Precipitation using High Altitude Balloons and X-ray Spectroscopy

by

Matthew Patrick

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

SPACE PHYSICS

CALGARY, ALBERTA

February, 2022

© Matthew Patrick 2022

Abstract

This work develops experimental methods to infer energetic electron precipitation using X-ray measurements from high altitude balloons. The problem of creating a mapping between the energies and fluxes of secondary X-ray radiation and the causative precipitating electron spectrum is solved using monte-carlo simulations and computer models of the detectors. The inverse problem of determining the electron spectrum from X-ray measurements is solved from first principles by treating the underlying ill-conditioning using regularization and preconditioning methods. Validation using simulations of the precipitating electrons and detector demonstrates that the inverse problem is solvable for realistic experimental noise levels in the measured X-ray spectra. The problem of overfitting is solved using cross-validation techniques, which produces electron spectra that do not depend on assumptions about the amount of information that can be reliably determined from the experimental measurements.

X-ray spectrometer measurements from two University of Calgary high altitude balloon flights and a NASA BARREL balloon flight are analyzed using the developed methods. The resulting electron spectra are compared with the electron spectra obtained using conventional fitting techniques to show that exponential or mono-energetic models can under-fit the available data. A comparison between the BARREL balloon data and a magnetically connected RBSP measurement of the precipitating electron spectrum prior to scattering shows that the approximate energy range over which the applicable scattering processes in the radiation belts operate can be determined from the balloon data. The average electron flux measured from the balloon is shown to be consistent with the RBSP measurements under a strong scattering assumption.

The combination of models produced using the developed analysis being consistent with previous experiments and correct based on simulated data is used to advance the case that balloon X-ray spectrometer measurements carry more information than is typically resolved using more simple techniques.

Acknowledgements

Paragraph 1

Paragraph 2

Paragraph 3

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Symbols	xii
1 Introduction	1
1.1 Like Astronomy, but Closer	1
1.2 The Earth's Atmosphere and Geospace	4
1.3 Getting Above the Atmosphere with High Altitude Balloons	9
1.4 Scope of this Thesis	12
2 Energetic Electron Precipitation and Geospace	13
2.1 Introduction to Connected Geospace	13
2.2 Electrons in Geospace	16
2.3 Single Particle Motion	18
2.4 Scattering Processes	20
3 Computer Modelling and the Forward Problem	25
3.1 Problem Statement	25
3.2 Scattering and Energy Deposition Physics	26
3.3 Previous results on Energetic Electron Precipitation	28
3.4 GEANT4 and Monte Carlo Simulations	30
3.5 Representing the Earths Atmosphere	34
3.6 GEANT4 Physics and Simulation Construction	38
3.7 Validation of Results	42
3.8 Problem Separability	42
4 The Inverse Problem	51
4.1 Definition and Properties	51
4.2 Framing the X-ray Inversion Problem	53
4.3 Creating Synthetic Data	58
4.4 The Least-Squares Regression and the Bias-Variance Trade-off	60
4.5 Biased Estimators and Improving the Condition Number	61
4.6 Tikhonov Regularization and the L-Curve Method	67
4.7 Higher Order Regularization	77
4.8 Preconditioning Techniques	81
4.9 Constrained Optimization	87
4.10 Evaluation on Synthetic Data	89
5 Experiment Data and Analysis	122
5.1 Overview of The Balloon Experiments	122
5.2 Detector Design	123
5.3 The ABOVE ² Campaign	127
5.4 The EPEX Campaign	133
5.5 The BARREL Campaign	135

5.5.1	Energy Selective Scattering	136
6	Conclusions	146
6.1	The Inverse Problem	146
6.2	Application to Experimental Data	150
6.3	Future Work	152
6.3.1	Simulations and Models	152
6.3.2	Experimental Data	154

List of Tables

List of Figures and Illustrations

1.1	This picture of the Aurora, taken from the International Space Station, shows that the space around Earth can have a lively and energetic character. Image by NASA, from nasa.gov.	3
1.2	Schematic illustration showing the scale and structure of the atmosphere, with the overall trend in electron density with height. Image by NASA Goddard, from nasa.gov.	8
1.3	Zero-pressure balloon in flight shortly after release. The lift gas will expand to fill the balloon volume during the ascent.	11
2.1	Model of the magnetic field of the Earth which includes effects due to external sources and the solar wind. Coordinates are measured in Earth radii and centered on the Earth. Near to the earth, a dipole field describes the geometry well. Farther away, this changes.	15
2.2	Cutaway view of AE9 model electron fluxes in the radiation belts, with some model magnetic field lines from Tsyganenko (1987) to show the overall geometry.	17
2.3	Frequency/time relation for a whistler-mode wave, observed by a Stanford University VLF radio receiver at Palmer Station, Antarctica.	24
3.1	X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 32 km.	31
3.2	X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 34 km.	32
3.3	X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 40 km.	33
3.4	GEANT4 architecture, reproduced from Pfeier et al. (2001). Class structure is in the vertical direction, with dependencies towards the bottom.	35
3.5	Evaluation of key MSIS model parameters as a function of altitude. Particularly important for the GEANT4 simulation are the density profiles of each atmospheric constituent (top).	37
3.6	Schematic diagram of information flow for a single GEANT4 simulation of energetic electron precipitation and the resulting X-ray radiation	40
3.7	Schematic diagram of information flow for a single GEANT4 simulation of energetic electron precipitation and the resulting X-ray radiation	41
3.8	Expected X-ray flux as a function of energy at a sample altitude of 34 kilometers using data from GEATN4 simulations, Berger and Seltzer (1972), and the BARREL SPEDAS model.	43
3.9	Polar angle of X-ray photon population at 34 kilometer altitude, as produced by precipitating electrons at different fixed energies. Units are normalized to account for the higher number of photons produced with greater beam energies.	45
3.10	Photon energy distribution for 1 MeV incident electron beams having isotropic and unidirectional angular distributions.	46

3.11	GEANT4 representation of the detector geometry. The lower plexiglass mount is shown in cyan, the aluminum body in grey, and the led collimator in purple. The height of the complete unit is approximately 30 cm.	47
3.12	49
3.13	Energy deposition events inside the detector crystal for beams of monochromatic photons at different set energies.	50
4.1	(Top): Example of three different assumed electron energy spectra (mono-energetic, exponentially decreasing, and Gaussian) for a precipitating electron flux of 1×10^9 electrons per cm^2 per second. (Bottom): Resulting simulated X-ray spectra for each electron spectrum, using the methods developed in Chapter 3.	54
4.2	Condition number of mapping \mathbf{G} as a function of X-ray spectrum and electron spectrum bin-widths in KeV. The white region is the domain where the condition number is infinite, and \mathbf{G} is singular.	56
4.3	Direct matrix inversion applied to simulated exponential electron spectra using bin widths of 10 (A), 50 (B), and 100 (C) KeV. The total incident electron count is 1×10^8	57
4.4	Example of over-fit and under-fit models to data. A random data set is generated and polynomial models are fit using least-squares regression. The reduced χ^2 statistic is shown for each model.	62
4.5	Plots of the singular values of the kernel matrix for the X-ray inversion problem for different electron spectrum bin widths, with singular values arranged from biggest to smallest.	64
4.6	Plots of the condition number of the kernel matrix for the X-ray inversion problem as a function of electron spectrum and X-ray spectrum bin width, for 1 (A) through 6 (F) truncated singular values.	65
4.7	Plots of simulated electron spectra for a total electron flux of 1×10^9 electrons/ cm^2 /second, and their reconstruction from the associated simulated X-ray spectra using a truncated singular value approximation of the kernel matrix. Electron bin widths are set to 50 KeV, and X-ray bin widths to 1 KeV. (A), (B), and (C) show reconstructions with the smallest 1, 5, and 10 singular values dropped from the expansion.	68
4.8	Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The X-ray spectrum corresponding to the assumed electron spectrum is shown in the top panel. The middle panel shows the assumed electron spectrum in blue, with retrieved spectra in grey (direct inversion) and green (Tikhonov regularization). The L-curve and selected value for α are shown in the bottom panel.	71
4.9	Random data set (blue) with under-fit (orange), and over-fit (green) representations. A particular data point (red) is removed prior to fitting. The removed data point is far from both the over-fit and under-fit representations of the data.	75
4.10	Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).	76

4.11 Application of second-order Tikhonov regularization to synthetic X-ray measurements corresponding to two mono-energetic electron beams at different X-ray fluxes and noise levels (A - 10^7 electrons / keV, B - 10^6 electrons / keV, C - 5×10^5 electrons / keV)	80
4.12 (left): Graph of the rows of the X-ray response matrix. A high degree of collinearity is apparent, as with a large scale change between rows. (right): Graph of the singular values of the X-ray response matrix, from largest to smallest.	84
4.13 (top): Contour plot of the condition number of the X-ray response matrix as a function of electron spectrum and X-ray spectrum bin widths, with no preconditioner. (bottom): The same contour plot with diagonal scaling preconditioner applied.	85
4.14 Response matrix for the X-ray inversion problem. Order of magnitude scale differences across both rows and columns of the matrix contribute to the high condition number. Units are normalized per incident electron.	90
4.15 Response matrix for the X-ray inversion problem with left and right preconditioners applied. Scale differences across both rows and columns of the matrix have been reduced, improving the condition number. The units on the color axis are no longer physical, having been scaled by the preconditioning matrices.	91
4.16 Response matrix condition numbers as a function of X-ray and electron spectrum bin widths for: (A) - no preconditioning, (B) - left scaling preconditioner only, (C) - right preconditioner only, (D) - both left and right preconditioners.	92
4.17 Retrieval of electron spectra from synthetic X-ray data across different simulated count rates. Left and right preconditioners, second-order regularization, and a non-negativity constraint are applied.	93
4.18 Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^9) from synthetic X-ray data.	94
4.19 Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^9) from synthetic X-ray data.	95
4.20 Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^9) from synthetic X-ray data.	96
4.21 Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^8) from synthetic X-ray data.	97
4.22 Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^8) from synthetic X-ray data.	98
4.23 Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^8) from synthetic X-ray data.	99
4.24 Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^7) from synthetic X-ray data.	100
4.25 Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^7) from synthetic X-ray data.	101
4.26 Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^7) from synthetic X-ray data.	102
4.27 Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^9) from synthetic X-ray data.	103
4.28 Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^9) from synthetic X-ray data.	104

4.29	Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^9) from synthetic X-ray data.	105
4.30	Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^8) from synthetic X-ray data.	106
4.31	Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^8) from synthetic X-ray data.	107
4.32	Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^8) from synthetic X-ray data.	108
4.33	Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^7) from synthetic X-ray data.	109
4.34	Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^7) from synthetic X-ray data.	110
4.35	Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^7) from synthetic X-ray data.	111
4.36	Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^9) from synthetic X-ray data.	112
4.37	Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^9) from synthetic X-ray data.	113
4.38	Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^9) from synthetic X-ray data.	114
4.39	Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^8) from synthetic X-ray data.	115
4.40	Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^8) from synthetic X-ray data.	116
4.41	Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^8) from synthetic X-ray data.	117
4.42	Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^7) from synthetic X-ray data.	118
4.43	Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^7) from synthetic X-ray data.	119
4.44	Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^7) from synthetic X-ray data.	120
4.45	Distributions of accumulated errors in the cross-validation process used to select the regularization parameter for inversions of 100 randomly generated electron spectra.j	121
5.1	X-ray spectrometer (white cylinder) encased in insulation foam and installed with flight systems in balloon payload. The DPU (digital processing unit) is contained in the metal rectangular enclosure on the left. The lead collimator assembly is hidden under the white styrofoam insulation.	125
5.2	Block diagram of X-ray scintillation detector system	126
5.3	ABOVE ² balloon being filled and prepared for launch in Saskatoon, Saskatchewan, August 2016. The payload and flight train are laid out on the ground, extending approximately 10 meters to the right. Image credit: Joel Kesler / RMD Engineering. 128	

5.4	ABOVE ² X-ray detector counts per second, summed across all energy channels, as a function of time since power-on of the system.	129
5.5	Background-subtracted X-ray energy spectrum of the precipitation event observed by ABOVE ² science flight 2. The integration time is 900 seconds, and the bin width is 1 keV.	131
5.6	Application of Tikhonov regularization to the background-subtracted precipitation event measured by ABOVE ² science flight 2. The algorithm applied is as described in Chapter 4. Second-order regularization with left and right preconditioning is applied together with a positivity constraint to generate the solution.	132
5.7	EPEX X-ray spectrometer data in spectrogram form. The precipitation event was observed at 07:00 UTC. The total counts per second are shown below the spectrogram as a function of time.	134
5.8	Background-subtracted X-ray event measured by the EPEX spectrometer. The model fit using Tikhonov regularization is shown overlaid.	139
5.9	Precipitating electron spectrum retrieved from EPEX X-ray data using the techniques of Chapter 4. The required value for α is low and well-defined through cross-validation.	140
5.10	Spectrogram and raw count rates of the event measured by BARREL and studied by Halford et al. (2015).	141
5.11	Background-subtracted X-ray event measured by the BARREL spectrometer. The model fit using Tikhonov regularization is shown overlaid.	142
5.12	Precipitating electron spectrum retrieved from BARREL X-ray data using the techniques of Chapter 4. The required value for α is approximately 5×10^2	143
5.13	Mono-energetic and exponential fits to the BARREL event from Halford et al. (2015), with the fit obtained by applying the techniques developed in Chapter 4 (preconditioned, constrained second-order Tikhonov regularization using cross-validation). The total electron flux for the regularized model is found to be 1.6×10^6 electrons/cm ² /sec.	144
5.14	600 second average differential electron energy flux measured by the RBSP-B MAGEIS instrument and used in the analysis by Halford et al. (2015).	145

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
E	Electric field
B	Magnetic field
P	Radiated power from a charge
q	Electric charge
m	Mass
β	Relativistic factor
c	Speed of light
keV	kilo electron volt
κ	Matrix condition number
\mathbf{G}	Response matrix mapping electron spectra to X-ray spectra
χ^2	Pearson statistic
σ	Matrix singular value
\mathbf{L}	Regularization operator
α	Regularization parameter
Ω	Solid angle
K	Degrees kelvin
PMT	Photomultiplier tube
BARREL	Balloon array for radiation belt relativistic electron losses
GEANT	Geometry and tracking software package
MSIS	Model atmosphere package
SPEDAS	Space Physics Environment Data Analysis Software
NMSE	Normalized mean squared error

Chapter 1

Introduction

1.1 Like Astronomy, but Closer

Space physics is a field of study that encompasses natural phenomena that occur in our own solar system. The basic difference between it, and the much older field of Astronomy is that in-situ measurements and direct experiments are possible. As a separate field of study, space physics began in earnest at the beginning of the space age with the flights of the first instrumented spacecraft. The discovery of the radiation belts that surround Earth was a profound example of the subsequently rapid increase in knowledge about the space surrounding Earth which is a trend that has continued to the present day. At the same time, and equally importantly, ground based and remote sensing techniques have advanced to the degree that robust, distributed, measurements of the properties of the space around Earth are readily available. The synthesis of these two orthogonal, but complimentary, sources of data gives the field of space physics a distinct image among the sciences, and motivates its pursuit not just for its own sake, but as a window to the workings of the wider universe which we cannot reach.

The space surrounding the Earth is full of activity. To get a picture of this, one can look at the Aurora Borealis, one of the dramatic and beautiful properties of this system. When they occur, the aurora aren't static features. Depending on the type of aurora, they can present as shimmering curtains, pulsing patches, or diffuse regions. These types are not exclusive. Dynamic structures like these, in space and time, are often indicative of a complex system. The old description - that the aurora are caused by the solar wind following the magnetic field of the Earth and hitting the atmosphere - falls flat in light of this complexity. In fact, the current state of knowledge paints a much different picture, with internal properties of the space and constituents surrounding Earth doing most of the "work" producing the aurora, which then carries much information about the

entire system.

There are two basic entities in near Earth space: fields and particles, which may be electrically charged. Classical electromagnetism provides a sufficiently accurate description of these in most cases. These are Maxwell's equations, in one of their most common forms:

$$\text{Gauss' Law} \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (1.1a)$$

$$\text{Gauss' Law (B Fields)} \quad \nabla \cdot \mathbf{B} = 0 \quad (1.1b)$$

$$\text{Faraday's Law} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.1c)$$

$$\text{Ampere's Law} \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (1.1d)$$

In these equations, \mathbf{E} and \mathbf{B} are the electric and magnetic fields, \mathbf{J} is electrical current, ρ is charge density, and μ_0 and ϵ_0 are constants. These have been known for a long time, but they contain the root of why the space around Earth is a complicated place. When you have charged particles, they move according to electromagnetic fields, but while they move, they also affect the fields themselves. In an environment like near Earth space, where matter mostly exists as a plasma of charged particles, this gives rise to collective, or self-consistent behaviour and is a source of complexity. Further, a simple analysis of Maxwell's equations shows that wave phenomena also occur. This is combined with the fluid character that particles in large numbers display to make what is called plasma physics.

Most of the matter in the space around Earth exists as a plasma. Plasma can be roughly defined as a gas of positive and negatively charged particles in approximately equal number, free to move, and at a high temperature. There is a more precise definition, which comes from statistical mechanics. In general, it is accurate to say that one of the defining characteristics of a plasma is the combination of the long-range electromagnetic wave interactions, with short-range collisional and fluid properties. Plasma waves, collective motion, and the propagation, absorption, and emission of radio waves are all observed in the space surrounding Earth. Based on all of this, the overall pic-

ture is not one of cold, dark, and empty space, but of a lively and energetic environment, constantly in motion and full of vibrations, with energy, matter, and information constantly flowing.



Figure 1.1: This picture of the Aurora, taken from the International Space Station, shows that the space around Earth can have a lively and energetic character. Image by NASA, from nasa.gov.

The active transport of particles through the space around Earth motivates questions about their dynamics. One of the most basic questions is how constituent particles enter and leave the system. Understanding this is central to forming an accurate picture of the overall system. Experimental measurements are required to achieve this. The fact that the Aurora occur is evidence that some fraction of the particles in near-Earth space eventually end up in the atmosphere, where they deposit energy and create visible light. The fact that we can directly observe this secondary radiation makes it a compelling target for experimental measurements. The question then becomes what information about the original particles can be extracted from their impact on the atmosphere. This type of question is not uncommon in the physical sciences. An analogous problem in geophysics is the reconstruction of seismic events from the vibrations caused at the surface of the Earth. This type of inverse problem naturally leads to a mismatch between the amount of data available and the amount required to generate unique solutions. In this thesis we address the problem of determining the population of electrons leaving geospace by impacting the atmosphere, using measurements of

the secondary radiation they produce. Rather than visible light, the secondary X-rays produced by the electrons decelerating in the atmosphere are used as the source of data.

The aim of this thesis is to develop the techniques necessary to obtain useful information from this indirect measurement. The combination of practical problems due to the need to obtain measurements at high altitudes of greater than 30 kilometers or so, and the mathematical problem of reconstructing useful models from incomplete secondary data lead to a challenging experiment. We show that a surprising amount of detail about the particles in geospace can be found despite these challenges.

1.2 The Earth's Atmosphere and Geospace

The experiments described in this thesis are conducted within the Earth's upper atmosphere, with detectors carried onboard stratospheric balloons. The processes that these experiments measure occur on the boundary between space and the atmosphere. An accurate description of this region is therefore needed to analyze data from the experiments.

At the surface of the Earth, the atmospheric constituents are well-mixed. Most of the mass of the atmosphere lies in the region defined as the troposphere, which extends from ground level to an altitude of around 10 kilometers. This dense region of the atmosphere is where life on Earth, and most terrestrial weather happens. The main structure of the troposphere is defined by a decrease in temperature and density with increasing height. Density decreases exponentially with altitude, which can be shown from the hydrostatic condition. Temperature is somewhat more complicated. Besides local effects (weather), temperature decreasing with height owing to the decreased gravitational potential. The lapse rate defines the rate at which temperature drops with height, and has a value of approximately 2° C per kilometer of altitude.

At the tropopause, the temperature of the atmosphere begins to increase with height. The density here is low enough that ultraviolet radiation penetrates and causes the disassociation of molecular oxygen, which results in free oxygen atoms combining with diatomic oxygen to form

ozone. The stratosphere is characterized by limited vertical mixing, and is very dry, with few weather effects from the troposphere reaching it. Winds in the stratosphere can be intense, reaching speeds of hundreds of 200 kilometers per hour. This presents interesting operational challenges to experiments carried on balloons in this region, particularly with regards to their post-flight recovery.

The density in the stratosphere is low, ranging from a few percent of that at ground level to nearly zero at the vertical limit. The low density in the stratosphere makes it quite transparent to certain kinds of radiation from space. In particular, X-rays, created from the slowing down of charged particles farther up in the atmosphere, can propagate deeply in this region. This is where remote sensing of space effects based on secondary radiation becomes possible using balloon-borne instruments, which are discussed in the following section.

At the end of the stratosphere, temperature once again begins to decrease with height. This occurs at an altitude of approximately 50 kilometers above the surface of the Earth. The mesosphere is the name given to this region of the atmosphere. Here, the different components of the atmosphere are still well-mixed owing to turbulence. This continues to be true to a height of around 80 kilometers. Meteorites burn up in this part of the atmosphere, and the temperature reaches a lower limit of approximately -100 degrees C at its highest point. Once the turbulent effects which occur in the mesosphere become less important, separation of the different atmospheric constituents occurs due to the different scales at which their densities decrease. The hydrostatic condition imposes exponential decreases in density with height, but the scale at which this happens depends on the particle mass. This gives rise to a structure in the upper atmosphere, with the proportions of the different components changes with height. This is important for modelling the propagation of radiation, which has interactions which depend on the chemical composition, overall density, and temperature.

Past the mesosphere lies a region called the thermosphere. In this region, temperatures climb sharply and then are steady with increasing height. Solar activity has a strong bearing on the

temperatures in this region, which can reach hundreds or thousands of degrees C. Turbulant mixing no longer has meaning at the low densities in the thermosphere. The international definition of the beginning of space, a boundary that begins at 100 kilometers altitude, is contained in this region. In the thermosphere, the density begins to become sufficiently low that the rate of collisions between individual atoms and molecules becomes small enough to start treating them individually, rather than as a fluid. This is the domain where plasmas can exist, since ionized particles and electrons do not immediately collide and become neutral once they are created. This region of space overlaps with the ionosphere, for this reason. Spacecraft orbit in the thermosphere; orbits with a practical decay time begin at heights of around 350 kilometers.

At heights of hundreds of kilometers, charged particles survive for long enough that electromagnetic interactions become important. Temperature takes on a different intuitive meaning, one based more on the statistical properties of individual and group particle motion, than the classical value that can be measured with a thermometer. The local magnetic field becomes the most important geometry and coordinate system, rather than the local vertical direction from Earth. Free from the dominating effect of collisions, particles undergo characteristic motions not seen in the lower regions of the atmosphere, which are determined by electric and magnetic fields. Figure ?? shows an illustration of some the different layers of the atmosphere, and how the population of charged particles changes with height.

It's not easy to draw a clear sharp line where the atmosphere ends. Such a definition probably needs to depend on the application or line of inquiry being pursued. For aircraft, heights past the stratosphere are unreachable owing to low density. Rockets can reach the intermediate heights between the stratosphere and the thermosphere, while orbiting satellites require further heights still, so that orbits decay on a sufficiently large timescale. Eventually, even the magnetic field of the Earth becomes less important than the magnetic field carried with the plasma of the solar wind, and farther past this point, the influence of the Earth and its magnetic field ceases. The exploration of the solar system, and beyond, by robotic spacecraft begins to blur the distinction between space

physics and astronomy. It is interesting, that the two fields are separated mostly by a practical limitation due to the vast distances involved, instead of a more fundamental, and natural boundary.

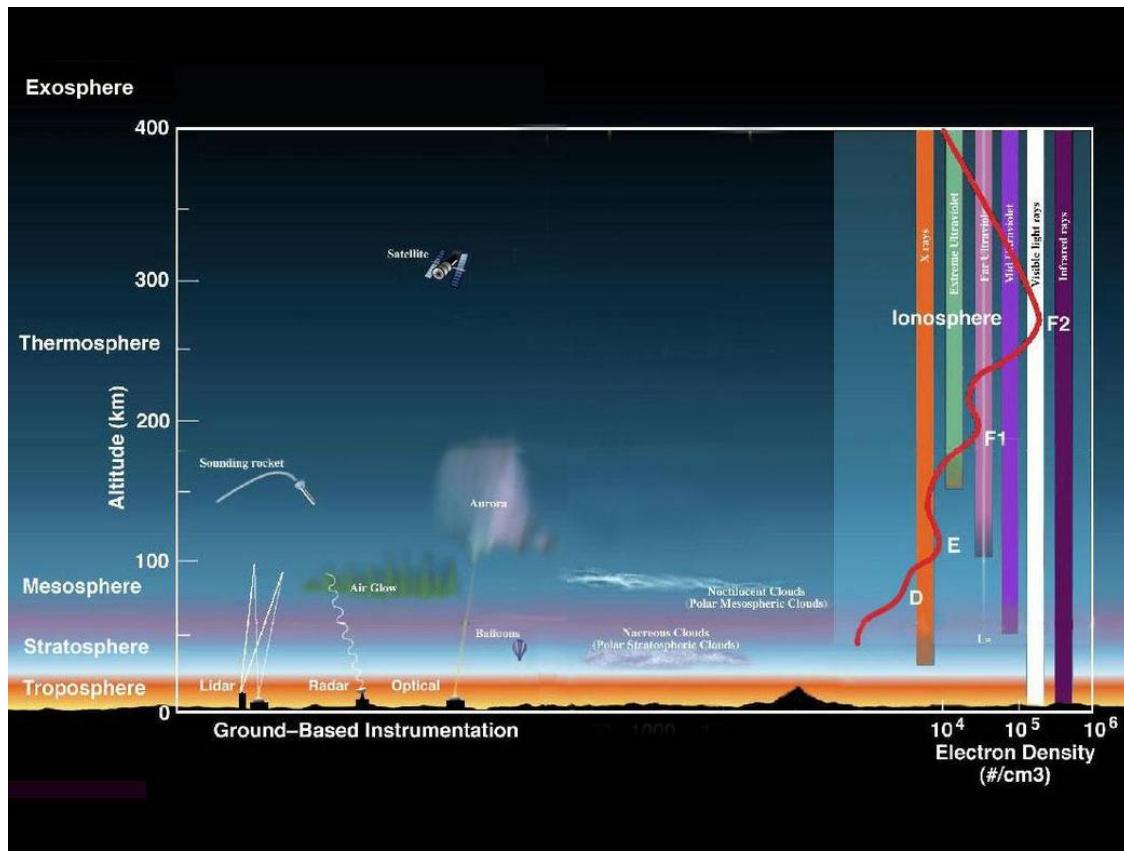


Figure 1.2: Schematic illustration showing the scale and structure of the atmosphere, with the overall trend in electron density with height. Image by NASA Goddard, from nasa.gov.

1.3 Getting Above the Atmosphere with High Altitude Balloons

Since the atmosphere shields most of the radiation from space, it is not possible to build ground-based experiments that are sensitive to X-ray and charged particle radiation. Rocket flights routinely reach heights where this is no longer a problem, but by their nature, the flight times are short, on the scale of minutes. Orbiting satellites are practical at altitudes far above where radiation from space deposits most of its energy, so there needs to be another kind of vehicle, to get instruments above most of the atmosphere by mass, and that can remain there for hours, or days. Balloons provide data which are distinct from satellites. In the types of balloon experiment conducted in this thesis, radiation detectors are designed with a field of view that extends above them and covers a circle in the sky. The detectors measure secondary radiation produced by energetic particles which enter the atmosphere. On the other hand, satellites do measurements along the path of their orbit, and see particles that have not entered the atmosphere. Additionally, balloons move slowly compared to satellites, and can therefore accomplish measurements of roughly the same region over extended periods.

Balloons are the vehicles used in this thesis to get above the shielding effects of the lower atmosphere, and measure radiation from space directly at an altitude of 30 to 40 kilometers. High altitude balloons are large, thin, membrane structures, filled with either hydrogen or helium, which carry payloads ranging from grams to hundreds of kilograms, and ascend until they reach an equilibrium height in the stratosphere, above most of the radiation shielding effects of the atmosphere. Balloons need no external energy source to maintain their flight, and so can remain at altitude for hours, days, or even weeks depending on their design and the desired mission profile. With the exception of the balloons needed to carry the largest, heaviest, experiments, high altitude balloons are also inexpensive, compared with rocket flights and satellites. This opens possibilities for extensive surveys and coincident missions which do not exist when using other platforms.

High altitude balloons follow two distinct designs. The zero-pressure balloon is named because it maintains an internal pressure equal to the surrounding atmosphere. This is the type of balloon

most frequently used for scientific experiments, because it has the property of passively remaining at peak altitude for extended periods of time. The design consists of a large polyethelyne bubble, with one or two vent ducts near the bottom, which are open to the atmosphere. The payload hangs beneath the balloon on a cable or rope. As the balloon is filled through the ducts, lift gas collects at the apex and causes hydrostatic lift. Once a predetermined amount of lift gas has been added, the balloon is released, and ascends, carrying the payload with it. As the balloon ascends, the internal lift gas expands to fill the balloon volume. Eventually, the balloon is completely filled with the low density lift gas, and excess is passively released through the vent ducts, causing the balloon to stop ascending and maintain a fixed altitude. When the balloon mission is terminated, cable cutting equipment is used to separate the payload from the balloon. The payload descends under parachute and is usually recovered. Upon release of the payload, it is typical that the payload rips a destruct panel from the balloon, causing it to release most of its lift gas and descend as well. The point at which the termination process happens is chosen in the design phase of the mission, and can be activated by timer, remote command, or both.

There is an ultimate limit to the flight duration which can be achieved using zero-pressure balloons, which is caused by internal temperature differences at day and night. During the day, solar radiation warms the balloon and the gas contained in it. When night occurs, the balloon begins to cool, which causes the gas inside to decrease in volume. The balloon has no rigid structure to maintain a fixed volume, so the balloon membrane then occupies less volume as well. This causes the balloon to start to descend, where it reaches a new equilibrium point. At sunrise, solar heating causes the balloon to rise again, although it will not reach the same altitude it started with. For long duration scientific experiments, this cycle is controlled through the release of ballast, usually sand. It can be shown that the ultimate limit for the duration of zero pressure balloon flights is approximately 7 days, even if most of the payload consists of expendable ballast. The exception to this rule occurs for flights in the polar regions of the Earth, where the day night cycle is longer. Figure 1.3 shows a zero-pressure balloon in flight, with the scientific payload being carried below.



Figure 1.3: Zero-pressure balloon in flight shortly after release. The lift gas will expand to fill the balloon volume during the ascent.

The second type of high altitude balloon is primarily used for meteorological experiments. Consisting of a relatively small latex sphere, this balloon is closed from the atmosphere and rises until the internal pressure causes it to rupture and descend. These meteorological balloons are launched regularly around the world several times per day to aid in the creation of weather models and forecasts. Typically carrying payloads of only a few hundreds of grams, to at most a few kilograms, these balloons reach heights comparable to the zero-pressure balloons, but do not maintain equilibrium altitude. This makes them less useful for the measurement of space radiation, although, some exotic ways to overcome this problem have been developed and tested.

1.4 Scope of this Thesis

The question is: what is the maximum amount of information that can be obtained about processes far out in the space surrounding Earth, using measurements of the resulting radiation, and what kinds of instruments do we need to use to get the requisite data? This question is interesting because the experiments take place on the boundary between what we can directly affect using active experiments, and what we only learn through passive observation. Right now, most experimentalists are earth-bound, and this is unlikely to change in the near future. On the other hand, the types of experiments we fly on high altitude balloons provide a window to the wider universe, in a way that, unlike orbiting spacecraft, which require years or decades of planning, can be rapidly iterated on and repeated. The upper atmosphere can be thought of as a giant laboratory, and a detector which is sensitive to the radiation incident upon it from space. We have the ability to send experiments to this region rapidly and at low comparative cost to spacecraft. There is therefore a scientific return on being able to maximize the information we obtain from these experiments. The problem of mapping the available measurements of secondary radiation back to the particles impacting the atmosphere is required for this to happen. The following chapters will show the development of a solution which can be used to increase the observational power of balloon-borne experiments.

Chapter 2

Energetic Electron Precipitation and Geospace

2.1 Introduction to Connected Geospace

The Earth is a complicated place. Natural processes, which occur on vastly different time scales, are all connected to each other and produce the state of the overall system. Only in the last 70 years or so, since the beginning of the space age, has the picture of the Earth as a dynamic system been extended to include the space around it. This space is not empty, and it has structure and dynamics which are very different than processes on Earth, mostly owing to the importance of the electromagnetic force between its constituents. The word used to describe the space around Earth is geospace, so-named because the Earth and its processes are not strictly separated from the space around it. In fact, information, matter, and energy flow between regions which can be separated by great distances, and cause effects which can echo throughout the entire system. In this chapter, I will describe one of the most dramatic ways which this can happen, where energetic particles from space enter the atmosphere and affect it directly.

Energetic electron precipitation is a process by which electrons depart trapped states in geospace and enter the Earth's atmosphere. One of the most brilliant effects that can be caused by this process is able to be seen from the ground on Earth - the northern lights, or Aurora Borealis. There has been a great deal of work done in the last half-century to understand, quantify, and model the processes responsible for creating the Aurora, but many unanswered questions remain. With the knowledge that the Earth and the space around it are both connected and affect each other, it becomes of great scientific interest to understand the ways which particles and energy move between them. The fact that energetic particle precipitation can have dramatic visual affects on the atmosphere motivates its use as measurable target which can provide a window to the dynamics of the connected system.

The earth and geospace are not stationary systems. One can draw a loose analogy between

their related structure and the anatomy of a living organism. While the system is always in a state of change, there is a consistent structure and organization at the highest levels. In this case, that structure is largely defined by the magnetic field of the Earth, and the magnetic field of the solar wind which interacts with it. Navigation on the surface of the Earth can be accomplished by measuring the direction of the local magnetic field in two dimensions. In geospace, the magnetic field changes significantly over all three dimensions, and, to an extent, in time as well. As charged particles, electrons in geospace are subject to the Lorentz force. This force is always orthogonal to the magnetic field, which is why the magnetic field is such a powerful organizational force in the system.

The magnetic field of the Earth is due to both internal, and external (space-based) current sources. This field can be conveniently described by the multipole expansion. Figure 2.1 shows a model of the magnetic field of the Earth, which takes into account both internal and external sources. Towards the surface, the dipole is a good approximation. Farther away, however, higher-order terms become important and result in a field has a vastly different structure than the dipole description. The magnetic field geometry in Figure 2.1 does not imply a steady-state to the system. Rather, this can be thought of as the overall topology of a system which is always in motion. In reality, only some modified version of this picture will be accurate at any particular time. Particularly, at distances which exceed a few Earth radii, the state of the solar wind, through which the Earth is always moving, becomes more important than the current sources internal to the Earth itself. A more accurate picture of the system than Figure 2.1 would actually be an animation, with magnetic field lines cascading on the dipole field of the Earth, and reconnecting to form the stretched-out “tail” from 10 to hundreds of Earth radii “downwind”. The description of that process is an entire field of study, and it too, has questions which are currently unanswered.

Given that electrons in geospace exist, and they are subject to a magnetic field geometry which is complicated and dynamic on the relevant scales, this suggests that the motion of electrons within the system is important to the state of the overall system. This, then, implies that the loss of these

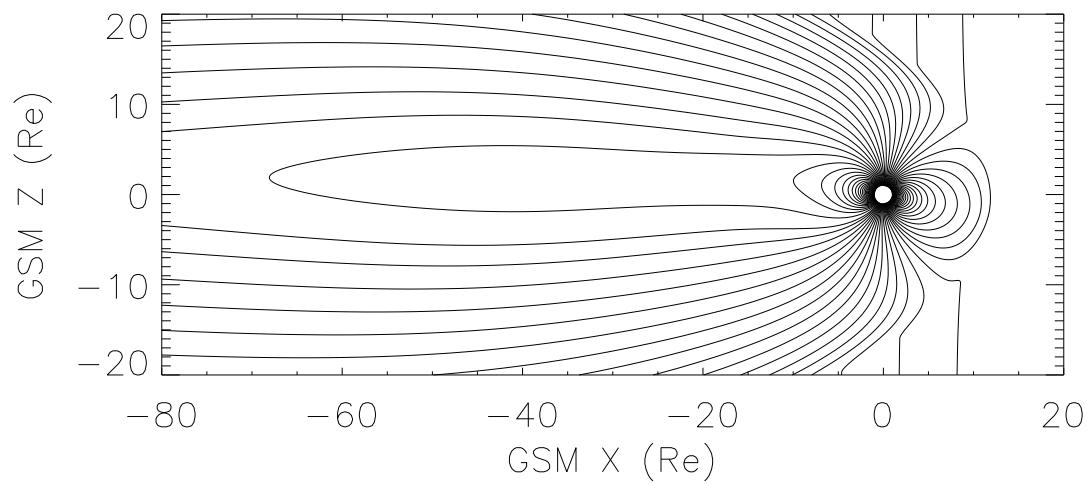


Figure 2.1: Model of the magnetic field of the Earth which includes effects due to external sources and the solar wind. Coordinates are measured in Earth radii and centered on the Earth. Near to the earth, a dipole field describes the geometry well. Farther away, this changes.

particles to the Earth's atmosphere, is also important. The goal of this chapter is to describe, in adequate detail, the connection between the precipitation of electrons into the atmosphere and the state of the larger geophysical system. The goal is to advance electron precipitation, which is a fairly dramatic event, as a process which is both worth understanding in its own right, but also as a tool, which provides a window in to the processes which occur in the larger system.

2.2 Electrons in Geospace

The first American artificial satellite, Explorer 1, was launched in 1958. Among other instruments, it carried a Geiger counter, which is a device sensitive to the energy deposition of charged particles like electrons. The expected background count rate due to cosmic rays was observed at some points during the flight, but at others, zero counts were reported. This was determined on later flights, to be due to the saturation of the device by a much larger than expected radiation flux. This led to the discovery of the radiation belts around the Earth. The question then became why the radiation belts existed, and what their properties are. A review of the historical development of knowledge about the radiation belts is Millan and Baker (2012).

The radiation belts have a spatial structure. This is shown in Figure 2.2, which is an evaluation of a model, AE9, generated from measurements from many different satellites (see Gin et al. (2013)). An approximate description of the Earth's magnetic field (Tsyganenko, 1987) is overlaid to show that it is the main geometry at play. There are three essential regions. The inner zone, and outer zone are distinct and separated by a region of empty space, called the slot region.

The main process which is thought to be responsible, at least partially, for the existence of the inner radiation belt is called cosmic ray albedo neutron decay (CRAND) (Singer, 1958). In the CRAND process, cosmic rays from space create neutrons in the upper atmosphere, which are then free to travel upwards, or downwards. Those which travel upwards, and subsequently decay into a proton and an electron, form the inner radiation belt. The gap that defines the separation of the inner and outer radiation belts is called the slot region, and is caused by a kind of radio wave called

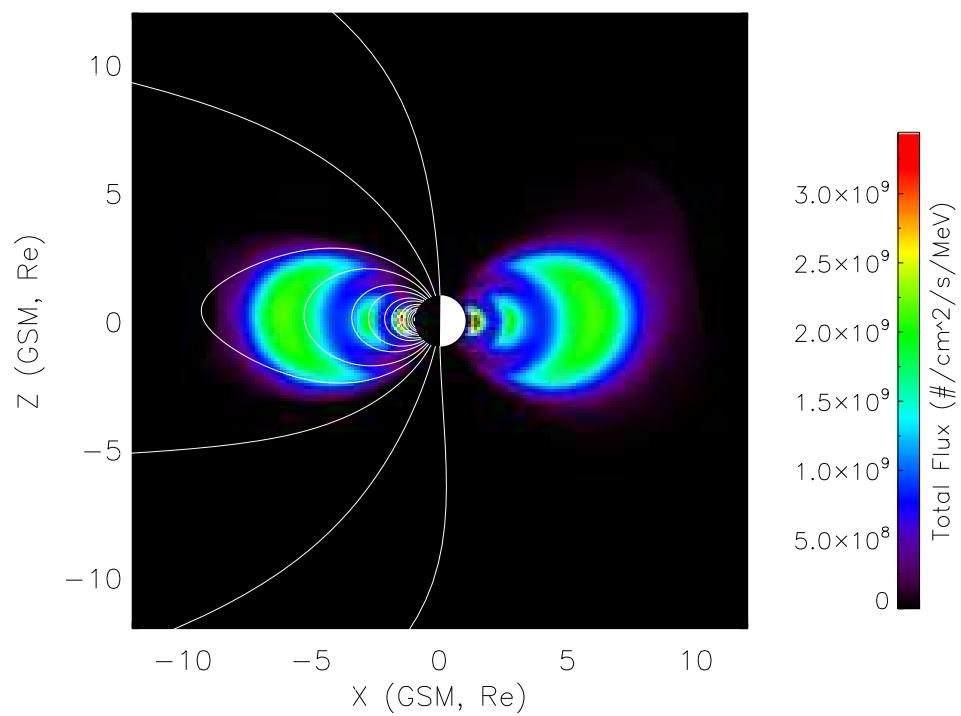


Figure 2.2: Cutaway view of AE9 model electron fluxes in the radiation belts, with some model magnetic field lines from Tsyganenko (1987) to show the overall geometry.

plasmaspheric hiss. As the radio wave interacts with particles trapped in the slot region it changes their motion in such a way that they precipitate and are lost in the Earth's atmosphere. This process can be partially understood through the adiabatic description of trapped particle motion, which is discussed in the next section. Even though steady-state and adiabatic expressions of the process by which the slot region is formed make predictions that agree with experimental data, they do not form a complete description.

2.3 Single Particle Motion

Some properties of the radiation belts, and electron precipitation, can be understood through a picture of their dynamics which considers particles independently. In this description, particle motions do not affect each other and we ignore the fact that a charged particle moving in an electromagnetic field also creates an electromagnetic field due to its own motion. This picture neglects the collective effects characteristic of space plasmas.

Electrons from the radiation belts move under the influence of the Earth's magnetic field. They are subject to the Lorentz force, $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, where q is the electron charge, \mathbf{v} is the velocity vector, and \mathbf{B} is the static magnetic field. The Lorentz force causes cycloidal motion about lines of constant magnetic force. The velocity vector of the electrons can be expressed in components which are parallel and perpendicular to the local magnetic field. The parallel component of the velocity vector will be constant provided the fields \mathbf{E} and \mathbf{B} are constant. Only the electric field does work and changes the kinetic energy of the electrons. In this case the particle motion is a helix, with a pitch angle defined by:

$$\alpha = \tan^{-1} \left(\frac{v_{\perp}}{v_{\parallel}} \right).$$

Particle motion in geospace can be described by adiabatic invariants, which are quantities that remain constant under slow changes to the system relative to the periodic motion of the particles. The action integral:

$$J = \oint pdq$$

is a constant of motion under these circumstances, where p and q are the generalized momentum and coordinate from Hamiltonian mechanics. It can be shown that the magnetic moment of an electron undergoing gyro motion is an adiabatic invariant. It follows that the magnetic flux encircled by the gyrating electron is also invariant. This quantity:

$$\Phi = \frac{2\pi m}{q^2} \mu$$

is called the first adiabatic invariant, where q is the electron charge, m is the electron mass, and μ is the magnetic flux enclosed by the electron gyro radius. The invariance of Φ implies that when an electron moves to a region of increased magnetic field strength, the gyro radius will decrease to encircle a constant magnetic flux. For electrons in near Earth space, this means that the electron gyro radius decreases with proximity to the poles. This has the effect of creating a magnetic mirror, since the only parameter that can change in the magnetic moment:

$$\mu = \frac{mv^2 \sin^2(\alpha)}{2B}$$

is the pitch angle α given that kinetic energy is constant. The pitch angle α changes between values α_1, α_2 for magnetic field strengths B_1, B_2 according to:

$$\frac{\sin^2 \alpha_1}{\sin^2 \alpha_2} = \frac{B_2}{B_1}.$$

When an electron drifts towards a region of higher magnetic field strength, there will be a point where the pitch angle reaches 90° and the parallel component of the velocity is zero. The force caused by the gradient in the magnetic field then accelerates the electron in the reverse direction. The point where this happens is called the mirror point. In the absence of other forces, electrons in near-Earth space undergo bounce motion between conjugate mirror points at the poles. When the mirror point occurs at a height inside the atmosphere, the electron will be absorbed and lost.

The minimum magnetic field strength for an electron undergoing bounce motion occurs at the magnetic equator. Electrons with a pitch angle of nearly 90° at the equator cannot move farther towards the poles, while electrons with pitch angles near 0° can move unimpeded by the mirror force. The set of pitch angles at the equator which result in mirror points within the atmosphere are called the loss cone. Electrons with equatorial pitch angles within the loss cone will not undergo subsequent bounce motion and will precipitate into the atmosphere due to collisions with its neutral constituents.

2.4 Scattering Processes

Interactions with electromagnetic waves can change the pitch angles of electrons so that they lie within the loss cone and subsequently precipitate. When a wave phenomenon has a doppler-shifted frequency ω which is resonant with the relativistic electron gyro frequency, energy can be exchanged in a resonant interaction. This condition is written as:

$$\omega - k_{\parallel} v_{\parallel} = \frac{n\Omega_e}{\gamma}$$

where k_{\parallel} and v_{\parallel} are the parallel components of the wave vector and velocity vector, n is an integer multiple, γ is the relativistic factor, and Ω_e is the electron cyclotron frequency. The resonant condition depends on the electron kinetic energy, and so different electromagnetic waves will resonantly interact with electrons that have a compatible energy range. This fact connects wave particle interactions in space to the population of electrons which precipitate into the atmosphere. Information about the scattering processes is carried by the precipitating electrons to the atmosphere. There are three main types of wave which are responsible for causing pitch angle scattering and electron precipitation. The first two we will discuss, plasmaspheric hiss, and whistler chorus, are different manifestations of the same type of wave propagation, called whistler-mode.

Whistler-mode waves are right-hand circularly polarized waves which occur below the electron cyclotron frequency, f_{ce} . The dispersion relation for whistler mode propagation determines how

the frequency ω , is related to the wavenumber k , of the whistler mode waves. The derivation is accomplished using the theory of radio propagation through cold plasmas and several approximations, and results in the relation:

$$\frac{c^2 k^2}{\omega^2} = 1 - \frac{\omega_{pe}^2}{\omega(\omega - |\Omega_e|)}$$

where c is the speed of light, k and ω are the wave number and frequency of the radio wave, ω_{pe} is the electron plasma frequency, and Ω_e is the electron gyro frequency. The plasma frequency ω_{pe} is given by:

$$\omega_{pe} = \sqrt{Ne^2/\epsilon_0 m}$$

where N is the electron number density, e is the electron charge, and m is the electron mass. The plasma frequency can be thought of as a natural frequency which the positive and negatively charged components of the plasma would oscillate at if there were separated by some small distance and released. A spectrogram showing how a typical whistler wave changes frequency in time is contained in Figure 2.4, reproduced from data by the Stanford University VLF group. The radio receiver used to record these data were located at Palmer Station, Antarctica.

A key driver of electron precipitation is plasmaspheric hiss, which is a particular kind of whistler-mode wave that is responsible for the formation of the slot region in the radiation belts (Lyons et al., 1972; Lyons and Thorne, 1973). The frequency range is broadband ELF, occupying 100 Hz to a few KHz (Millan et al., 2007). The generation mechanism is not fully understood. Theoretical modelling is consistent with generation by the electron cyclotron instability (Millan et al., 2007; Huang and Goertz, 1983; Church and Thorne, 1983; Meredith et al., 2004). The electron cyclotron instability, described by Thorne et al. (1979), was known in plasma physics before plasmaspheric hiss was discovered (Kennel and Petschek, 1966), and is a way that the local plasma population can create hiss. At the time of writing, it is thought that there are several distinct mechanisms which combine to produce plasmaspheric hiss in different situations (Thorne et al., 2010; Bortnik et al.,

2008; Albert and Bortnik, 2009).. Hiss below frequencies of 150 Hz is thought to be generated by amplification of whistler-mode noise due to substorm-injected electrons (Meredith et al., 2018). At higher frequencies, up to approximately 2 Khz, hiss likely comes from short bursts of chorus emissions (discussed next), which exist farther out in space, propagate in, and become trapped in the plasmasphere (Meredith et al., 2018). Finally, at higher frequencies, hiss becomes more associated with land mass and geographic location, which suggests an association with terrestrial lightning.

Another important wave phenomenon which causes precipitation is whistler mode chorus. Chorus is a wave phenomenon consisting of discrete whistler-mode emissions, which are observed outside the plasmasphere, with a frequency range of 100 Hz to 5 Khz (Millan et al., 2007). The waves occur in two distinct bands, one below, and one above, half the electron gyro frequency (Tsurutani and Smith, 1974; Thorne et al., 2010). It is believed that chorus is generated by the electron-cyclotron instability from injected plasma sheet electrons near the equator (Kennel and Petschek, 1966; Millan et al., 2007). Chorus waves cause a phenomenon called microburst precipitation (Millan et al., 2007). Microburst precipitation is characterized by short, quasi-periodic bursts of electron precipitation that occur on timescales of 10 milliseconds. Microbursts were first discovered at relatively low energies (approximately 100 KeV), but they do occur at relativistic, MeV energies as well. The relation between microburst precipitation and chorus waves is suggested by the fact that they both occur between 0300 and 1500 magnetic local time (Lorentzen et al., 2001; Millan et al., 2007), and spacecraft measurements of both phenomena showed a strong correlation. An investigation by Thorne et al. (2005) used data from a 1998 geomagnetic storm event to estimate the microburst precipitation rate due to whistler mode chorus waves and found general agreement between between observed, and model, loss rates at wave amplitudes consistent with observations Millan et al. (2007).

Chorus waves are an important loss mechanism for the radiation belts. Pitch angle scattering caused by chorus waves occurs over a broad range of energies Thorne et al. (2010), from tens of keV to MeV. There are limits to the applicability of the quasi-linear description of wave-particle

interactions between electrons and chorus waves, which are relevant, because very large amplitude chorus waves have been observed, with electric field strengths of more than 240mV/m (Cattell et al., 2008; Thorne et al., 2010). These cases, in particular, highlight the need for experimental measurements as constraints for implemented models.

A different electromagnetic wave that causes precipitation is the Electron Ion Cyclotron (EMIC) Wave. EMIC waves are a distinct type of electromagnetic wave, which are left-hand circularly polarized with frequencies below the proton gyro frequency. EMIC waves occur in bands separated by multiples of the ion gyrofrequency (Thorne et al., 2010). Different ion species give rise to different frequencies for EMIC waves. These waves are thought to originate near the equator (Fraser et al., 1996; Loto’aniu et al., 2005; Millan et al., 2007), where they are created by a temperature anisotropy in the ring current with respect to the local magnetic field (Jordanova et al., 2001). EMIC waves are known to be a major source of loss for electrons in the radiation belts. At energies of greater than approximately several 100 keV, they are effective at scattering electrons into the loss cone Thorne et al. (1981); Millan et al. (2007). The resonance condition requires that the wave, which is left-hand circularly polarized, be seen in the electron frame as right-hand polarized. This happens when the electron overtakes the wave, and so requires the electrons to have a minimum energy Millan et al. (2007). This minimum energy required can be calculated from the dispersion relation of the EMIC wave.

Satellite observations suggest that, most of the time, precipitation caused by EMIC waves occurs at higher energies (Millan et al., 2007). This can be seen as a result of the fact that owing to their left-handed polarization, an electron scattered by an EMIC wave needs to have a kinetic energy high enough that it overtakes the wave and sees it as right-handed polarized in its reference frame. The survey conducted by (Meredith et al., 2003) using measurements of EMIC waves by the CERES satellite found that the resonant energy required for electron scattering was less than 2 MeV in only approximately 11 percent of cases. The first balloon with a detector capable of measuring these higher energies was flown from 1996 in Kiruna, Sweden (Millan et al., 2007). This

balloon measured electron precipitation with a peak characteristic energy of 1.7 MeV (Foat et al., 1998). The observed precipitation was later shown by (Lorentzen et al., 2000) to be associated with nearby spacecraft observations of EMIC waves Millan et al. (2007).

The different types of wave that result in electron scattering and precipitation are energy selective, applying to populations of electrons with distinct energy ranges. This implies that information about the scattering processes can be found by measuring the properties of electrons which precipitate to the atmosphere. The focus of this thesis is on experimental techniques that can be used to accomplish this.

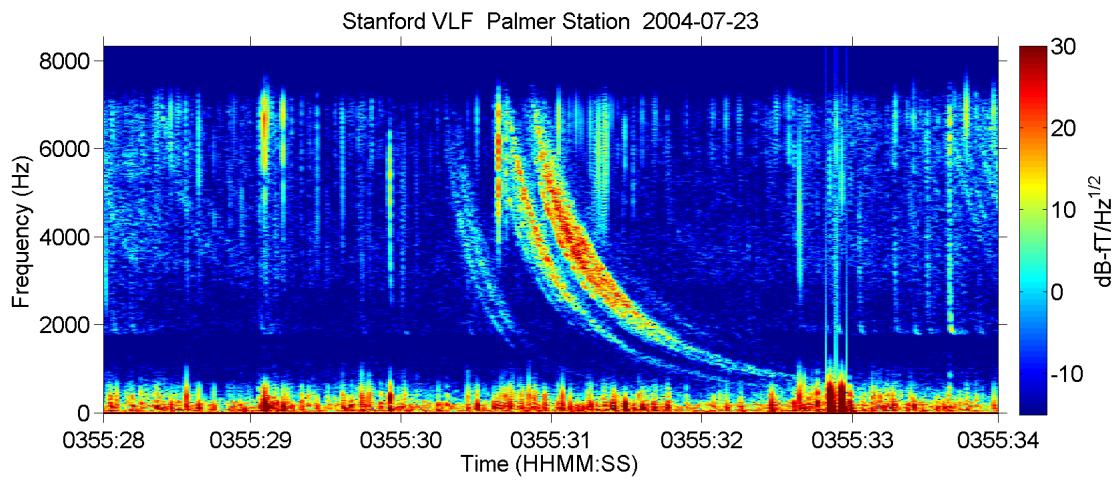


Figure 2.3: Frequency/time relation for a whistler-mode wave, observed by a Stanford University VLF radio receiver at Palmer Station, Antarctica.

Chapter 3

Computer Modelling and the Forward Problem

3.1 Problem Statement

The goal is to obtain measurements of energetic electron precipitation spectra and fluxes. Direct measurements of the electrons are not always practical, except for instrumented sounding rockets, since the precipitating electrons deposit their energy in a region of the atmosphere (approximately 80 to 100 kilometers) that is difficult to reach for durations longer than a few minutes (Berger and Seltzer, 1972). Satellites in low Earth orbit can accomplish measurements of precipitation, but only along their relatively short passes through the region of interest.

The X-ray photons emitted by the precipitating electrons as they decelerate penetrate deeply into the atmosphere and are detectable at altitudes above approximately 30 kilometers from the ground. This region of the atmosphere is accessible by stratospheric balloons, capable of flights which can extend for days or weeks, which makes them an attractive platform for longer term measurements. In this chapter, a model is created which predicts the photon measurements a high altitude balloon would observe for a given electron precipitation event.

This is not a new problem, measurements of X-ray spectra from high altitude balloons have an extensive heritage. The purpose of this chapter is to review the models which already exist and then to develop a similar model, tailored to the instrumentation discussed in Chapter 3. There will be an emphasis on computation speed, so that the model can be evaluated over large parameter spaces quickly on ordinary hardware. This will be particularly important when the inverse problem - attempting to reconstruct electron spectra from X-ray measurements - is attempted in Chapter 5. Additionally, the model we create will be designed to separate dependent variables as much as possible, so that it can be applied to different experiments and different instruments with few modifications.

3.2 Scattering and Energy Deposition Physics

For the problem at hand, classical physics (with relativistic effects) offers a sufficient description of the energy deposition and radiation caused by electrons as they travel through the atmosphere. Electrons from space enter the upper atmosphere at relativistic energies, ranging from less than 100 keV to several MeV. As the electrons penetrate the atmosphere, they interact with the neutral constituents of the atmosphere. These interactions are elastic and inelastic scattering.

Elastic (Rutherford) scattering was discovered through an experiment consisting of a beam of alpha particles incident on a sheet of thin gold foil. This lead to the discovery of the atomic nucleus by observing that few alpha particles were deflected, and when this happened, they were mostly deflected through large angles, implying that most of the atomic mass was concentrated in a small region in the centre of each atom. In this interaction, neither the nucleus or the scattered alpha particle is excited or changes state, implying that both kinetic energy and momentum are conserved. For electrons scattering from nuclei at keV to MeV energies, relativistic effects need to be taken into account. The name Mott scattering is used in this case.

When inelastic scattering occurs between an electron and an atomic nucleus, kinetic energy which is lost by the electron is emitted as a photon. The term used for this process is Bremsstrahlung, a german word for the term “braking radiation”. The Larmor formula, and its relativistic generalization are the well-known descriptors of the radiation emitted by this process. The power radiated by an accelerating charged particle is given classically as:

$$P = \frac{2}{3} \frac{q^2}{m^2 c^3} |\dot{\mathbf{p}}|^2$$

where q is the particle charge, m is its mass, and $\dot{\mathbf{p}}$ is the time derivative of its momentum. The relativistic generalization of this formula is:

$$P = \frac{2q^2\gamma^6}{3c} (\dot{\beta}^2 - (\beta \times \dot{\beta})^2)$$

where β is the velocity as a fraction of the speed of light, v/c and γ is the relativistic factor.

There are also expressions for the angular distribution of the radiated power.

Interaction probabilities for electron scattering are a function of energy, and are characterized by a quantity called the cross section. The definition is constructed by considering a particle incident on a target, which is scattered through an angle θ . Define the impact parameter b as the distance of closest approach to the target. Define the area element $d\sigma = bd\phi db$, in the plane of the impact parameter. The differential cross section is the ratio $\frac{d\sigma}{d\Omega}$, where $d\Omega$ is the differential solid angle through which the incident particle is scattered.

Cross sections can be experimentally measured, and also derived from theory. For Bremsstrahlung photons emitted by inelastic scattering from atomic nuclei, the electron cloud of the atom complicates the resulting expressions. Usually the process is complex enough that specific assumptions need to be made for derivations which apply in different energy ranges. For the problem of Bremsstrahlung emission by energetic electrons in the atmosphere, there is a synthesis of theoretical and experimental measurements which can be used to obtain electron energy loss as a function of depth in the atmosphere (Berger and Seltzer, 1972). Once a Bremsstrahlung photon is produced, it then propagates through the atmosphere and undergoes its own interactions. Photoelectric absorption, multiple scattering, and Compton scattering all have to be taken into account to predict the paths that photons will take in the atmosphere (Berger and Seltzer, 1972). This complicates the analysis of the causative electron spectrum significantly. In the photoelectric process, a photon impacts a neutral atom and causes an electron to be emitted. This process is dominant at lower energies (less than 511 keV approximately), beyond which Compton scattering becomes more relevant. Compton scattering is inelastic scattering between a photon and electron, which results in a longer wavelength photon and a recoiling electron. Finally, at energies beyond 1.022 MeV, pair production can occur where a photon is converted into an electron positron pair. The pair subsequently annihilates, releasing a pair of photons.

3.3 Previous results on Energetic Electron Precipitation

Approximations and descriptions of electron precipitation, the resulting production of X-ray photons, and their subsequent transport through the atmosphere have existed for a long time. An early comprehensive review was carried out by Brown (1966). The first approximations of this processes ignored multiple scattering (Anderson and Enemark (1960), Brown (1965), Christensen and Karas (1970), Barcus and Rosenberg (1966), kamiyaha h. (1966), Polk (1965)). The first attempt with the effects of multiple scattering included was done by Rees (1963). The method used was based on empirical measurements of the energy dissipation of electrons in air.

There are some findings from Rees (1963) which can be used to make our analysis of the problem more simple. They note that the effect of the Earth's magnetic field, which can roughly be thought of guiding incident electrons, shouldn't have a significant effect on the energy deposition altitude profiles. This is a big simplification, because the Monte Carlo simulations we will employ are highly sensitive to the simulation spatial scale. Ignoring the effects of the magnetic field allows effects which occur on the scale of the electron gyrofrequency to be ignored, making the simulations much faster. As noted by Berger and Seltzer (1972), the number of photons which survive to detection is small compared to the number of incident electrons. This implies that Monte Carlo simulations of the resulting photon spectra will be expensive, so to obtain statistically meaningful results, simulation speed is important.

Another important result from Rees (1963) is that the initial angular distribution of the precipitating electrons has only a minor effect on the energy deposition profiles, once the total path length through the atmosphere is accounted for. This has two implications: on one hand, the problem of finding X-ray photon spectra corresponding to precipitating electron spectra becomes a single dimensional problem in energy, rather than a two dimensional one in angle and energy. This represents a large simplification. On the other hand, this also implies that obtaining information about the angular distribution of the precipitating electrons from X-ray measurements is a difficult problem. Our approach in this chapter will be to verify this fact through simulation, and then apply the

problems insensitivity to electron angular distribution as a useful simplifying assumption.

Berger and Seltzer (1972) provide the number of X-ray photons created per incident electron as a function of electron and X-ray energy, for a given altitude in the atmosphere. These curves are reproduced in Figures 3.3, 3.3, 3.3 for altitudes of 32, 34, and 40 kilometers above ground. The strong dependence on the photon population to the sample altitude is apparent, as well as the characteristic shape of the photon spectra. Their results show that one aspect of the problem that cannot be ignored is the dependence on detector altitude. For a given experiment, this will need to be recorded, and the model run at the same altitude.

It is prudent to again emphasize at this point that modelling the production of X-rays in the atmosphere caused by precipitating electrons is not a new problem. We need to make a model similar to those of Rees (1963) and Berger and Seltzer (1972), not because their results are incorrect, but because we need the flexibility of repeated model runs under different scenarios. Further, when the inverse problem of determining causative electron spectra based on X-ray measurements is attempted in the next chapter, the energy resolution needed in the model output will exceed that of the data provided by these authors.

In Berger and Seltzer (1972), resultant X-ray spectra are shown which correspond to mono-energetic beams of electrons. This can be thought of as a Green's function approach to the modelling problem. Evaluating the model for every input electron spectrum of interest is not necessary, provided that the mapping between the input and output is linear. Since the state of the atmosphere is not altered significantly by incident beams of precipitating electrons, different electron beams will produce X-ray spectra which combine additively. For the same reason, the intensity of the resulting X-ray beam will scale linearly with the intensity of the beam of precipitating electrons. In the Green's function approach, beams of mono-energetic electrons are modelled, and the resulting X-ray energy spectra are used to form the columns of a matrix. This matrix then maps between electron energy spectra and X-ray energy spectra for a given altitude. Provided the angular distribution of the incoming electrons can be neglected, and the responses to individual mono-energetic

beams can be evaluated to sufficient statistical purity, then model runs for arbitrary precipitating electron spectra can be produced with a single matrix multiplication.

3.4 GEANT4 and Monte Carlo Simulations

There are several computer software packages used to simulate the transport of radiation through matter. Two of the most popular are GEANT4 (GEometry And Tracking), and MCNP (Monte Carlo N Particle). In this project, I use the GEANT4 package, since it is open source, and well-documented. The scope of the GEANT4 package is vast, and is used in fields ranging from medical physics to nuclear reactor and detector design. The three main references for the package are Agostinelli et al. (2003), Allison et al. (2006), and Allison et al. (2016). GEANT4 is implemented as a set of C++ libraries which provide object-oriented abstractions to nuclear particles and transport processes. The modularity of this design allows for the creation of simulations with a complexity that can be adjusted to the problem at hand. The simulation I created for predicting X-ray spectra from electron precipitation spectra uses only the necessary components of GEANT4. This is important both for making the resulting program understandable and maintainable, but also for speed. Subjectively, there is no upper limit to the amount of complexity that can be incorporated into this type of software. The aim is to represent only the relevant physics and collect only the data needed for the problem being solved.

Programs that use the GEANT4 toolkit do so through a defined interface. The structure of a GEANT4 program follows the class diagram shown in Figure 3.4, which is reproduced from Pfeier et al. (2001). The major components of a GEANT4 simulation are as follows:

- Detector geometry and readout: The sensitive component of a GEANT4 simulation is called a detector. This is a representation of the physical object in which particle interactions are measured. An example is the sodium iodide crystal in a scintillation detector. The detector is associated with the necessary infrastructure to read measurements, such as, for example, the energy deposition by an incident particle

Atmospheric Bremsstrahlung Spectra for Monochromatic Beams
Berger and Seltzer (1972)

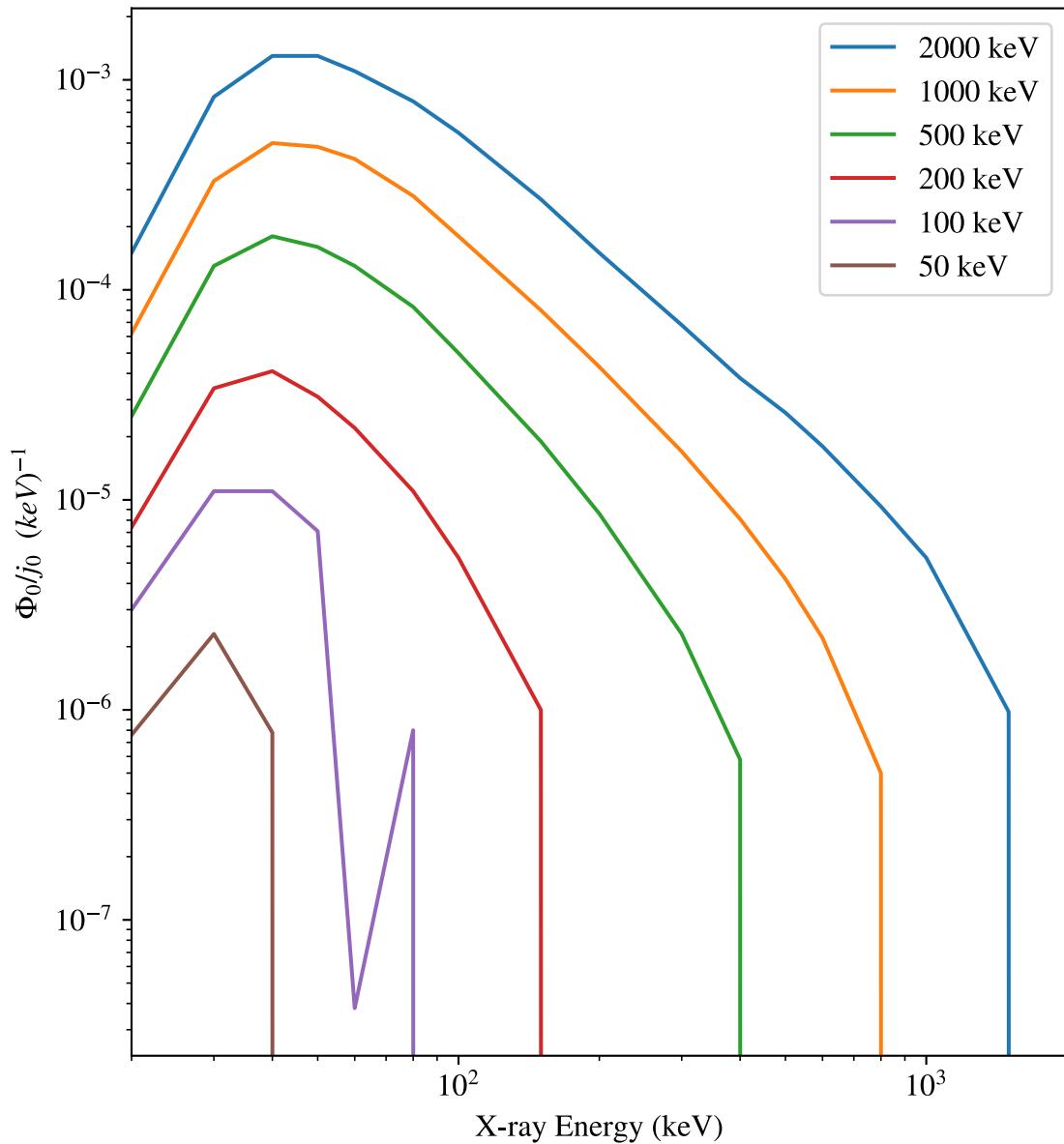


Figure 3.1: X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 32 km.

Atmospheric Bremsstrahlung Spectra for Monochromatic Beams
Berger and Seltzer (1972)

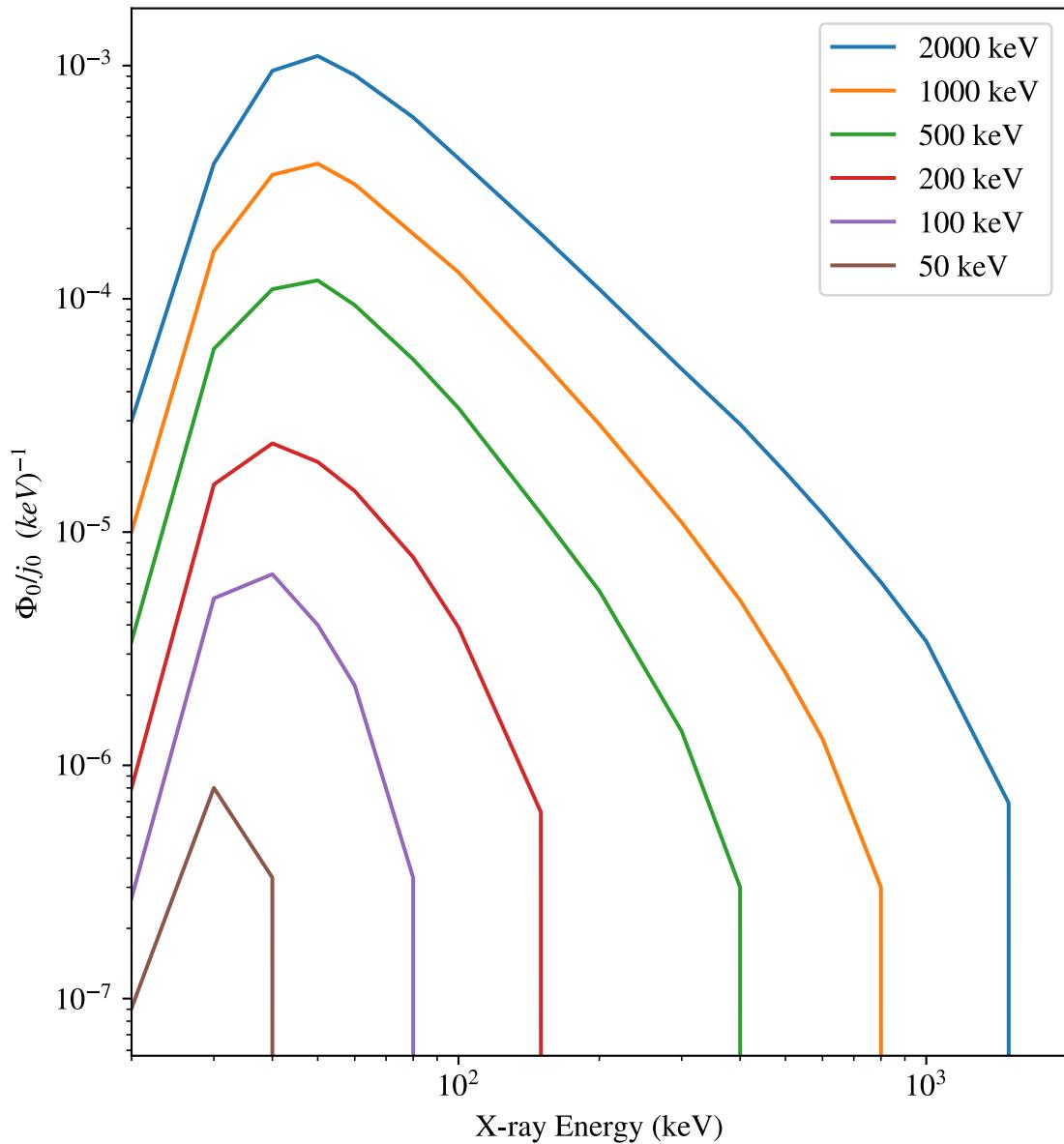


Figure 3.2: X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 34 km.

Atmospheric Bremsstrahlung Spectra for Monochromatic Beams
Berger and Seltzer (1972)

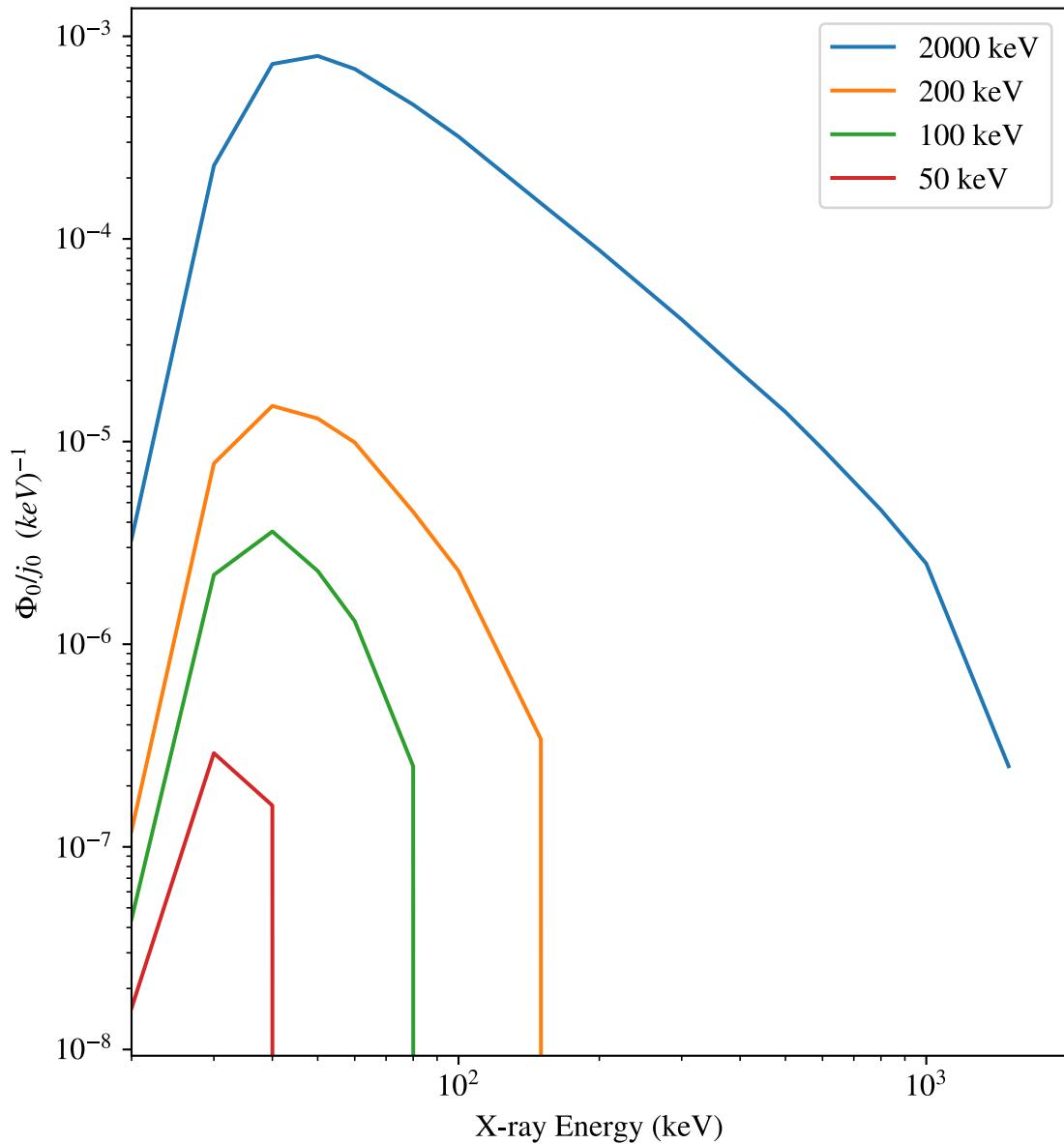


Figure 3.3: X-ray photon flux per incident electron flux for different incident electron beam energies, from Berger and Seltzer (1972) for an altitude of 40 km.

beam.

- Run / Run Action: An instance of a GEANT4 simulation is called a run. Each run usually consists of a set number of incident particles on the detector.
- Event: Particle interactions are called events. There are many events for a given run.
- Tracking: Particles are associated with tracks. Each track is a straight line, which starts and ends with an event. When an event occurs, the simulation physics determine the probability of each interaction along a track.
- Hits and Processes: A hit represents a specific instance of a particle interaction with the physical materials represented in the simulation. Each hit has parameters such as energy deposition, incident angle, and others associated with it. These quantities can be stored in a histogram when they occur.
- Particles and Materials: A particle follows a track in GEANT4 and interacts with the materials represented in the simulation.

For the simulation of electron precipitation I make the representation of the Earth's atmosphere the GEANT4 detector. This choice makes it possible to track and record, among other quantities, energy deposition as a function of penetration depth in the atmosphere. The choices made when designing a GEANT4 simulation for a particular experiment need to be done with the understanding that increasing complexity eventually results in diminishing returns. The main focus of this chapter is to develop a simulation which is sufficiently accurate to represent our experiment using a minimal parameter space.

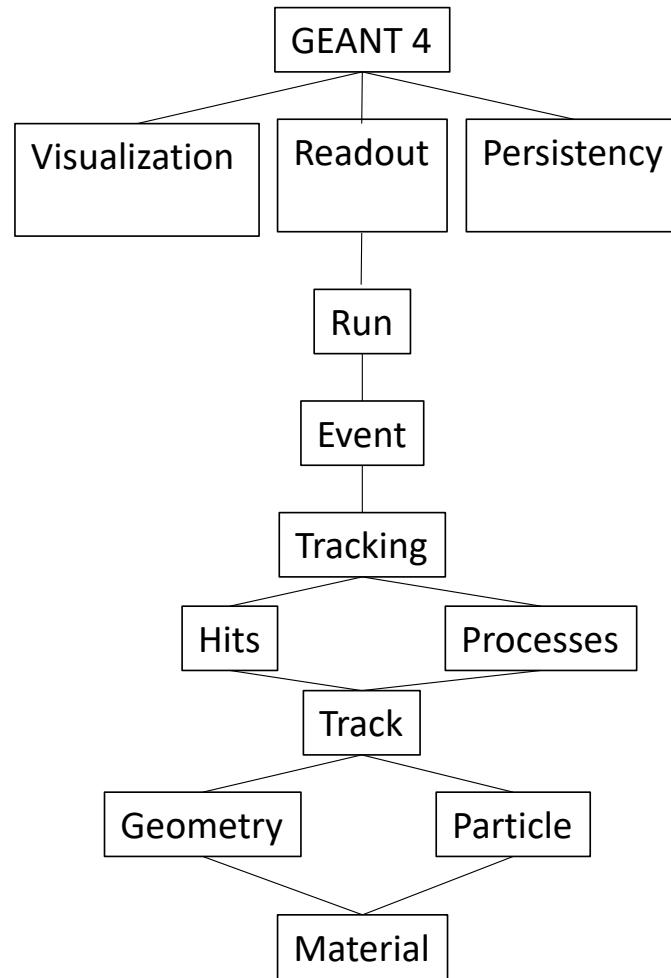


Figure 3.4: GEANT4 architecture, reproduced from Pfeier et al. (2001). Class structure is in the vertical direction, with dependencies towards the bottom.

3.5 Representing the Earths Atmosphere

Precipitating electrons interact with the atoms and molecules which constitute the atmosphere. The MSIS-E-90 model (Picone et al., 2002) contains a representation of the main neutral constituents of the atmosphere, their density, and temperature, parameterized by altitude and geographic location. The specific implementation of the model we use in this chapter is written in C++, which makes it compatible with the GEANT4 framework, and is available at <https://www.brodo.de/space/nrlmsise/index.html>. This implementation encodes the same data as the original FORTRAN version. Figure 3.5 shows the evaluation of this model as a function of altitude for a specific location on Earth.

The altitude profiles of Figure 3.5 must be discretized and encoded as a GEANT4 geometry. This is where approximations must be made to keep the model simple enough to be computationally tractable. Each particle interaction in a GEANT4 simulation takes time to calculate and store, but additionally, an “interaction” occurs at the end of each particle track. The particle tracks end when a particle reaches a sufficiently low energy, but also when it crosses a physical boundary in the simulation, which exist between the faces of the simulation geometry. It is therefore critical to keep the complexity of the geometry as simple as possible. We choose to define a rectangular, Euclidean world, with vertical slices at discrete points representing the atmosphere as a function of altitude. The vertical extent of each slice is chosen to be 1 kilometer, since the major constituents of the atmosphere change density only a small amount on that scale. The horizontal extent of the simulation is set to an arbitrary large number, which for practical purposes can be taken as infinity. This ensures that particle tracks will not terminate by leaving the simulated world through the side boundaries. The model atmosphere is represented to a height of 500 kilometers above ground, beyond which the relevant particle interactions with the atmosphere very rarely occur.

Each slice of atmosphere in the GEANT4 model is treated as a “sensitive detector”, which is an abstraction that permits quantities such as energy deposition, entry angle, and the creation of secondary radiation to be recorded as they occur in the simulation.

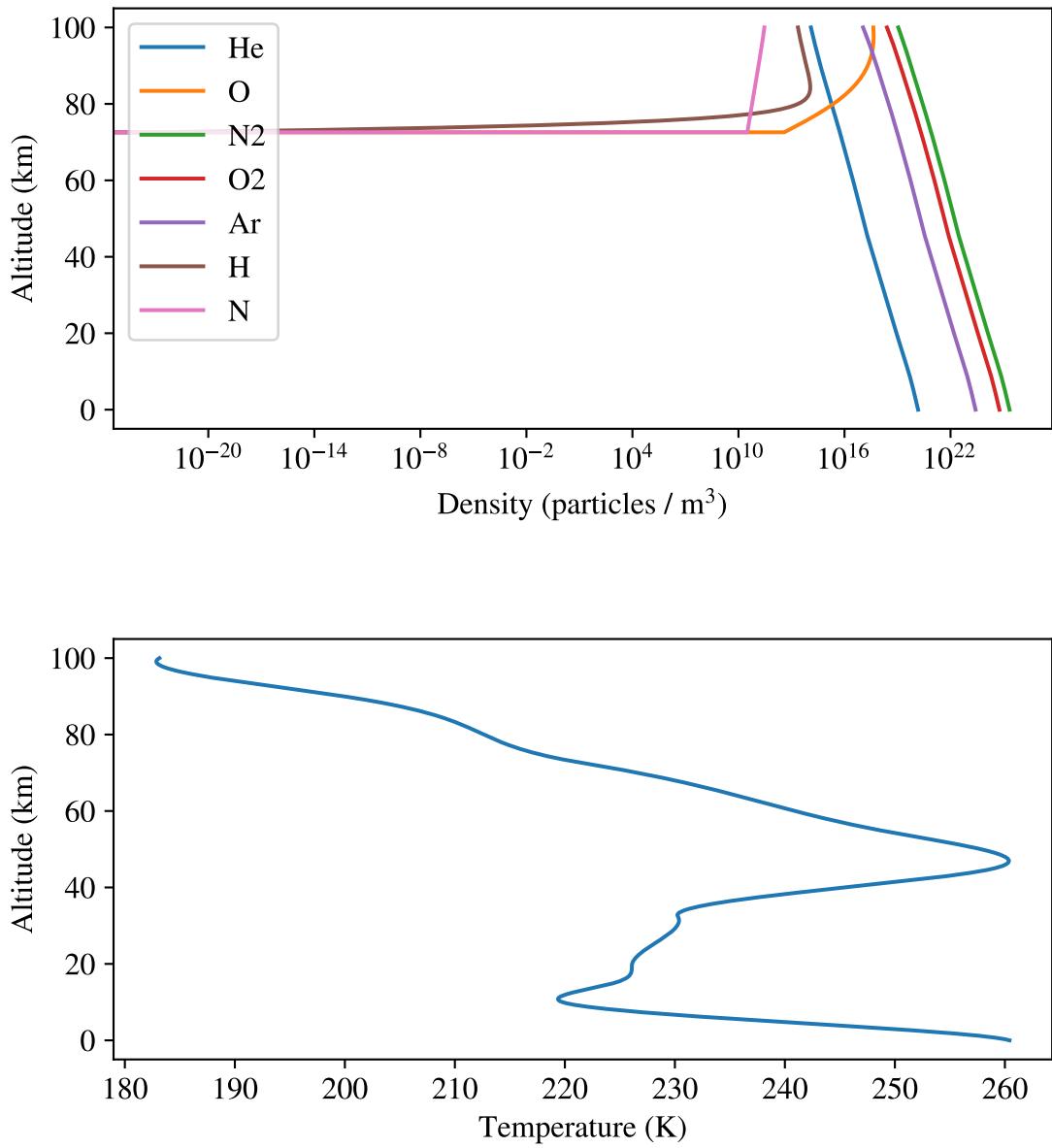


Figure 3.5: Evaluation of key MSIS model parameters as a function of altitude. Particularly important for the GEANT4 simulation are the density profiles of each atmospheric constituent (top).

A GEANT4 volume is constructed using a stack of three abstractions. First, a G4VSolid is created, representing the shape and dimensions. The G4VSolid can be created from either a small set of geometry primitives (cylinders, boxes, and spheres), or it can be imported from an arbitrary mesh of points. The next component is called a G4LogicalVolume. This abstraction holds a reference to a G4Material, which defines the physical substance that the volume is composed of. The G4Material has the density, temperature, and molecular and atomic components of the volume material specified. The G4LogicalVolume also contains flags which control whether it is a G4SensitiveDetector, or a passive component of the simulation, through which particles are simulated but which does not return data. Finally, a G4PhysicalVolume is created which wraps the rest of the stack into one reference to one object. This object is then placed in the simulation during the initialization routine.

3.6 GEANT4 Physics and Simulation Construction

Once the physical properties of the GEANT4 are fully specified, a selection of possible interactions between particles is chosen. There are many GEANT4 physics packages available, which are applicable over different energy ranges and for different types of particles. For this project, I chose the standard electromagnetic physics package (Burkhardt et al., 2004). This package represents interactions between electrons, photons, and atomic nuclei from energies ranging from MeV down to keV. Effects that occur primarily below a single keV, are not represented (Guatelli et al., 2004). Single keV photons and electrons do not travel far in the relevant parts of the atmosphere, and are quickly stopped by the materials surrounding the X-ray instrument.

The input to the atmospheric simulation is specified by electron distributions in incident angle, and energy. Each of these are specified in GEANT4 macro files, which are text files that are used to script GEANT4 applications, and are used to avoid the need to load all the relevant data and inputs at compile time. Topside electron energy and angle distributions can be specified through a list of values, and there are also pre-built functions for the most common forms. For the Green's

function approach used for this task, the input energies are a single value which is swept over the relevant energy range (100 keV to 5 MeV).

Which angular distribution to use is not obvious, since it depends on the injection process responsible for causing the electron precipitation, the local magnetic field, and possibly other parameters. A Green's function approach here would be complicated, and the resulting simulation would need to span discrete representations of at least two dimensions. It is more practical to fix the angular distribution for a given simulation, and apply a Green's function approach only over the relevant energies. Only the electron energy spectrum can be retrieved unambiguously from the secondary Bremsstrahlung radiation (Brown et al., 2006). This indicates that the effect of the angular distribution is small. I chose an isotropic distribution when creating this model.

The inputs to the GEANT4 simulation are then the precipitating electron distribution and total flux. The angular distribution and the model of the atmosphere are fixed. The outputs from the GEANT4 simulation are the distributions, in both angle and energy, of the produced X-ray fluxes, along with the total number of X-ray photons in the selected vertical slice of atmosphere. To obtain statistical results from the model, it is run for a large, arbitrary number of precipitating electrons. The results are represented in units per incident electron flux. The production of secondary radiation happens less at lower energies, so the chosen incident flux increases towards this limit. Because of this, the most expensive part of the simulation happens when evaluating the effect of low energy (less than 100 keV) precipitating electrons. Figure 3.6 shows the inputs and outputs of the GEANT4 simulation schematically.

The diagram in Figure 3.6 is the flow of information for a single GEANT4 simulation run. Since we are applying a Green's function approach, this process needs to be repeated across different monoenergetic electron beams. When this is completed, the responses to arbitrary precipitating electron distributions can then be represented as a linear combination of the recorded monoenergetic responses, without requiring further simulation runs for a fixed sample altitude and electron angular distribution. The workflow for applying the resulting data is shown in Figure 3.6.

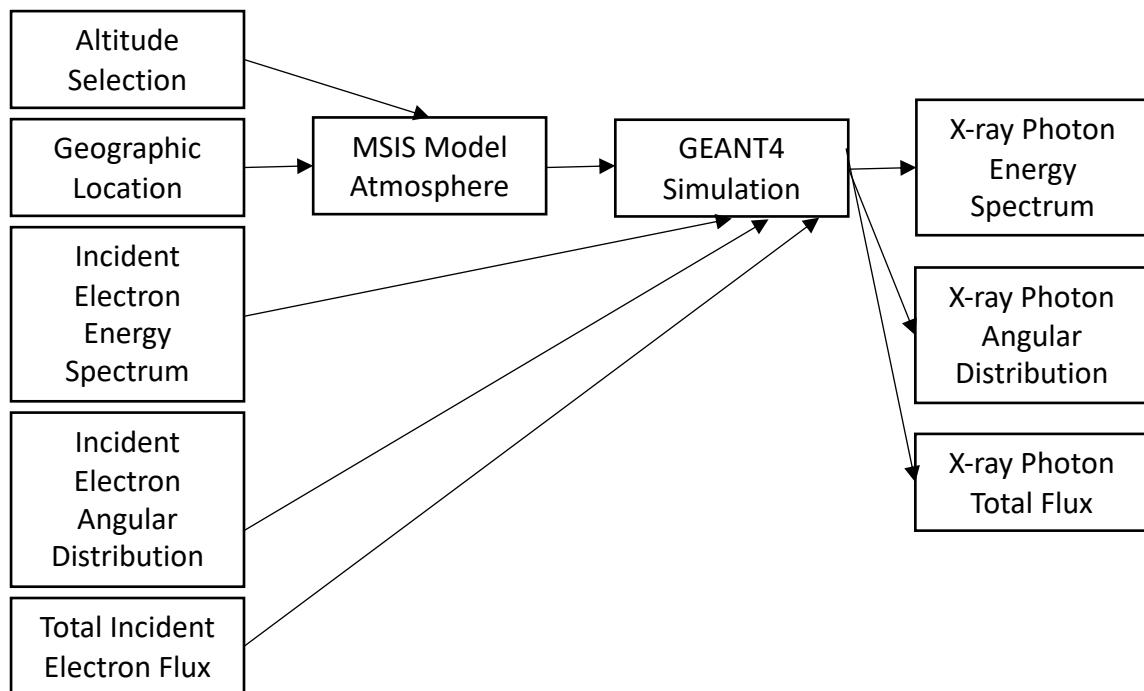


Figure 3.6: Schematic diagram of information flow for a single GEANT4 simulation of energetic electron precipitation and the resulting X-ray radiation

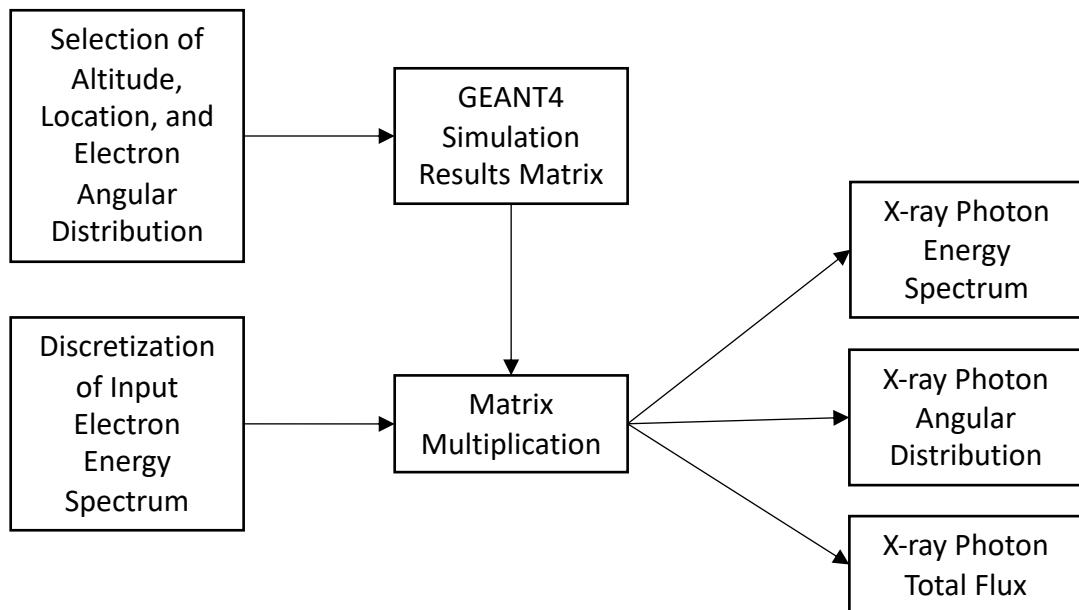


Figure 3.7: Schematic diagram of information flow for a single GEANT4 simulation of energetic electron precipitation and the resulting X-ray radiation

The simulation of all the GEANT4 runs required is computationally expensive. Using the University of Calgary ARC computer cluster, each run was scripted to be run across the required electron energy ranges, and across unidirectional, omnidirectional, and cosine angular distributions. The simulation ranges were then evaluated again over altitudes from 30 to 34 kilometers above ground, which correspond to where viable balloon measurements of the X-ray photons can take place. In total, several weeks of computation across over 100 CPU cores were required. Once this was completed, the results were aggregated and stored in plain text files. While the computation was expensive, the results are compact, and are stored in less than one gigabyte of data.

3.7 Validation of Results

As stated at the beginning of this chapter, the simulation problem is not new. We are reproducing the work of others so that it may be run over the parameter space we need, especially when attempting to solve the inverse problem. The first published results which we can compare to are by Rees (1963). This author presents a semi-empirical form for the results, with a high degree of normalization to the properties of the atmosphere. The work done by Berger and Seltzer (1972), on the other hand, is presented in numerical tables in physical units that are close to the output of the GEANT4 simulation. This makes them a good target for comparison to our work.

The NASA BARREL (Balloon Array for Radiation-belt Relativistic Electron Losses) project has made their modelling work on this problem available through the SPEDAS software package (Angelopoulos et al., 2019). The BARREL project will be discussed in Chapter 4. For now, we will only use their atmospheric simulation as a comparison to our own work. It is distributed in a format comparable to that used by Berger and Seltzer (1972), and can be shown in the same parameter space. Figure 3.7 show the expected X-ray flux, in normalized units, for isotropic incident electron beams at varying energies. The output of my model agrees with the published results in terms of the produced X-ray flux and the shape of the resulting distribution.

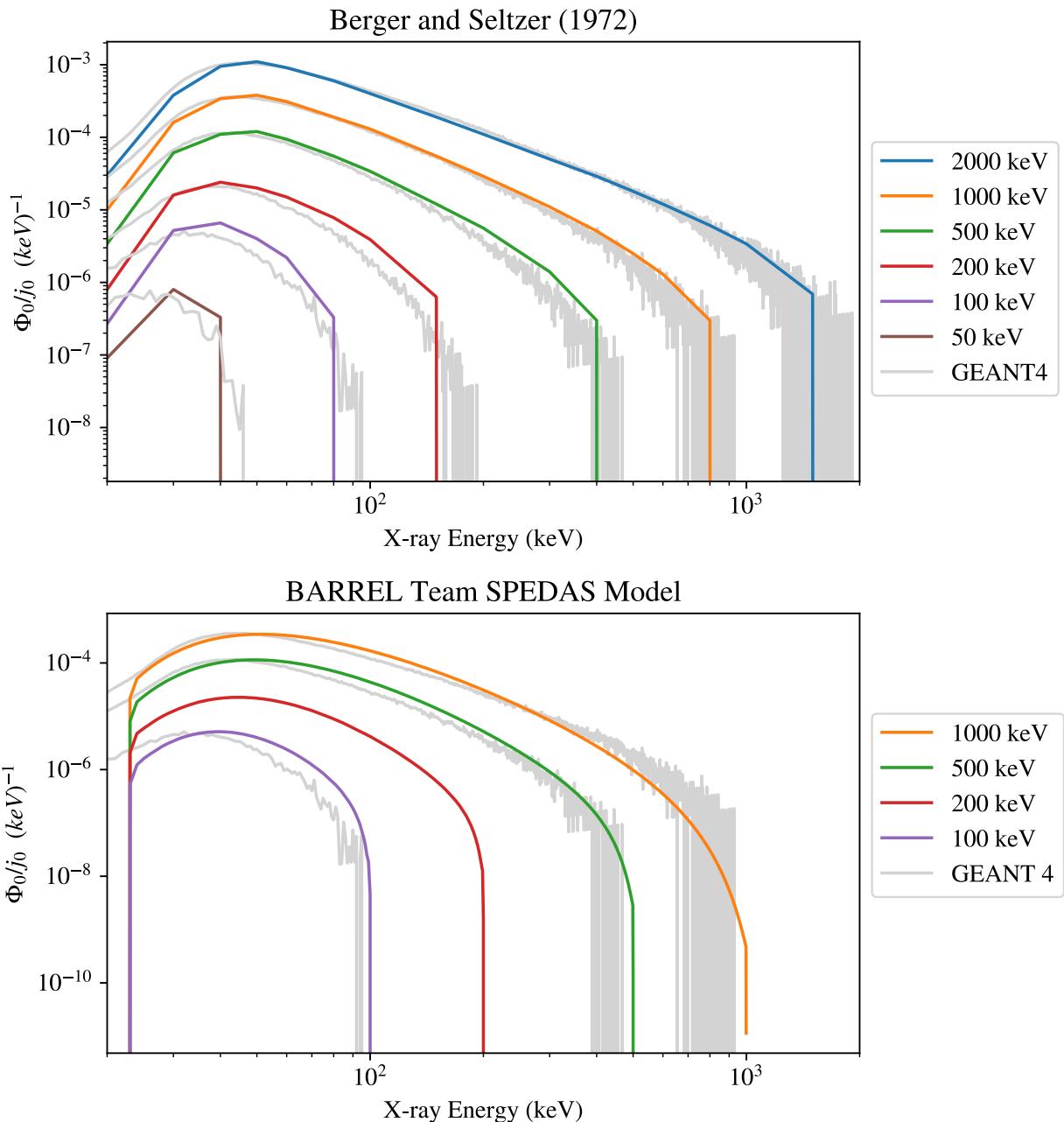


Figure 3.8: Expected X-ray flux as a function of energy at a sample altitude of 34 kilometers using data from GEATN4 simulations, Berger and Seltzer (1972), and the BARREL SPEDAS model.

3.8 Problem Separability

The distributions of photons of a given polar angle and kinetic energy are outputs from the simulation. We will show that the polar angle distribution does not depend on the energy of the precipitating electrons. This allows the results of the simulation to be expressed through their impulse responses in energy only, which is a large simplification. Figure 3.8 shows the polar angle distribution for X-ray photons created by monoenergetic beams of precipitating electrons. Once the total photon counts are expressed in the same units, there is no significant difference in the resulting distributions.

We further show that the photon distributions in kinetic energy do not depend on the angular distribution of the precipitating electrons. This result was originally found by Rees (1963), and implies that experiments can not recover the angular distribution of the parent electron spectrum. For the attempt at the recovery of the parent electron spectrum in the following chapter, this is useful, since this represents a degenerate dimension of the inverse problem that can be ignored. Figure 3.8 shows GEANT4 simulations of the X-ray energy distributions caused by precipitating beams of electrons following unidirectional (downward), and isotropic angle distributions. The distributions are the same at low energies (≤ 200 keV), and .

The final simplification for the simulation problem comes from the parameterization of the detector response. So far we have focused on the creation and validation of the atmospheric model, which gives the angle and energy spectra of the X-ray photons which exist at the detector altitude. The detector output when it is exposed to this population will be a function of this population, and must be modelled through further simulation. The general operating principles and design of the scintillation detectors used in this experiment were described in the previous chapter. For the detector simulation, a GEANT4 model of the geometry is constructed (Figure 3.8.)

We examine how the detector geometry selects a population from the X-ray angular distribution. In the detector simulations, an isotropic source of X-ray photons at a fixed energy are created. We run two simulations independently: the first has the entire detector geometry, with its construc-

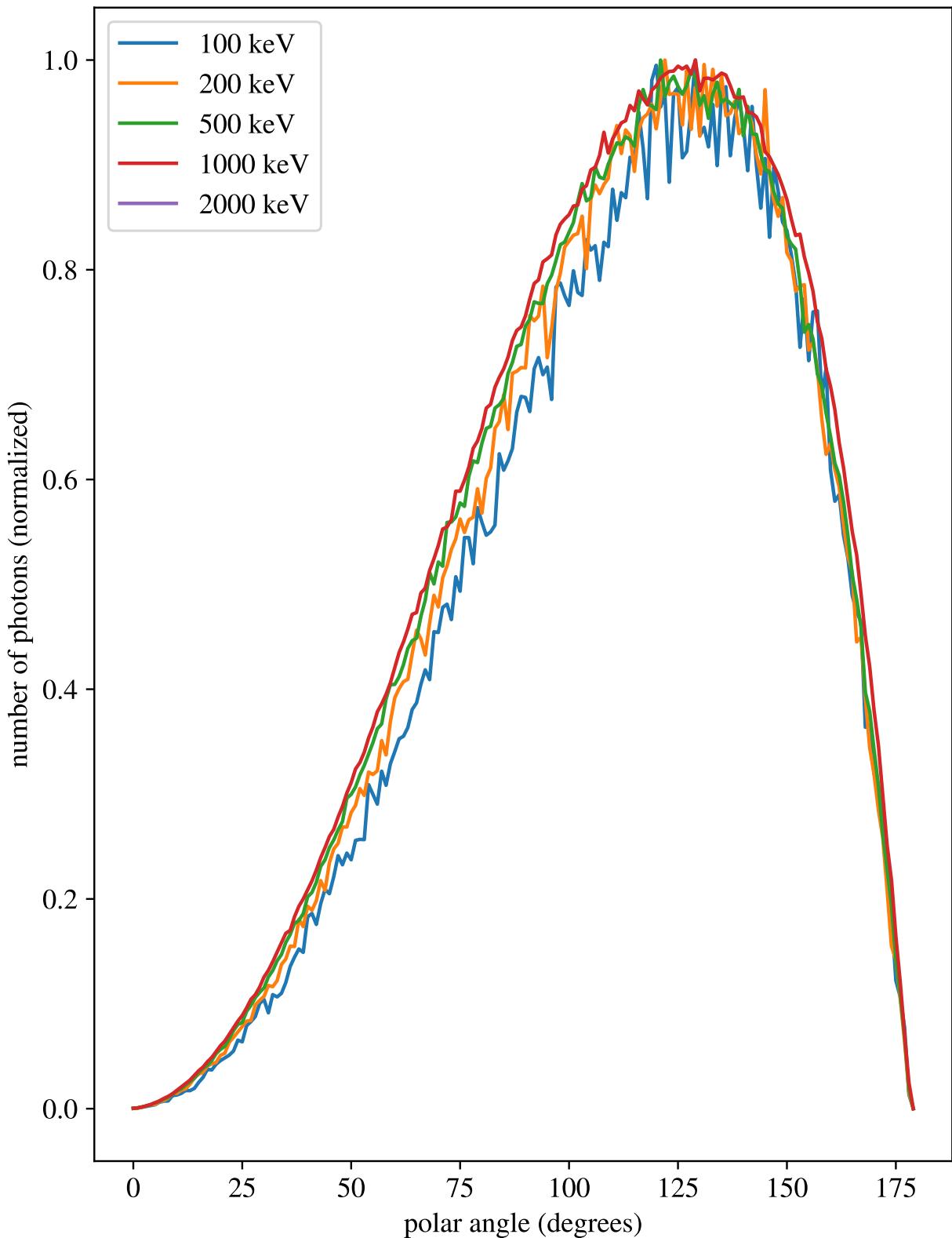


Figure 3.9: Polar angle of X-ray photon population at 34 kilometer altitude, as produced by precipitating electrons at different fixed energies. Units are normalized to account for the higher number of photons produced with greater beam energies.

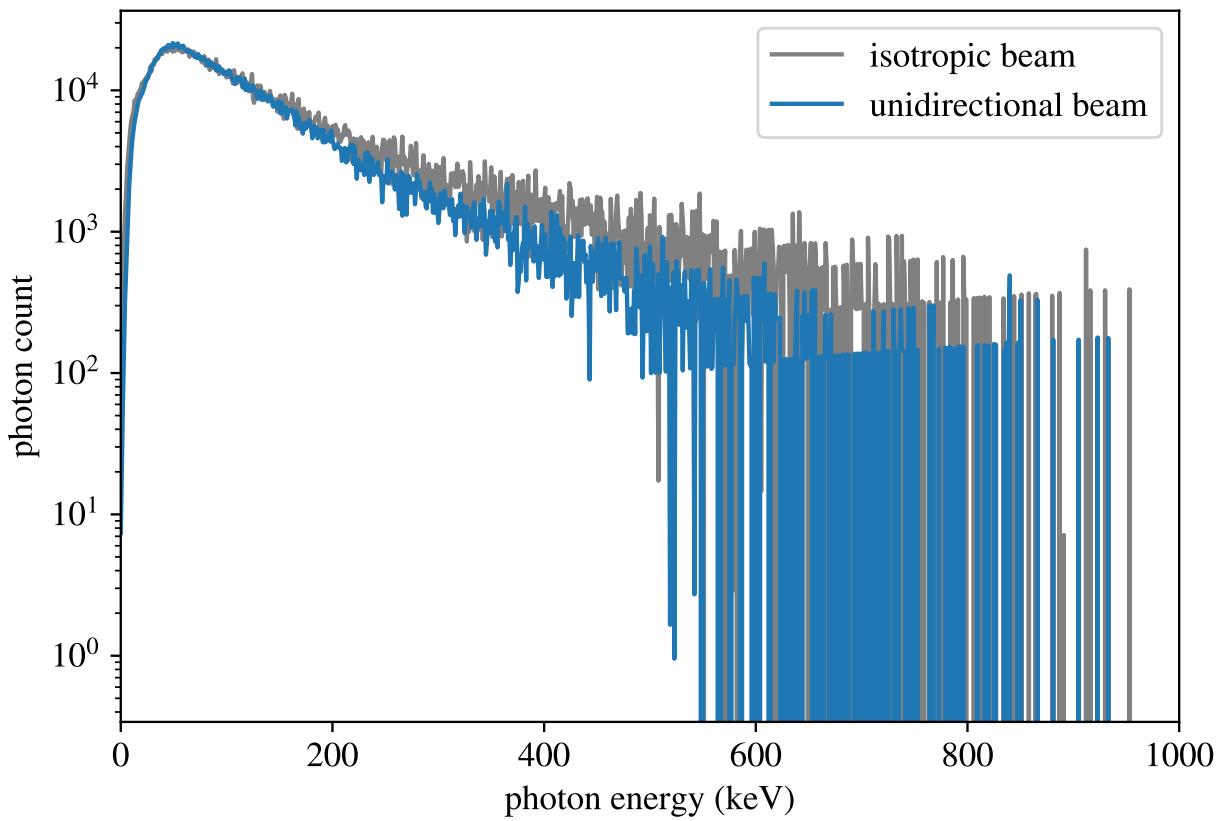


Figure 3.10: Photon energy distribution for 1 MeV incident electron beams having isotropic and unidirectional angular distributions.

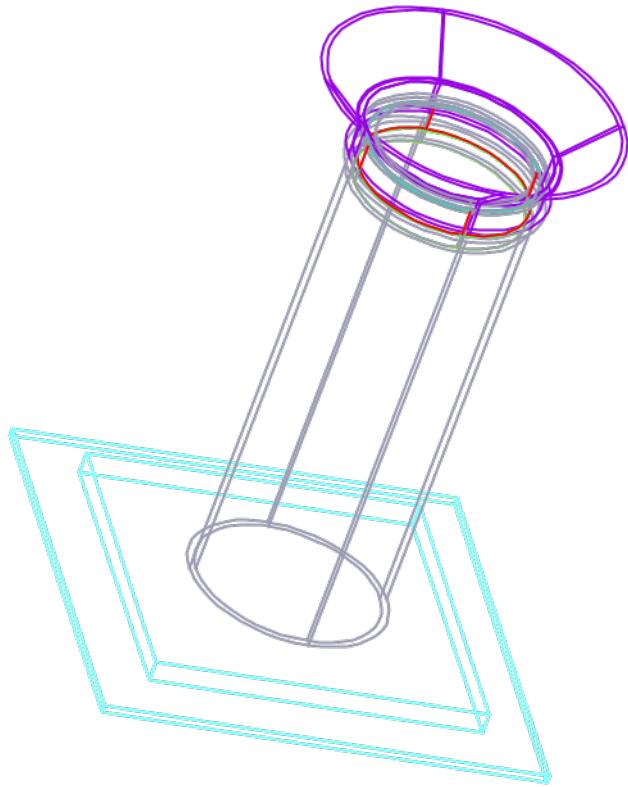


Figure 3.11: GEANT4 representation of the detector geometry. The lower plexiglass mount is shown in cyan, the aluminum body in grey, and the led collimator in purple. The height of the complete unit is approximately 30 cm.

tion materials, enabled. The material corresponding to the sodium iodide crystal, however, is set to a vacuum. This is done to measure the population of photons that reach the crystal, rather than the dynamics of the energy deposition events that happen inside of it. The second simulation is the same, except that the detector materials are set to a vacuum as well as the sodium iodide detector. This provides a baseline which we compare against the first simulation. The distribution of photons which enter the detector geometry will be subset to those recorded when the detector materials are set to a vacuum. Figure 3.8 shows the results these two simulations. The lead collimator assembly at the top of the detector has limited the field of view. Some backscatter from the detector materials can also be seen in the first simulation.

The atmospheric simulation showed that the angle distribution of the X-ray photons does not significantly depend on either the angular distribution of precipitating electrons or their energy distribution. Figure 3.8 indicates that the effect of the lead collimator of the detector is essentially to limit the angle from which photons may enter the sodium iodide crystal. Together, this means that the angular portion of the complete problem from electron precipitation, to the creation of x-rays, and finally, to detection, can be largely treated as trivial. To a good approximation, and at the relevant energies (100 keV to 2 MeV), a constant scalar which accounts for the detector field of view can be used to map recorded counts to physical units.

The behaviour of the sodium iodide crystal itself to the absorption of X-ray photons is important and can also be answered through simulation. For this, the detector geometry is fully enabled in the simulation, including the sodium iodide crystal. Rather than the distributions of photon energies and angles within the crystal volume, energy deposition events will be recorded. The energy deposition events correspond to the flashes of light which are amplified by the photomultiplier tube and recorded by the detector electronics. Figure 3.8 shows histograms of the energy deposition events for a set of monochromatic photon beams. The responses can be described by delta functions at the corresponding photon energies, with interactions at lower energies which become more complex as the beam energy increases. The main interaction in this energy range is the single

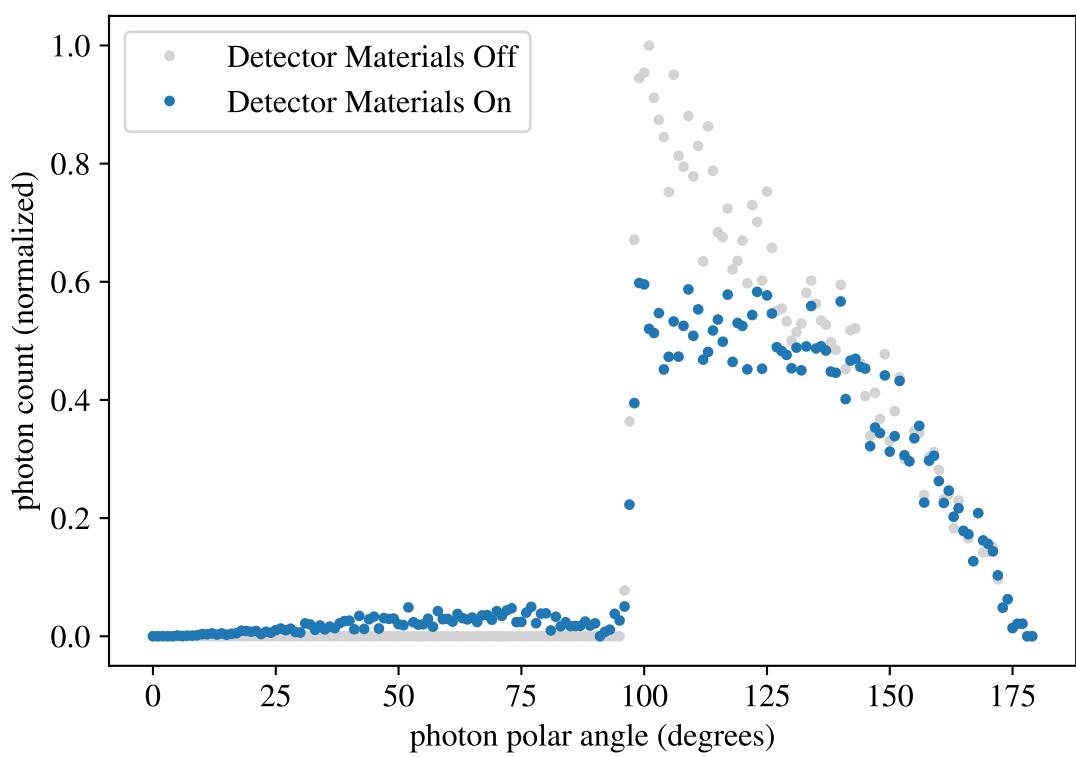


Figure 3.12:

deposition of all the photon energy at once, which is why the peaks are easily detectable.

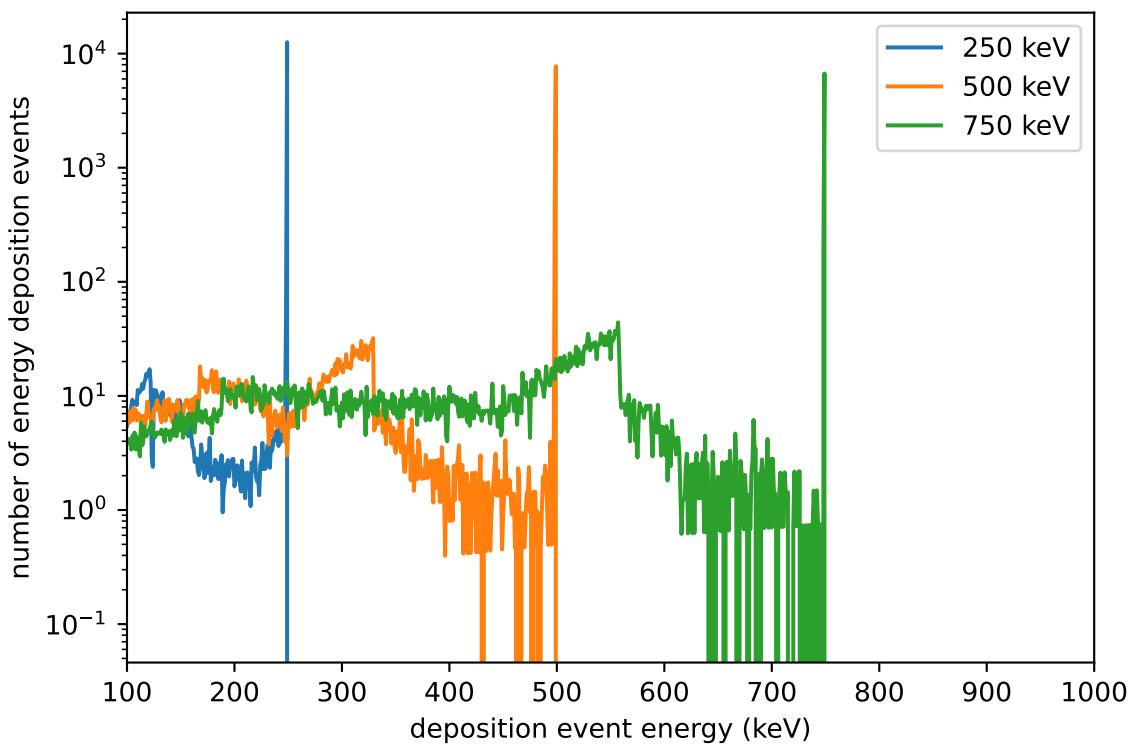


Figure 3.13: Energy deposition events inside the detector crystal for beams of monochromatic photons at different set energies.

Chapter 4

The Inverse Problem

4.1 Definition and Properties

In this chapter the problem of obtaining information about the causative electron precipitation directly from X-ray measurements will be examined. We will term this the “X-ray inversion problem.” This type of analysis belongs to a broad class of what are called inverse problems, which attempt to reconstruct physical events based on limited measurement data. A classic example of an inverse problem is the determination of media velocities within the Earth based on seismic measurements at discrete points. The “forward problem” of predicting instrument responses based on an assumed physical event can be handled by, for example, computer simulations, as was done in the last chapter for the problem of predicting the measured X-ray distributions which result from a given electron precipitation event. Unlike forward problems, inverse problems have mathematical properties which make them difficult or impossible to solve without significant a-priori assumptions. For the X-ray inversion problem, we will show the reasons for this, and demonstrate some measures which can be taken to mitigate these effects.

We will use the terms used by Tarantola (2004). Define the forward problem by the mapping:

$$\mathbf{F} : \mathbf{M} \rightarrow \mathbf{D} \tag{4.1}$$

where \mathbf{D} is the space containing data from measurements, and \mathbf{M} is the space of models, or solutions. The mapping \mathbf{F} may be either linear, or non-linear. Additionally, the definition of the model \mathbf{M} is generally not unique (Tarantola, 2004). For the forward problem, we start with an assumed model and evaluate its predictions against the available data. The metric by which this is

done is another choice which may not be unique for the given problem. The inverse problem is:

$$\mathbf{F}^{-1} : \mathbf{D} \rightarrow \mathbf{M} \quad (4.2)$$

In addition to uncertainty in the measurements \mathbf{D} , knowledge of the mapping \mathbf{F} may be incomplete, or based on uncertain experimental data or simulations. To describe the challenges associated with inverse problems some definitions are needed.

Definition 4.1.1 (ill-posedness). A problem is ill-posed if it fails to have any of the following properties (Hadamard, 1902) :

1. A solution exists.
2. The solution depends continuously on the input data.
3. The solution is unique.

Inverse problems are ill-posed as a general rule. This requires that we shift the analysis from a search for a particular solution, as with forward problems, towards a search for families of solutions, and methods of determining which produce the best description of the available data, often requiring the use of a-priori information about the particular problem at hand. Based on the results of the last chapter, the X-ray inversion problem certainly satisfies property 1, a solution exists. The difficulties lie in a failure to satisfy properties 2 and 3. We will show that without prior treatment, finding a unique solution to the X-ray inversion problem is, in most realistic cases, not possible. Another definition is needed to see why:

Definition 4.1.2 (ill-conditioning). A problem is ill conditioned if a small change in the input data results in a large change in the solution, that is, the condition number (Rice, 1966):

$$\kappa = \lim_{\epsilon \rightarrow 0} \sup_{\|\delta x\| \leq \epsilon} \frac{\|\delta f\|}{\|\delta x\|}$$

is large for small changes in the kernel function δf and input data δx .

The X-ray inversion problem is ill-conditioned. Figure 4.1 illustrates this by showing three different model precipitating electron spectra, and a simulation of their resulting X-ray spectra in log-linear space. Each model electron spectrum was assumed, and the detector response was produced using the forward modelling of the previous chapter. Different model electron spectra produce very similar resulting X-ray spectra. The treatment of this instability will be the main problem addressed in this chapter.

4.2 Framing the X-ray Inversion Problem

The forward problem of determining expected instrument responses from electron precipitation events is linear, discrete, and can be written as:

$$\mathbf{G}\mathbf{m} = \mathbf{d}$$

where \mathbf{G} is a matrix representing the discretization of the kernel function mapping precipitating electron spectra \mathbf{m} to resulting measurements of the X-ray spectra \mathbf{d} . Techniques for computing an approximation of \mathbf{G} were discussed in the previous chapter. There is a result that allows the calculation of the condition number of \mathbf{G} as the ratio of its maximal and minimal singular values. As a rule, if a given linear problem has condition number κ , then we can expect to lose 10^κ digits of precision in calculating the solution (Cheney and Kincaid, 2008). Techniques to frame the X-ray inversion problem in such a way as to reduce the effect of the ill-conditioning of \mathbf{G} will be discussed in the this section.

A representation of matrix \mathbf{G} was developed in Chapter 3 using monte-carlo simulations. In that analysis, a Green's function approach is used to write the X-ray spectra resulting from arbitrary input electron spectra as a linear combination of the effects of mono-energetic electron beams. The dimensions of matrix \mathbf{G} are determined by the binning scheme used to represent electron and X-ray energy spectra. For the forward problem, the particular binning scheme chosen for both spaces is not critical, provided it is fine enough to capture the scale of the details of interest in both spectra,

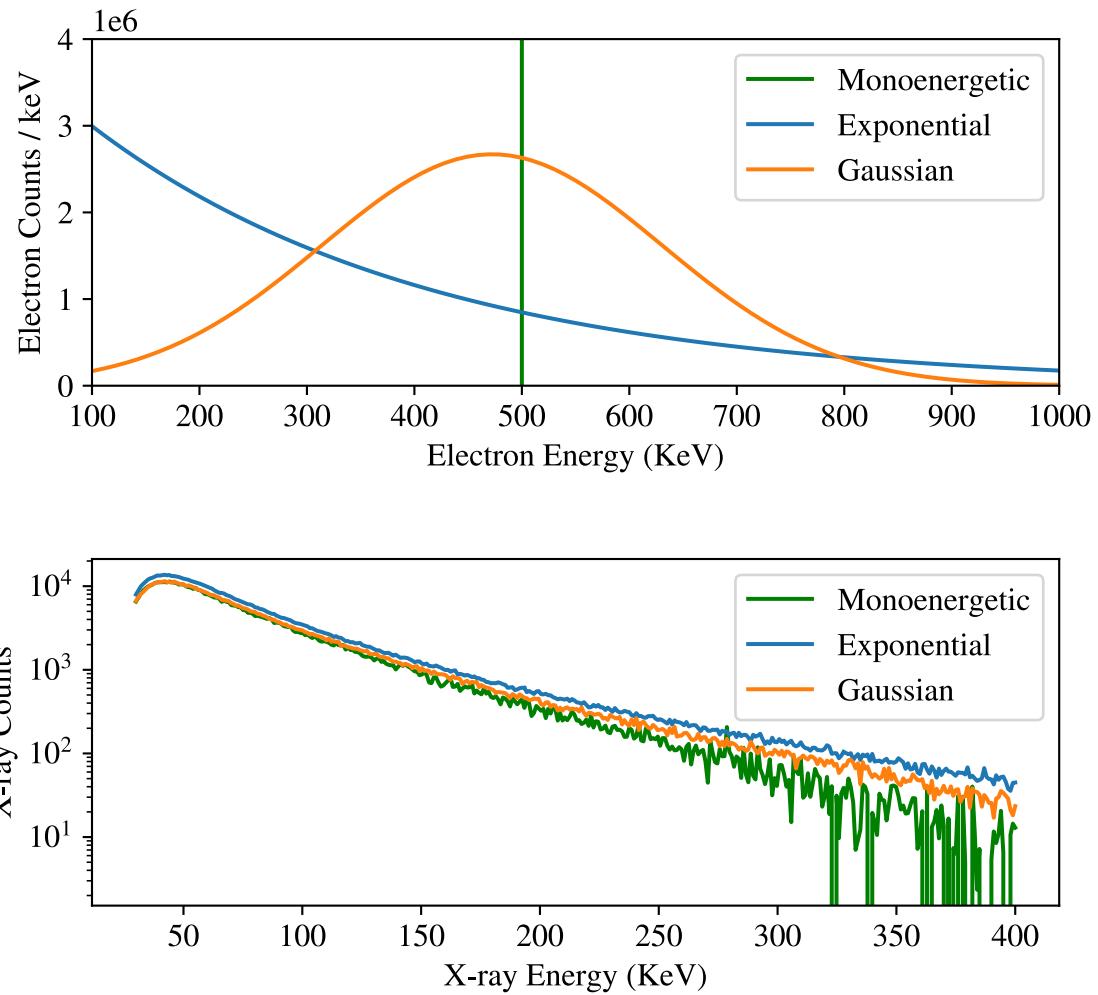


Figure 4.1: (Top): Example of three different assumed electron energy spectra (mono-energetic, exponentially decreasing, and Gaussian) for a precipitating electron flux of 1×10^9 electrons per cm^2 per second. (Bottom): Resulting simulated X-ray spectra for each electron spectrum, using the methods developed in Chapter 3.

but coarse enough to support obtaining adequately pure statistics for the resulting X-ray spectra in a practical amount of computing time. For the inverse X-ray problem, however, the binning scheme chosen for \mathbf{G} has a critical effect on the conditioning of the inverse problem, and can determine whether solutions are possible to obtain entirely. To show the effect of choosing different binning schemes for the rows and columns of \mathbf{G} , we plot the condition number as a function of bin widths in KeV for the input electron spectra and output X-ray spectra in Figure 4.2.

Figure 4.2 shows that there is a set of bin widths for the electron and X-ray spectra which make \mathbf{G} singular, which implies that it has no inverse. There are also some general trends apparent. Smaller bin widths for the X-ray spectrum tend to reduce the condition number of the problem, as do larger bin widths for the electron spectra. This can be viewed as the effect of supplying the problem with more information (more X-ray data points), and requiring less information from the output, corresponding to larger electron bin widths. The singular region of Figure 4.2 corresponds to binning schemes where too much information is being sought from a limited input data set for there to be a unique solution.

The matrix \mathbf{G} is not necessarily square. The inverse used in the non-square case is the Moore-Penrose pseudo inverse, which corresponds to solving the linear system in the least-squares sense (Dresden, 1920; Bjerhammar, 1951). Results from attempting to solve the inverse problem by directly inverting \mathbf{G} are shown for simulated X-ray and electron spectra across different binning schemes in Figure 4.2.

The direct inversion attempts in Figure 4.2 are an example of a fundamental trade-off encountered when solving inverse problems. Because the information contained in the available measurements is subset to the information contained in the physical event being studied, the best solution, in terms of minimizing residuals between predictions of the resulting model and the data, often has a high variance and contains little useful information. This corresponds to case (C) in Figure 4.2, where the electron spectrum bin widths are small enough that the model obtained is mostly random noise. On the other hand, choosing wide bin widths as in case (A) can leave relevant information

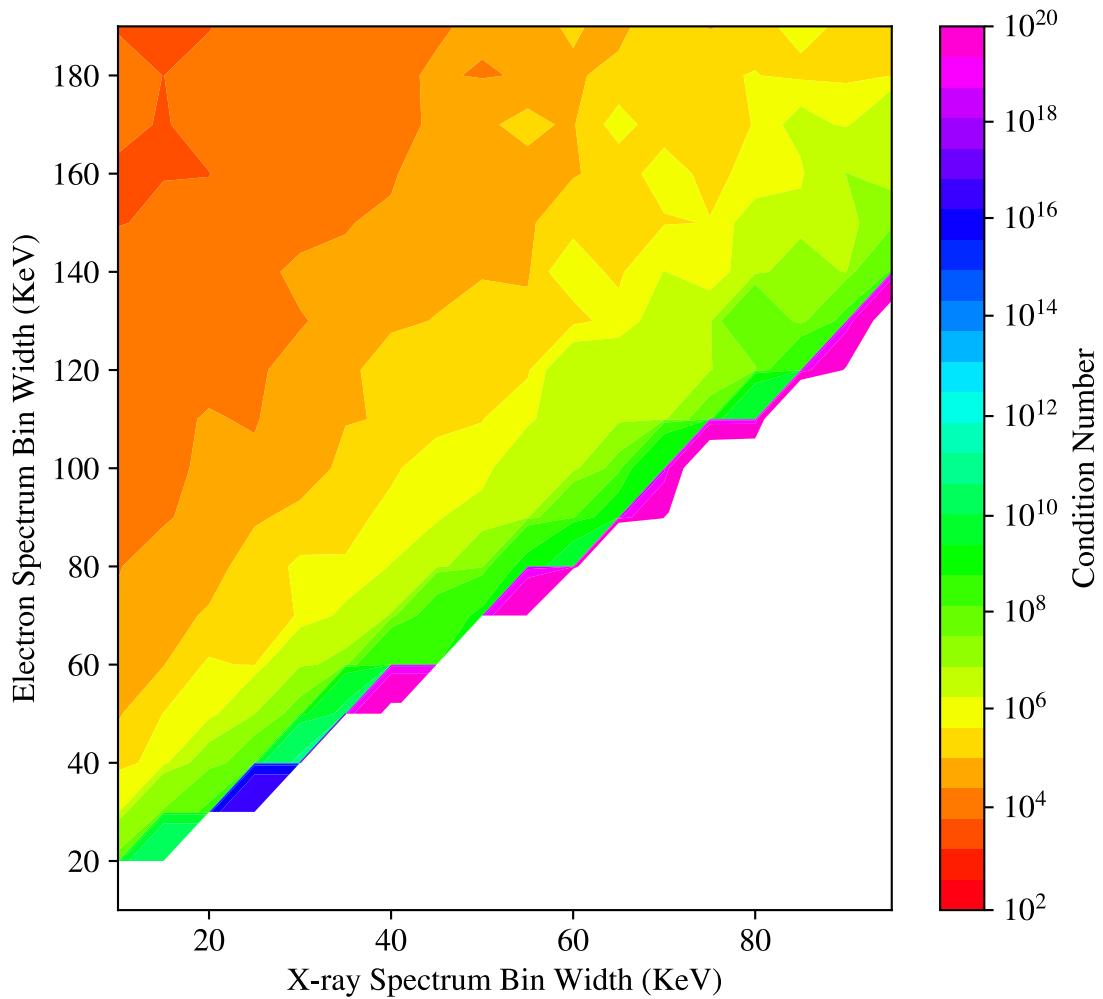


Figure 4.2: Condition number of mapping \mathbf{G} as a function of X-ray spectrum and electron spectrum bin-widths in KeV. The white region is the domain where the condition number is infinite, and \mathbf{G} is singular.

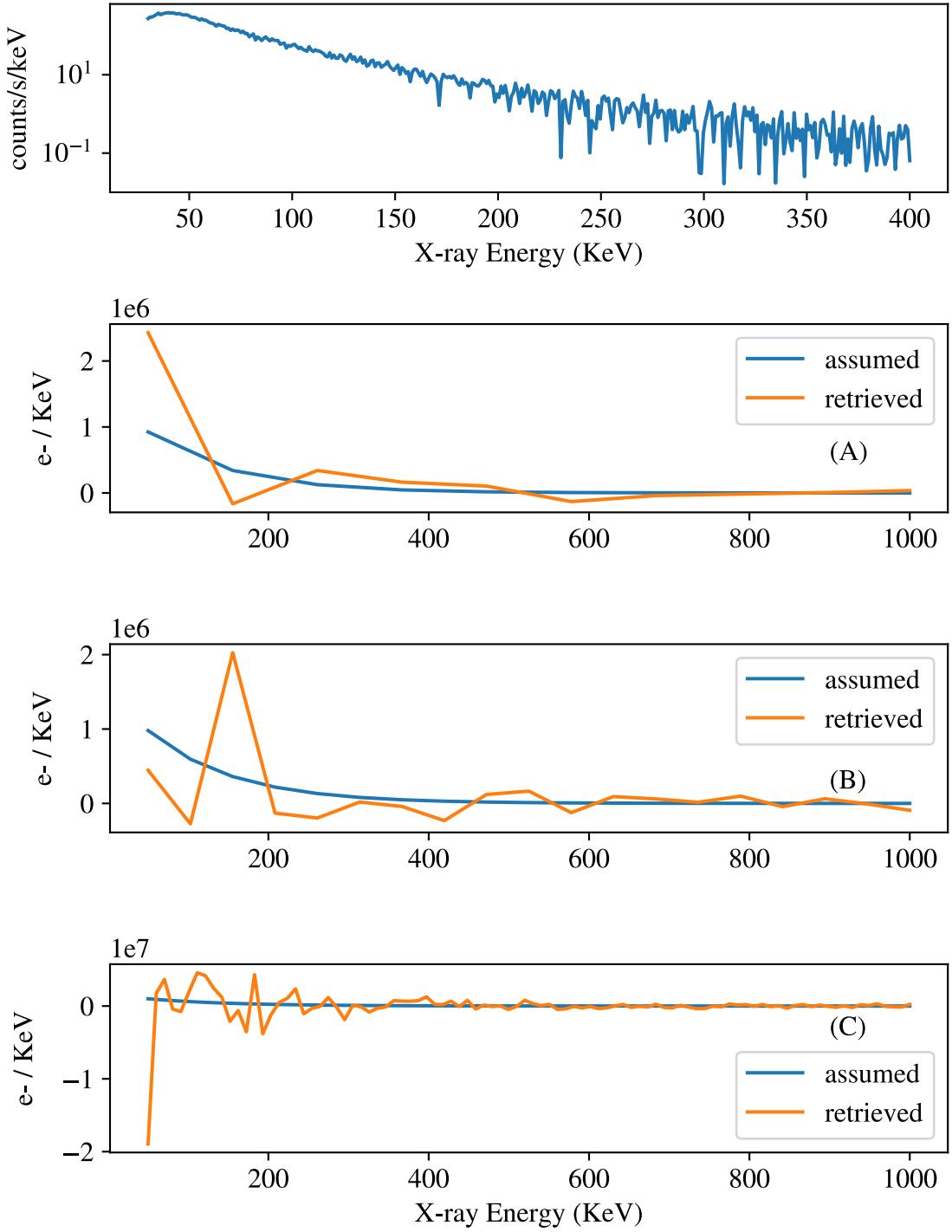


Figure 4.3: Direct matrix inversion applied to simulated exponential electron spectra using bin widths of 10 (A), 50 (B), and 100 (C) KeV. The total incident electron count is 1×10^8 .

contained in the measurements out of the model, so that while the variance in the solution is low, the model is still not descriptive of the event being studied. This trade-off can be handled through optimization techniques, and the application of these will be the main focus of this chapter.

4.3 Creating Synthetic Data

The evaluation of techniques to solve inverse problems often depends on the use of synthetically generated, or simulated, data which correspond to solutions which are already known. For the X-ray inversion problem, synthetic data are generated using the monte-carlo techniques discussed in Chapter 3, closely following the approach used by (Xu and Marshall, 2019). The techniques are not new, the forward problem having been solved by (Berger and Seltzer, 1972), which is based on work going back to (Rees, 1963).

Realistic synthetic data sets always have an error term added, which must be consistent with the statistics of the experiment being simulated. Measuring X-ray spectra is a counting experiment, and therefore the random error associated with the recorded count for a given energy bin has a Poisson distribution. This distribution can be treated as approximately normal for total counts greater than around 10 in each energy bin due to the central limit theorem. This property can be used to simulate an X-ray spectrum which corresponds to an assumed count rate, by adding appropriate noise to the signal produced using the techniques of Chapter 3.

In a real experiment, a measured X-ray spectrum must be background subtracted. The background subtraction is difficult to simulate and is a source of systematic error. Natural sources such as the 511 KeV annihilation line couple with noise inherent to the detector to make a background energy spectrum with a characteristic shape. We will use a sample drawn from experimental data sets to generate a characteristic background spectrum, and then add an appropriately scaled version to synthetic data generated for a given electron distribution and count rate. This approach neglects any systematic error caused by, for example, detector calibration drifting due to temperature changes.

Due to the sensitivity of the X-ray inversion problem to small perturbations in the measured spectrum, care must be taken when applying tests using synthetic data so as not to introduce a sampling bias in the measured performance of different inversion techniques. It is possible to combine a particular background subtraction with a simulated spectrum and corresponding measurement noise to obtain a solution to the inverse problem which very nearly matches the assumed causative electron spectrum through random chance. This may not occur when the same test is subsequently run with a different sample of the background spectrum, or a different random seed for the simulated detector noise. For this reason, single tests of an inversion method on one synthetically generated spectrum have no significance. Evaluating a solution technique to the inverse problem on synthetic data is itself an experiment, and must be treated using a large enough sample size.

Simulated X-ray spectra which are generated using the techniques of Chapter 3 are created without a background component. To account for this in the generation of synthetic data, the average of the available measured background spectra is added to the simulated spectrum, and then a randomly chosen background spectrum is subtracted. We will assume that the differences across all of the measured background spectra can be used as a proxy for the systematic error expected when a particular background subtraction is applied.

There needs to be an objective measure for the relative performance of a given inversion method. The metric should be a scalar, and should increase with the amount of detail that the inversion method captures about the assumed electron spectrum when it is applied to synthetic data. The normalized mean squared error (NMSE) is a simple metric which satisfies these properties. The NMSE is the squared total difference between the retrieved electron spectrum and the assumed electron spectrum, but normalized to the total signal magnitude to account for the fact that it is applied across different count rates. The definition follows.

Definition 4.3.1 (NMSE). The normalized mean squared error between assumed spectrum \mathbf{y}' and expected spectrum \mathbf{y} is (Hanha, 1988):

$$\text{NMSE}(\mathbf{y}, \mathbf{y}') = \frac{\|(\mathbf{y} - \mathbf{y}')\|}{\|\mathbf{y}\|}$$

The NMSE is zero for a perfect agreement between a simulated and retrieved electron spectrum, and increases as the agreement becomes worse. Tests on different inversion techniques will produce different distributions of the NMSE.

4.4 The Least-Squares Regression and the Bias-Variance Trade-off

Due to the ill-posed nature of the X-ray inversion problem, solutions often take the form of estimators. These estimators may be either unbiased, as in the case of least-squares regression, or contain a bias term towards solutions with certain properties. In Figure 4.1, it is apparent that directly solving the X-ray inversion problem using least-squares regression can be ineffective. This is because the problem has a high condition number (Figure 4.2). There is a trade-off apparent: the condition number can be reduced by selecting coarse binning in model space, which reduces the variance in the solution, but also reduces the amount of information about the data the solution represents. The optimization of this trade-off for a given data set is a basic problem in data analysis. A solution is said to be over-fit when it aligns well with one data set, but fails to usefully describe the data under small perturbations or additional measurements. On the other hand, a solution is under-fit when it fails to reflect significant changes in measurement data. The variance in the data set is used as a proxy for how much variance is expected in well-fit models which describe the data. The χ^2 statistic is a standard descriptor for this relative variance (Pearson, 1900).

Definition 4.4.1 (χ^2 Statistic). The χ^2 statistic is the quantity

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{\sigma_i^2}$$

where \mathbf{O}_i are the observed, or measured quantities, \mathbf{E}_i are the expected, or model, quantities, and σ_i^2 are the variances in the measured quantities.

The χ^2 statistic is usually normalized to the number of degrees of freedom which a particular problem has, which gives the reduced χ^2 statistic.

Definition 4.4.2 (Reduced χ^2 Statistic). The reduced χ^2 statistic is the χ^2 per degree of freedom, v :

$$\chi_v^2 = \frac{\chi^2}{v}.$$

If the reduced χ^2 statistic is much greater than 1, then it indicates a poor model fit. When the reduced χ^2 statistic is much less than 1, it indicates over-fitting to the data (Barlow, 2013). An example of this is shown in Figure 4.4 for polynomial models fit to a randomly generated data set.

Managing the balance over-fitting and under-fitting models is related to the condition number of the problem. If a problem is badly conditioned, then it may be necessary to somewhat under-fit the data to generate a solution with a sufficiently small variance. Another problem which frequently occurs for the X-ray inversion problem is the existence of negative particle fluxes in predicted electron spectra, an example of which can be seen in Figure 4.1. There is no clear way to interpret models with these features. Eliminating them from predicted models is possible using constrained optimization techniques, and will be discussed in the next section. Constrained optimization, combined with other modifications to the least-squares estimator to improve conditioning of the X-ray inversion problem, will be the topic of the remainder of this chapter.

4.5 Biased Estimators and Improving the Condition Number

The condition number for a given linear problem can be expressed as the ratio of the maximum and minimum singular values in the kernel matrix. Using the singular value decomposition, it is possible to write an alternative version of the problem with smaller singular values suppressed or removed. The aim is to exchange the original problem for a similar one which is better conditioned. The singular value decomposition writes the kernel matrix \mathbf{G} as:

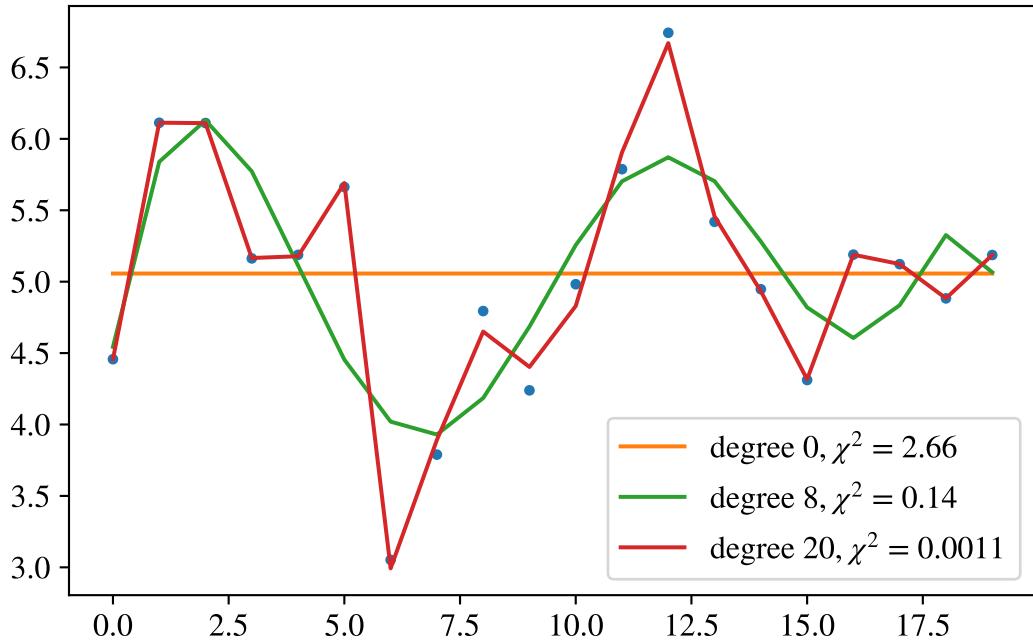


Figure 4.4: Example of over-fit and under-fit models to data. A random data set is generated and polynomial models are fit using least-squares regression. The reduced χ^2 statistic is shown for each model.

$$\mathbf{G} = \mathbf{U}\Sigma\mathbf{V}^T$$

where \mathbf{U} and \mathbf{V} are unitary, and Σ is a diagonal matrix with the singular values on the diagonal. Since the decomposition can be chosen such that the singular values are in descending order, an approximation and reduction in dimensions to the original kernel matrix can be generated by truncating the expansion. Processes like this produce biased estimators, or imperfect representations of the kernel matrix, in exchange for more stable solutions. In this section we will examine the effects of these approximations on the X-ray inversion problem using synthetically generated data from the techniques in Chapter 3.

Figure 4.2 shows that the binning scheme for the X-ray inversion problem with the smallest condition number minimizes the bin sizes for the data (X-ray spectra), and maximizes the bin size for the models (electron spectra). This choice must be made within the limits of the statistics of the data set at hand, and the amount of detail desired in the model electron spectrum. As discussed in the previous section, the χ^2 statistic is used to avoid over or under-fitting the problem. Unfortunately, this test is valid only for unbiased estimators, such as the least-squares regression. For a biased estimator, determining the right solution size becomes a much more complicated problem that will be addressed in the next section.

Figure 4.5 shows the distribution of singular values across different electron spectrum bin widths. The X-ray spectrum bin widths are set at 1 KeV. Most of the information contained in the kernel matrix is captured in the first few terms of the singular value decomposition. This suggests that reducing the dimensions of the problem by truncating the singular value decomposition might improve its condition.

To justify the use of a biased estimator in solving the X-ray inversion problem, two questions need to be answered. The first is whether or not the use of the biased estimator significantly improves the conditioning of the problem. The second is whether the bias significantly degrades the amount of information contained in the solutions. The first question is easily answered by

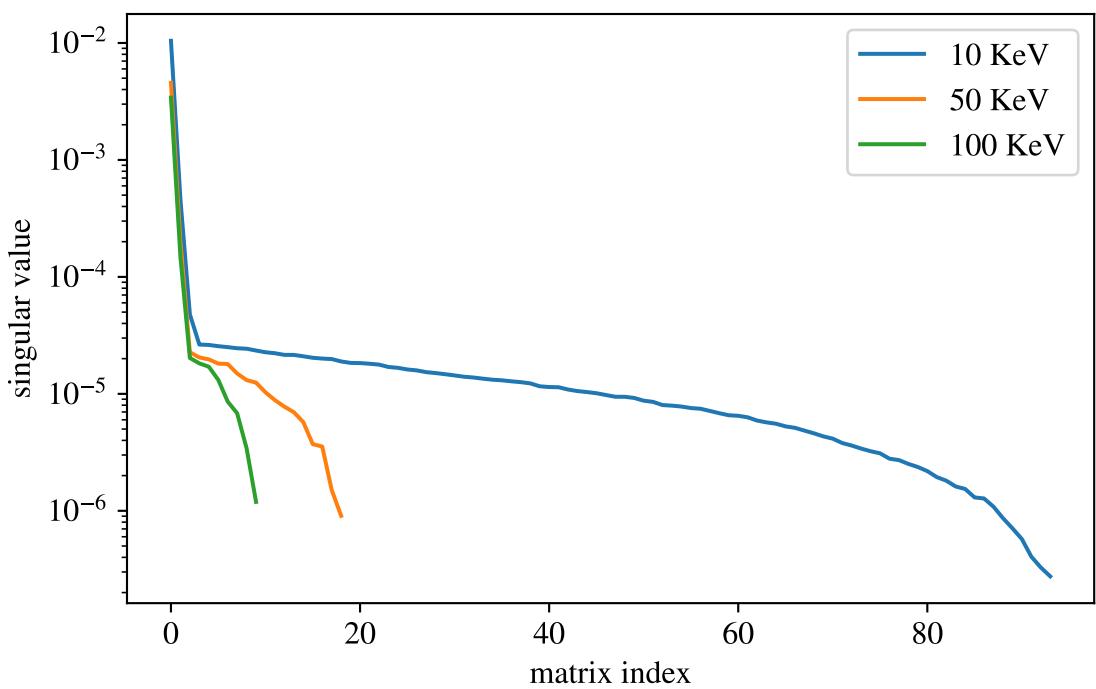


Figure 4.5: Plots of the singular values of the kernel matrix for the X-ray inversion problem for different electron spectrum bin widths, with singular values arranged from biggest to smallest.

graphing the condition number of the kernel matrix as a function of truncation index. This is shown as a function of binning scheme in Figure 4.6.

Compared to Figure 4.2, the plots in Figure 4.6 show that truncating the smallest singular values of the kernel matrix has a positive effect on the conditioning of the X-ray inversion problem. The number of binning schemes which produce a singular kernel matrix, with no inverse, is reduced with increasing truncation index. In addition, the regions with low condition numbers where the inverse problem is tractable (less than 10^3) are expanded.

The improved conditioning of the inverse problem by truncating small singular values comes at a significant cost by introducing solution bias. There is no simple method to describe the amount of bias this process introduces, or the effect it will have on a particular data set, however, we can apply the least squares inversion to sample data, as in 4.1, using the truncated kernel and see what effect the bias has.

Before applying a biased estimator to the inversion problem, there is an important fact which must be made explicit. The main symptom of an ill-conditioned problem is a high sensitivity to small perturbations on the input data. Because of this, one obtains almost no information about the overall behaviour of a given estimator on the X-ray inversion problem by running single tests on synthetic data (which can be generated using the techniques of Chapter 3.) X-ray measurements are counting experiments, and they are subject to statistical noise in their measurements. The statistical noise will follow a Poisson distribution in a well-designed experiment without significant uncorrected detector bias. Since the data have a random component, it would be an error in methodology to attempt to solve the X-ray inversion problem on one specific instance of synthetically generated data. Rather, one should sample a given synthetic data set many times, and then use the inversion technique of choice to see the resulting distribution of computed electron spectra.

This technique is applied to synthetic electron spectra in Figure 4.7. Exponential electron distributions and their resulting X-ray spectra are simulated, and then the X-ray inversion problem is

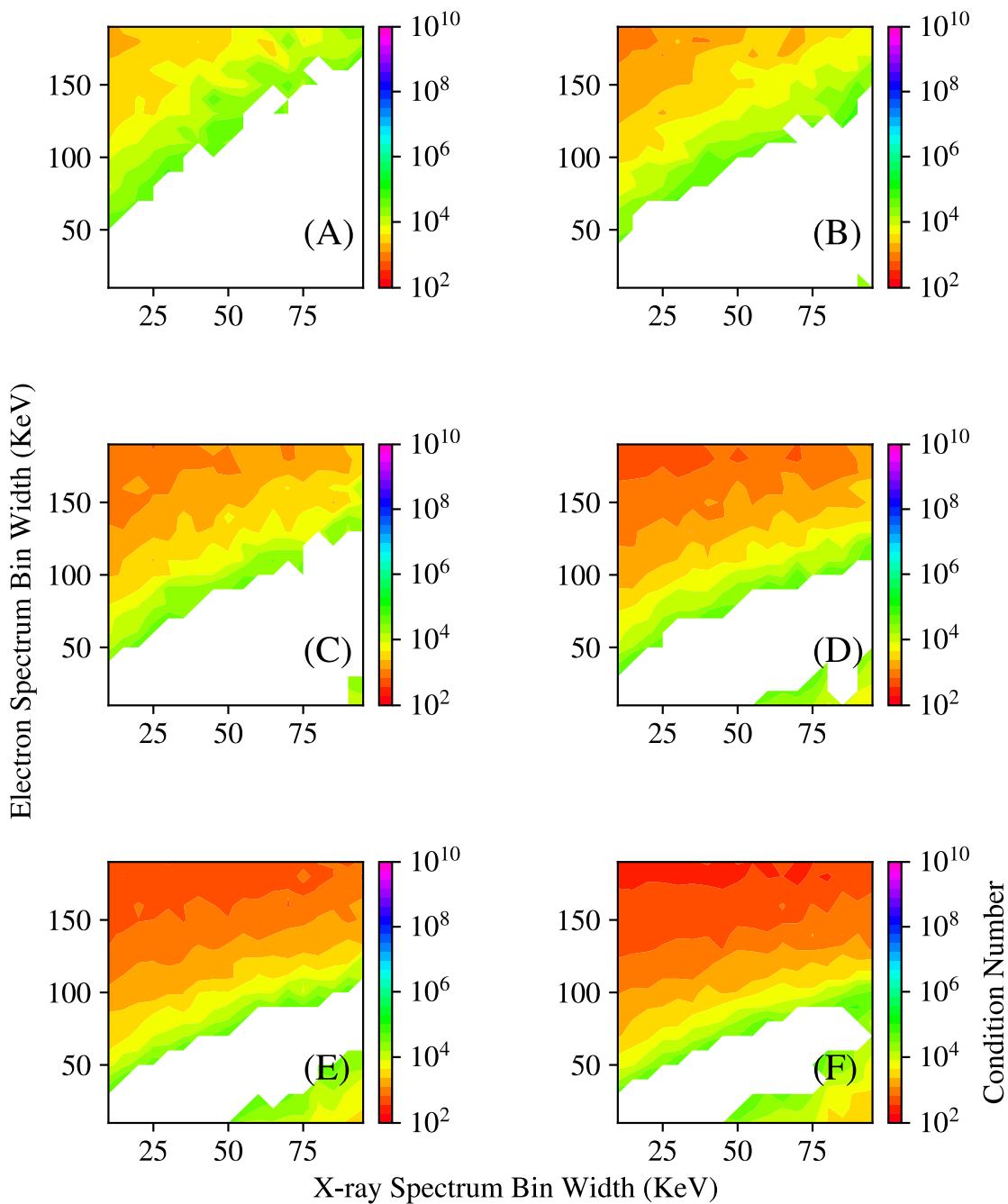


Figure 4.6: Plots of the condition number of the kernel matrix for the X-ray inversion problem as a function of electron spectrum and X-ray spectrum bin width, for 1 (A) through 6 (F) truncated singular values.

solved using the truncated singular value decomposition and least-squares regression for different truncation indices. The stabilizing effect the truncated SVD approximation has on solutions to the X-ray inversion problem is apparent, though, it is important to note that the spread in the reconstructed spectra due to statistical noise in the simulated X-ray measurements remains significant. In a real experiment, one obtains only one measurement of the X-ray spectrum and the underlying statistical distribution in the noise. This presents a significant problem for reconstructions based on biased estimators, since there is no a-priori way to predict how their characteristic bias will map the noise statistics in measurement space to model space. Despite this problem, and the associated difficulty assigning meaningful error estimates to reconstructed spectra, Figure 4.7 motivates the search among biased estimators for solutions to the X-ray inversion problem. It is entirely possible, for example, that a measured X-ray spectrum has statistics which only support a binning scheme which makes the inverse problem impossible. A biased estimator, such as truncated singular value decomposition, which effectively improves solution resistance to noise, can be the only way to recover a useful model in these scenarios.

The use of the truncated singular value decomposition as a biased estimator introduces the truncation index as a free parameter. This leads to an important question: when is the problem well-conditioned enough for the data set being analysed? Higher degrees of truncation improve the conditioning of the X-ray inversion problem (Figure 4.5), but only at the cost of less information being represented in the inverse mapping. The introduction of a free parameter which controls the bias-variance trade-off is a general feature of biased estimators used to solve inverse problems. This too represents a cost, since the choice of this free parameter must be somehow justified. The question then becomes whether there is a method to choose this parameter based on data from a given experiment. This question is central to the X-ray inversion problem, and is the focus of the remainder of this chapter.

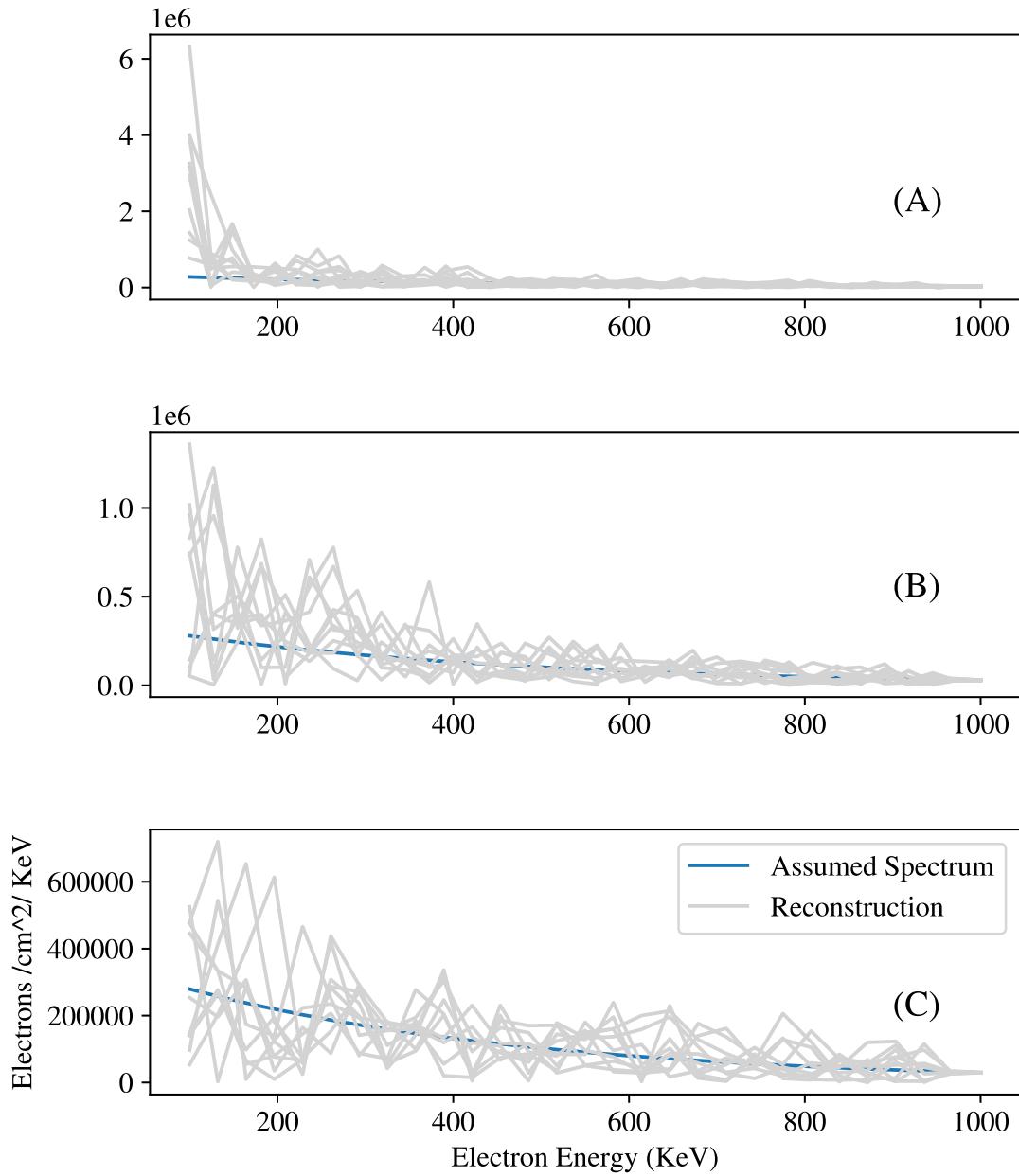


Figure 4.7: Plots of simulated electron spectra for a total electron flux of 1×10^9 electrons/cm²/second, and their reconstruction from the associated simulated X-ray spectra using a truncated singular value approximation of the kernel matrix. Electron bin widths are set to 50 KeV, and X-ray bin widths to 1 KeV. (A), (B), and (C) show reconstructions with the smallest 1, 5, and 10 singular values dropped from the expansion.

4.6 Tikhonov Regularization and the L-Curve Method

The truncated singular value decomposition explored in the last section is not the only method which has been developed to recondition inverse problems. For the optimization problems that arise from the introduction of a free parameter with biased estimators, it is desirable to have a continuous variable which controls the degree of reconditioning, rather than a discrete truncation index. Further, it is useful to have an estimator which exhibits a bias towards certain families of solutions, using a-priori knowledge we might have about the shape of the expected solution to the inverse problem. The method of Tikhonov regularization has these properties (Tikhonov et al., 1995).

In Tikhonov regularization, the unbiased least-squares regression:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|$$

is replaced with the modified version:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\| + \alpha \|\mathbf{x}\|.$$

The parameter α is a positive real number which controls the strength of the regularization. When α is zero, the least-squares solution is recovered. The addition of $\alpha \|\mathbf{x}\|$ to the least-squares minimization represents a penalty term on the size of the solution norm. There is a close relationship between the effect of adding this penalty term and the suppression of small singular values in the kernel matrix. Write the singular value decomposition of the kernel matrix as:

$$\mathbf{G} = \mathbf{U}\Sigma\mathbf{V}^T.$$

It can be shown (Hansen et al., 2006) that the solution to the Tikhonov regularization can be written as:

$$\mathbf{x} = \mathbf{V}\mathbf{D}\mathbf{U}^T \mathbf{b}$$

where the diagonal values of \mathbf{D} are:

$$\mathbf{D}_{ii} = \frac{\sigma_i}{\sigma_i^2 + \alpha^2}$$

where σ_i are the singular values of \mathbf{G} . The regularization term applies a filtering effect which smoothly suppresses the smallest singular values in the kernel matrix. This effect improves the condition number of the kernel matrix, and stabilizes solutions to the inverse problem. This is accomplished at the cost of a bias towards solutions with a small norm, which scales with the degree of regularization applied.

Tikhonov regularization has a heuristic interpretation. The deviation from a solution with a small norm can be viewed as a measure of its complexity, and the amount of information which it contains. The trade-off between solutions to inverse problems with stability but high bias, and solutions with small residual errors but high variance, is controlled by the regularization parameter α . There is no single method to choose an appropriate value for α for a given problem. Techniques to select α can include heuristic arguments, a-priori knowledge based on the properties of the particular problem at hand, and statistical arguments. Combinations of these methods are also possible, but all of them represent an implicit cost to the analysis of the problem at hand, namely, that a choice in α ultimately needs to be made, and that choice will always depend, to an extent, on the preferences of the analyst. This is a danger to the analysis of the X-ray inversion problem, which we will mitigate through careful testing on synthetic data across multiple techniques, and objective scoring of different techniques using clearly defined metrics.

The classic method for choosing α is the so-called “L-curve” (Hanke, 1996; Hansen, 1992; Hansen and O’Leary, 1993). This heuristic method works by plotting the solution norm $\|\mathbf{x}\|$ vs the error term $\|\mathbf{Ax} - \mathbf{b}\|$ in log-log space. Typically, but not always, this plot has a characteristic L shape, with an inflection point which locates the value for α that produces the most significant reduction in solution norm for the smallest error term. If the error term is thought of as the cost of solution stability, then the inflection point locates the most economical value of α .

An example of this applied to the X-ray inversion problem is shown in Figure 4.8. I use a synthetic X-ray spectrum based on an exponential beam of electrons with a folding energy of 300 KeV. The L-curve is shown and used to select $\alpha = 2.0 \times 10^{-5}$. The assumed electron spectrum is plotted along with spectra determined using direct matrix inversion, and Tikhonov regularization for the selected value of α . Multiple runs across synthetic data generated with different random seeds are used to provide an accurate representation of the variance in each solution.

The reduction in variance gained from the regularization process is evident in Figure 4.8. The spread in solutions using direct inversion is large enough, that besides a general trend towards a decrease with lower energies, few other features are readily apparent. Additionally, some of the direct solutions show negative particle fluxes in the lower energy bins. Without careful analysis of the variation in solutions caused by the counting statistics, inversions attempted using experimental data could even be misleading, showing apparently significant features which are in fact only caused by the mapping of random noise through the ill-conditioned kernel matrix of the problem.

The bias introduced by regularization is also apparent in Figure 4.8. Towards lower energies the solution trends downward, which is a feature the assumed electron spectrum does not have. The spectrum retrieved using regularization is also not fully exponential, and a folding energy retrieved by fitting an exponential curve would have a non-zero error when compared with the assumed data. Although the variance in the direct solution is large enough to be untenable when applied to experimental data, it remains exponential, and on average, the folding energy will match the assumed model.

Unfortunately, there is no general method to assign useful confidence intervals to solutions from biased estimators, including those generated from Tikhonov regularization (Goeman et al., 2012). The effect of the variance in solutions is simple to capture, as done in Figure 4.8 by running the inversion multiple times under different random seeds, but a true confidence interval would include the effects of both the bias and the variance in the solution. The use of a biased estimator such as Tikhonov regularization makes an otherwise unsolvable problem tractable, but only at the

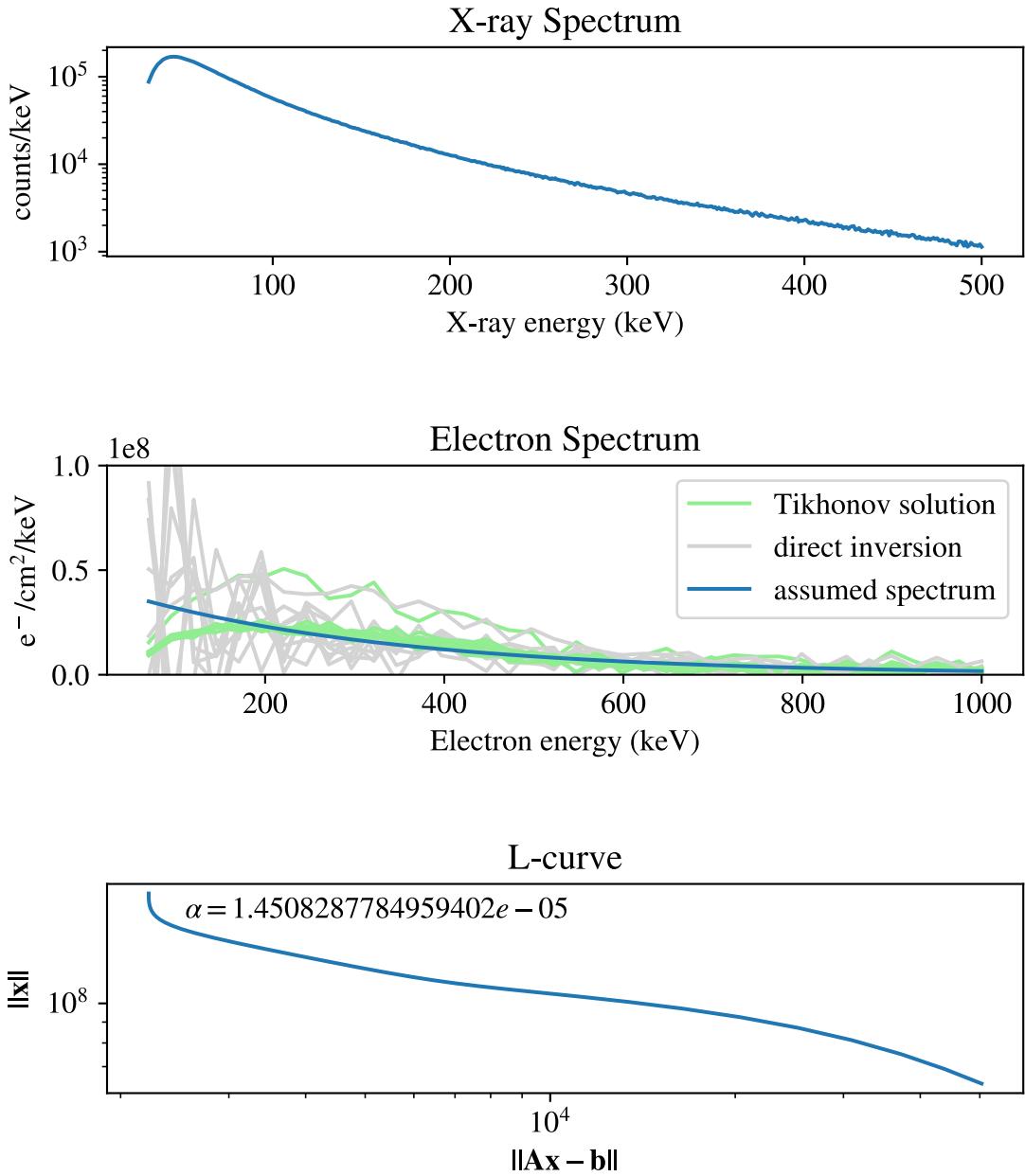


Figure 4.8: Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The X-ray spectrum corresponding to the assumed electron spectrum is shown in the top panel. The middle panel shows the assumed electron spectrum in blue, with retrieved spectra in grey (direct inversion) and green (Tikhonov regularization). The L-curve and selected value for α are shown in the bottom panel.

cost of the certainty that can be placed on the solutions. In a sense, an ill-posed problem represents a mismatch between the amount of information contained in the solution, and the amount of information in the available data. Tikhonov regularization allows us to control the manifestation of this fundamental lack of information, but there is no process by which it can be eliminated.

Tikhonov regularization has a critical property which the unbiased methods lack - namely that it is *adaptive*. The method works by balancing the variance in the solution against the norm, which means that as data become noisy and harder to resolve, the bias is increased to compensate and maintain stability. This is in sharp contrast to, for example, choosing a binning scheme or parameterization which remains fixed for a given problem. Some data sets are “easier” to invert through a given kernel matrix than others, which can be for a multitude of reasons. For example, the data set could represent a monochromatic solution, which incurs the least variance-inducing resistance in the inversions. Tikhonov regularization is sensitive to this fact, and the L-curve will have a shape which corresponds to the particular problem and data set being used.

The L-curve method is computationally expensive, since solutions need to be generated for every value of α being examined. For relatively small matrices, say, less than 1000 by 1000 rows and columns, this is a small problem, but it becomes significant as the problem being examined scales. Because of this, when using Tikhonov regularization, the most sensible binning scheme for the X-ray inversion problem is not necessarily the one which provides the finest resolution in X-ray spectra, as Figure 4.2 would suggest. Practical considerations based on computational limits need to be taken into account.

The fact that the L-curve method is somewhat heuristic motivates us to look for another way to select a value for α . The “corner” of the L-curve plot, if it appears, is not always sharp. This represents a way that the expectations of the human doing the analysis can enter into the solutions created when applying Tikhonov regularization. The potential deficiencies of the L-curve method are examined in depth by Hanke (1996); Hansen and O’Leary (1993), and Vogel (1996). There is an alternative family of related methods which can be used to select α automatically, based on

the statistical properties of the data alone. Cross-validation methods (Golub et al., 1979) use the idea that the artificial removal of parts of the input data produce different solutions and can give information about the stability of the problem being analysed. The hypothesis is that the effect of the removal of a single data point, such as a single energy bin in the X-ray inversion problem, should have only a small effect on the retrieved electron spectrum when the correct degree of regularization has been applied. Iterating over the available energy bins, and suppressing each one from the problem in turn, will produce a spread of solutions, whose variation gives a picture of the overall stability of the solution. If this entire process is then repeated over different values of α , then the total mean squared error between the suppressed data points and the regularized solutions can be compared and numerically minimized. This process is motivated with a sample of data in Figure 4.9. In both the over-fit and under-fit cases, the missing data point is represented poorly in the solutions. Minimizing this effect represents a balance between these cases and produces a corresponding value for α .

Iteration over a single suppressed data point is called leave one out cross-validation. It is also possible to leave out any number of data points up to $n - 1$, where n is the total number of points in the data set. For leave n out cross validation, one evaluates the solutions generated by every possible way to suppress n data points. This is computationally expensive.

In Figure 4.10, the synthetic data set from Figure 4.8 is reproduced, but α is selected using leave-one-out cross validation. Minimizing the mean-squared error using the cross-validation method suggests $\alpha = 7.8 \times 10^{-2}$. This is a different value from what was used in Figure 4.8, but the retrieved spectrum is approximately the same. This suggests a degree of stability across values of α for the application of Tikhonov regularization to the X-ray inversion problem, but testing across many data sets is required for verification. This will be done at the end of this chapter, where the different regularization methods are compared across a set of synthetic data sets.

Unlike the “L-curve” method, cross-validation leaves no room for the human analysing the problem to have input on the regularization parameter α . Instead of looking for an inflection point

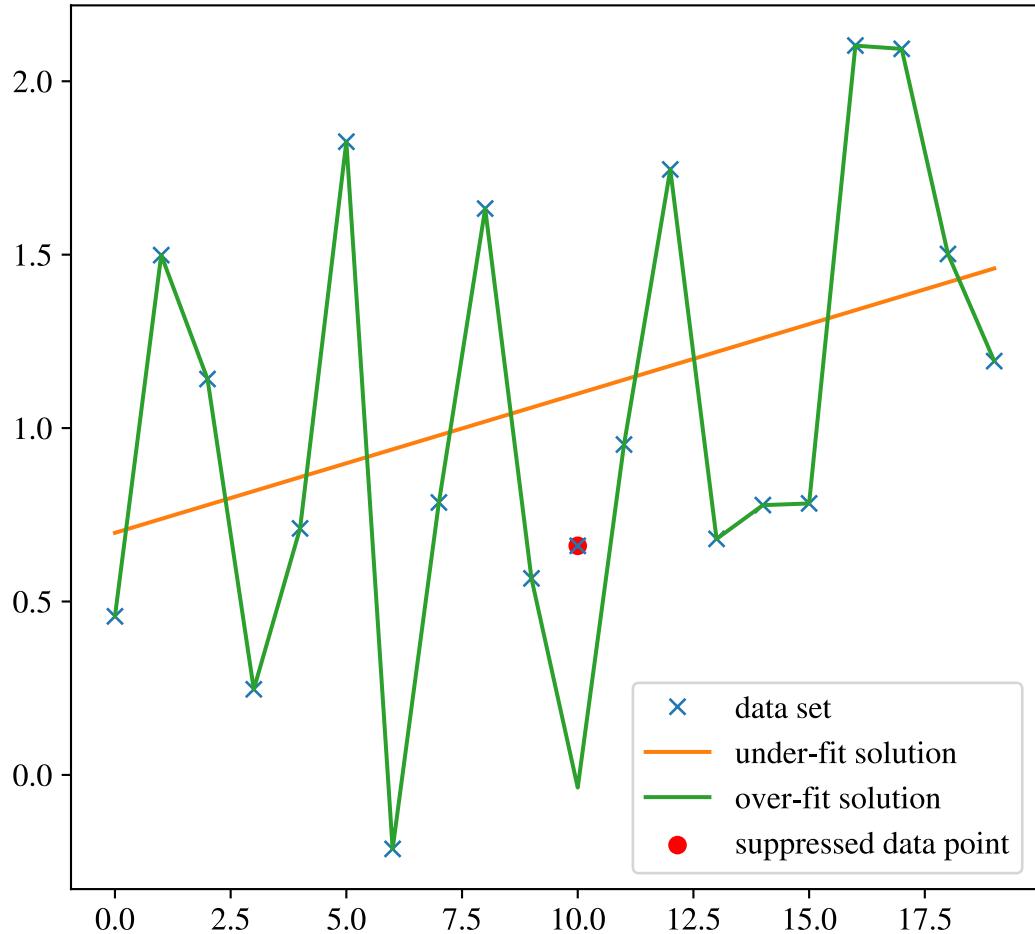
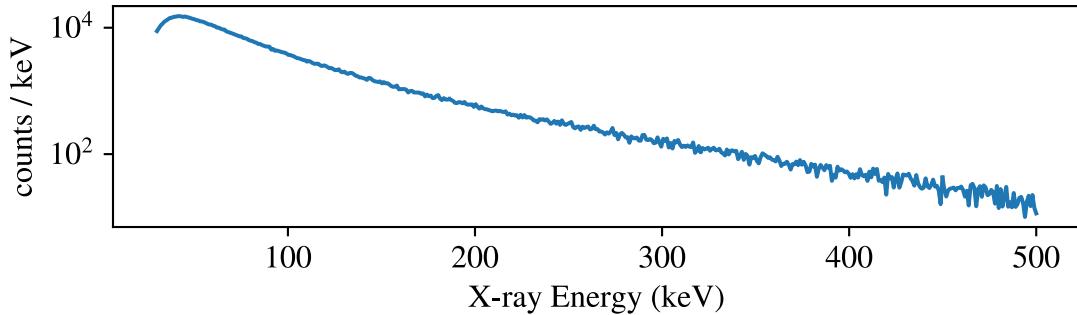


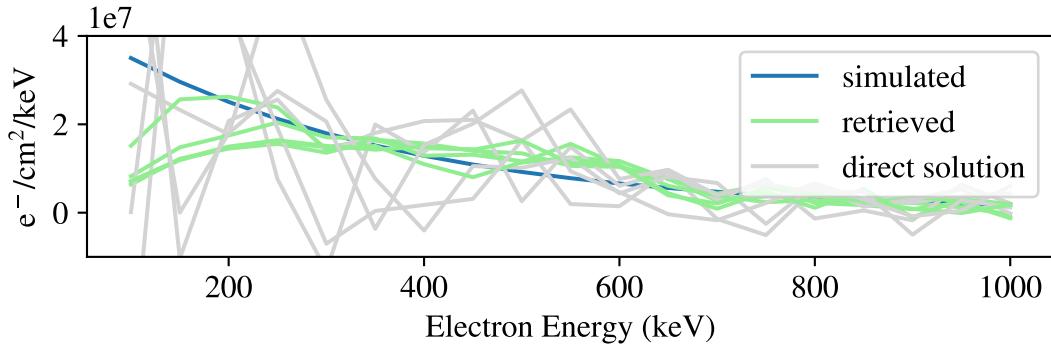
Figure 4.9: Random data set (blue) with under-fit (orange), and over-fit (green) representations. A particular data point (red) is removed prior to fitting. The removed data point is far from both the over-fit and under-fit representations of the data.

on a log-log graph, which has some room for interpretation, the minimization of squared errors across values of α for cross-validation is carried out numerically. Since the minimization process requires no human intervention, it can be applied to large bodies of synthetic data automatically. This is useful both for running on synthetic data for testing, as well as different sets of experimental data. Since the objective characterization of the performance of an inversion technique to the X-ray inversion problem will depend on its application across to synthetic data, this strength is important.

Simulated X-ray Spectrum



Simulated Electron Spectrum and Retrieval



LOOCV Squared Residuals

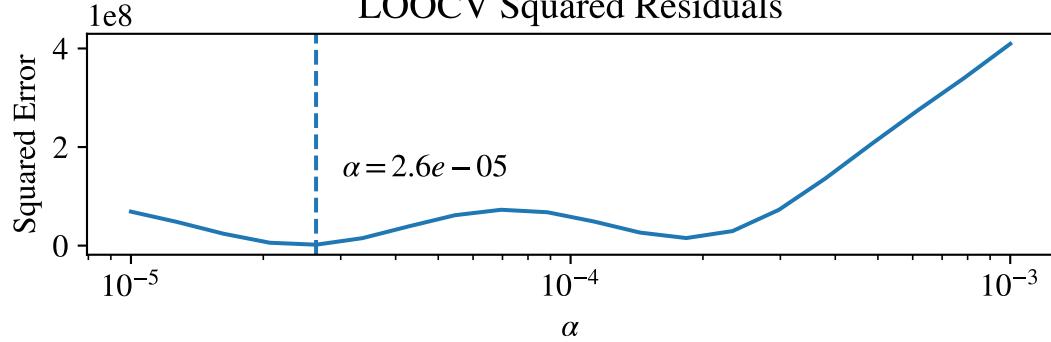


Figure 4.10: Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).

4.7 Higher Order Regularization

The process of Tikhonov regularization can be interpreted as the addition of a-priori information to the problem (Tikhonov et al., 1995). In Tikhonov regularization, the modified least-squares problem:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\| + \alpha \|\mathbf{x}\|$$

has a penalty term $\alpha \|\mathbf{x}\|$ which introduces a preference for solutions with a smaller norm. This was motivated by the fact that more complex solutions typically have a larger number of non-zero terms than simpler solutions, and so, to introduce stability, the strength of the regularization should increase with the solution norm, which is thought of as corresponding to the amount of information in the solution. This was indicated graphically through the L-curve method. There are other forms for the penalty term, which can be chosen to align with the properties of physically reasonable solutions to the problem at hand. This section will examine these higher-order regularization terms as applied to the X-ray inversion problem.

Write a new least-squares problem as:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\| + \alpha \|\mathbf{Lx}\|$$

where \mathbf{L} is a linear operator. If \mathbf{L} is the identity matrix, then the ordinary regularization is recovered. This is termed zeroth-order Tikhonov regularization. In first-order Tikhonov regularization, \mathbf{L} is chosen to be the matrix which approximates the first derivative operator:

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & \dots \\ 0 & 1 & -1 & \dots \\ \vdots & \ddots & & \\ 0 & \dots & 1 & -1 \end{bmatrix}$$

Unlike zeroth-order regularization, which prefers small solutions, this choice of \mathbf{L} penalizes solutions which change - the preference is for constant solutions. For problems where we have an a-priori expectation that solutions are constant, this operator would be the clear choice. A useful and interesting feature of first-order regularization is that no advance knowledge of the best constant solution needs to be applied.

For the X-ray inversion problem, we do not have any strong expectation that solutions will be constant. In fact, it is more reasonable that electron fluxes decrease as a function of energy. This is motivated by both satellite observations, and balloon flights where models of exponentially decreasing electron spectra tend to best describe measured X-ray data. This motivates the search for a different penalty term for the regularization problem, which better matches this expectation.

Second-order Tikhonov regularization uses a linear operator which approximates the second derivative:

$$\mathbf{L} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ \vdots & \ddots & & & \\ 0 & \dots & 1 & -2 & 1 \end{bmatrix}$$

This operator introduces a preference for smooth solutions. There is an intuitive reason why using this operator for the X-ray inversion problem might be useful - as the noise in the measured X-ray spectrum increases, the number of features detectable in the model electron spectrum will go down. A relevant example would be an electron spectrum which is well-described by two beams at different energies. For low noise levels in the corresponding measured X-ray spectrum, the two beams can be resolved through direct inversion, without applying regularization. As the noise in the measured spectrum is increased, second-order regularization will dull and blend the two sharp peaks until eventually they are not distinguishable. This is a desirable property, since the bias tends towards solutions which reflect the lack of information a high noise measurement represents.

This is illustrated in Figure 4.11, where a test electron spectrum consisting of two mono-energetic beams is assumed, and the resulting simulated X-ray spectra are inverted under different assumed noise levels, corresponding to gaussian noise consistent with total counts of electrons.

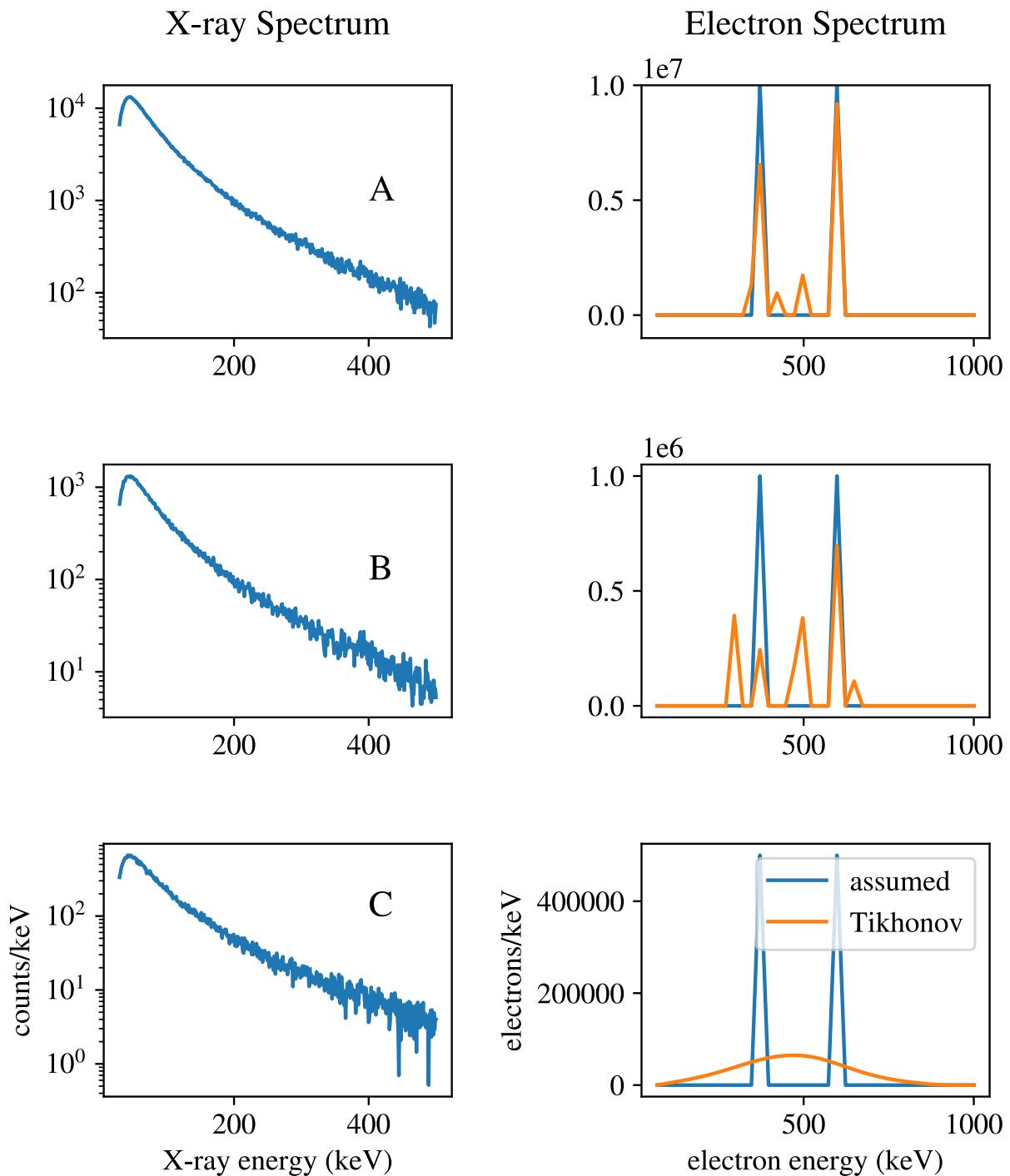


Figure 4.11: Application of second-order Tikhonov regularization to synthetic X-ray measurements corresponding to two mono-energetic electron beams at different X-ray fluxes and noise levels (A - 10^7 electrons / keV, B - 10^6 electrons / keV, C - 5×10^5 electrons / keV)

An additional argument for applying second-order Tikhonov regularization to the X-ray inversion problem is that completely sharp, mono-energetic features aren't realistic in natural electron spectra. Thermal processes tend to relax such distributions to have a characteristic width, and over a long enough time period, the central limit theorem tends them towards Gaussian distributions regardless. Since sharp solutions aren't generally expected, it makes sense to choose \mathbf{L} such that they are penalized. This represents further injection of a-priori information to the problem.

4.8 Preconditioning Techniques

The application of regularization techniques makes the X-ray inversion problem tractable, with realistic and imperfectly measured input data. In using a biased estimator, we have replaced an impossible problem with a condition number approaching 10^{10} , with a nearby problem that is better behaved. The cost of this is an artificial tendency towards families of solutions influenced by a-priori information, such as smooth solutions, for second-order Tikhonov regularization. The effect of this bias increases with noise in the input data. In this section, we explore methods which can be used to apply simplifying transformations to the X-ray inversion problem prior to the regularization process. Provided that the transformations applied have an inverse, they can be used to re-cast the X-ray inversion problem as an equivalent linear problem in a different, and more orthogonal, space. If the transformations are chosen carefully, the equivalent problem can have a condition number which is significantly reduced. This reduces the amount of regularization that needs to be applied for stability, and allows more useful information to be obtained from noisy measurement data.

As a motivating example, take the linear problem $\mathbf{A}\mathbf{x} = \mathbf{b}$, with \mathbf{A} defined as:

$$\mathbf{A} = \begin{bmatrix} .0001 & 1.0001 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The condition number of \mathbf{A} is approximately 2.0×10^4 . Instead of calculating \mathbf{A}^{-1} , first left and right multiply by linear transformation:

$$\mathbf{L} = \begin{bmatrix} 1 \times 10^4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The new matrix is:

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 1.001 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with a condition number of 2.62. The same transformation can be applied to \mathbf{b} , and the transformed problem is then solved to obtain the solution, \mathbf{x} , of the original problem. This preconditioning can be applied before using regularization methods, reducing the amount of regularization required for stability with a particular problem and data set.

In general, the left and right-multiplication can be done using different matrices (Bouwmeester et al., 2015). Label the left preconditioner \mathbf{L} and the right preconditioner \mathbf{R} . The linear problem:

$$\mathbf{Ax} = \mathbf{b}$$

can then be recast as:

$$(\mathbf{L}\mathbf{A}\mathbf{R}^{-1}\mathbf{R})\mathbf{x} = \mathbf{L}\mathbf{b}.$$

Since the X-ray inversion problem is linear, there is no inherent cost to applying preconditioning methods before solutions are found. Provided matrices \mathbf{L} and \mathbf{R} are not singular, the solution to the original problem can be recovered.

There is no known general method to determine optimal values for \mathbf{L} and \mathbf{R} . There are some specific cases, such as for symmetric matrices, where techniques exist, but the matrix which describes the X-ray inversion problem does not fall into any of these categories. This problem makes finding an appropriate preconditioner for the X-ray inversion problem non-trivial. We will try different combinations of known preconditioners, and stop when a selection is found that significantly reduces the condition number of the problem.

From a computational perspective, the ideal matrix representing the X-ray inversion problem would be the identity, with a condition number of one. Compared to the identity matrix, the real matrix representing the X-ray inversion problem has two main undesirable properties. The first is that the rows, which represent X-ray spectra from mono-energetic beams of electrons, have a high degree of collinearity. This is the main cause of the high condition number of the matrix. The second problem is that the rows and columns have vastly different norms. This is because low energy beams of electrons do not penetrate far into the atmosphere, and their impact in terms of measurable X-rays is small. On the other hand, high energy beams of electrons need only have a small total particle flux to create a very large measurable X-ray response. Figure 4.12 (top) shows this for a particular binning scheme on the response matrix though graphs of its rows. The spectrum of singular values is also shown. We will show that preconditioning can have a significant effect on the singular values of the resulting matrix.

One left-preconditioner is diagonal scaling (Pini and Gambolati, 1990). A diagonal matrix is constructed such that the entries are equal to the inverse of the norm of the rows of the matrix representing the linear problem being solved. This has the effect of making the row norms of the transformed matrix closer. Figure 4.12 (bottom) shows the effect of applying this preconditioner. The row norms and the spectra of singular values span fewer orders of magnitude. For this example, diagonal scaling has improved the response condition number from a value of 1.6×10^4 to 1.3×10^4 . The effect of the diagonal scaling preconditioner can also be seen across different binning schemes on the X-ray response matrix. The effect is not constant across binning schemes. This is

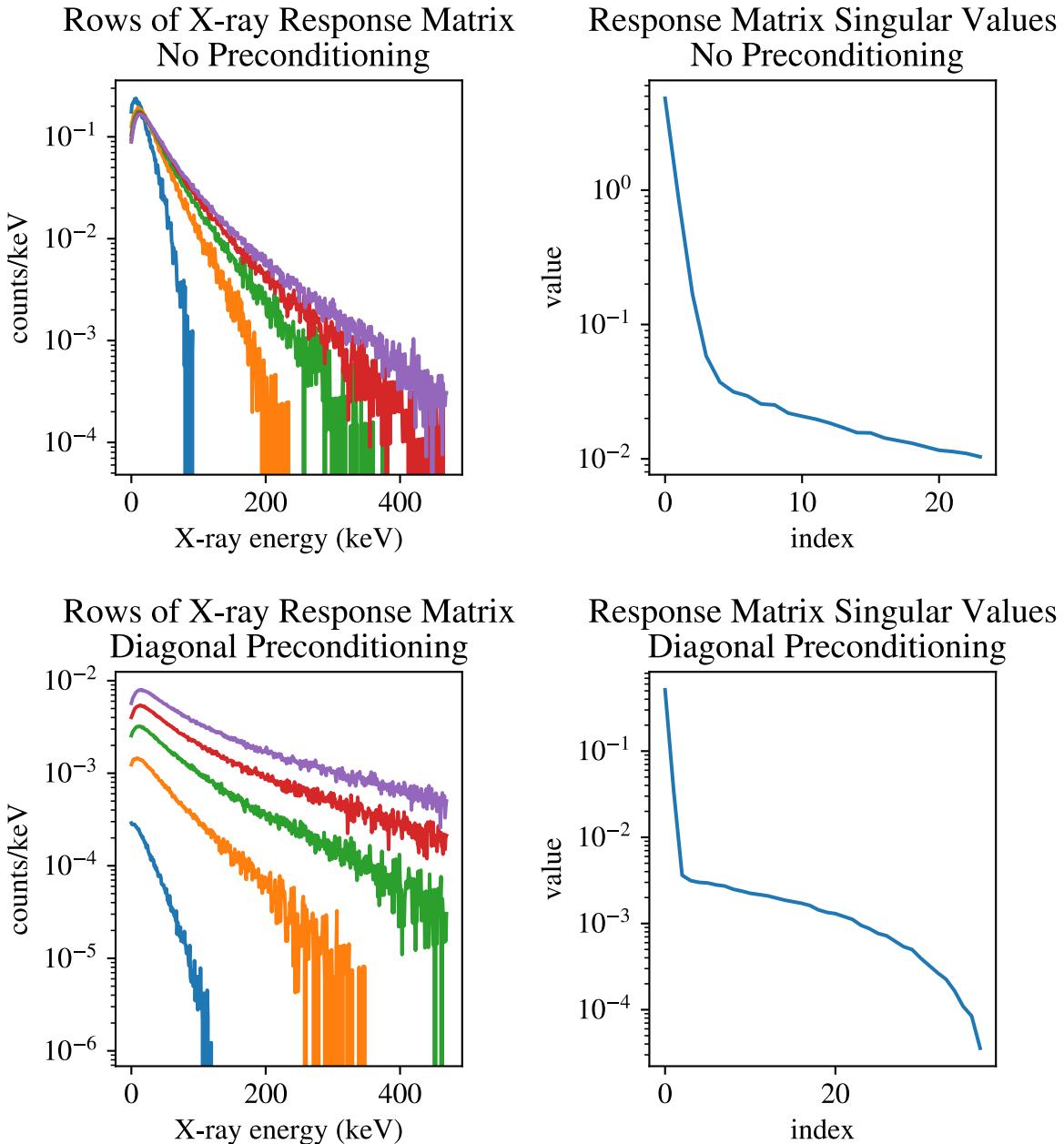


Figure 4.12: (left): Graph of the rows of the X-ray response matrix. A high degree of co-linearity is apparent, as with a large scale change between rows. (right): Graph of the singular values of the X-ray response matrix, from largest to smallest.

shown in Figure 4.13.

Calculating the condition number of a matrix requires computing the singular value decomposition, which is a well optimized process and can be done very quickly on modern computers. There are different matrices that can be selected to improve the condition of the X-ray inversion problem. Since one can generally expect to lose k digits of precision finding the solution of a linear problem with a condition number of 10^k , the search for an optimal preconditioner for the problem at hand isn't very sensitive. Significant benefits will only be returned if order of magnitude improvements in the condition number can be found. This can be achieved through regularization, but comes at the cost of a bias which is difficult to quantify. Because of this, the best approach to the X-ray inversion problem is a combination of left and right preconditioners, chosen for a given selected binning scheme, combined with a regularization method. This dual approach obtains the full benefit of any preconditioners, with the stability, at a cost, which regularization provides. Since the regularization techniques are adaptive to both the noise in the measurement data and the inherent conditioning of the problem, the result is that regularization is used “as little as possible”, to obtain stability in the solution.

Choosing left and right preconditioning through optimization techniques generally works to reduce the condition number of the problem, but there are opportunities for it to fail, such as the numerical optimizer failing to converge. Additionally, there is a risk that the optimization adjusts preconditioning based on noise in the response matrix, which is generated through simulations, rather than the inherent shape of the response curves the matrix contains. The danger with this effect is that the collinearity of the component vectors of the response matrix is reduced only for synthetic data generated with the same response matrix. Real experimental data will not benefit from this reduced collinearity, since their noise statistics are independent of the response matrix.

An intuitive and useful pair of left and right preconditioners for the X-ray inversion problem can be created based on the observation that the response matrix is poorly scaled in both rows and columns. This is illustrated in Figure 4.14, which shows that the row and column norms

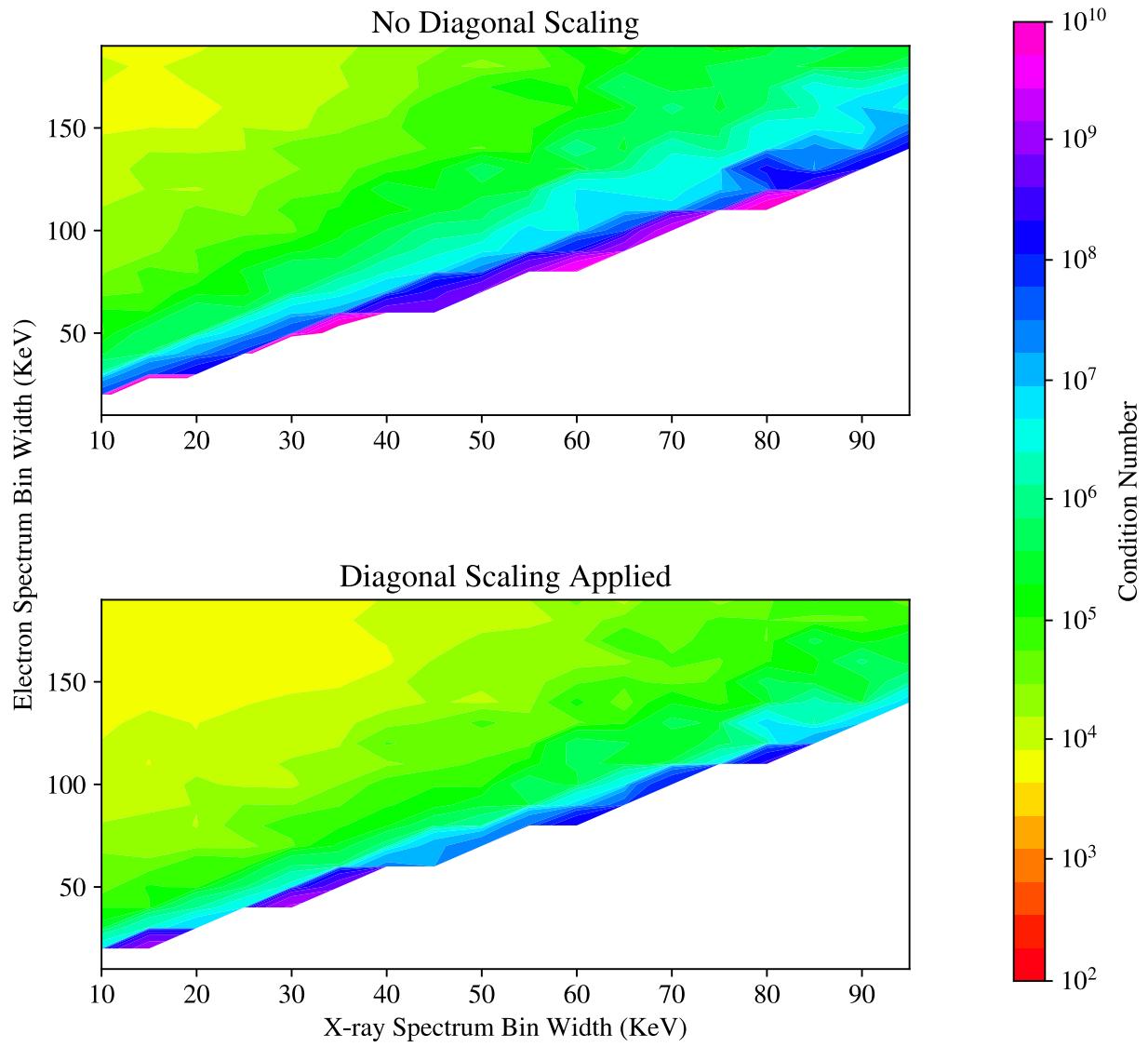


Figure 4.13: (top): Contour plot of the condition number of the X-ray response matrix as a function of electron spectrum and X-ray spectrum bin widths, with no preconditioner. (bottom): The same contour plot with diagonal scaling preconditioner applied.

differ through orders of magnitude across the response matrix. Motivated by this, we will try a left preconditioner which scales the rows of the response matrix, and a right-preconditioner which scales the columns. Both will be constructed as diagonal matrices, with dimensions compatible with the response matrix for a given binning scheme. The diagonal entries of the left and right preconditioning matrices will be the inverse of the row, and column norms of the response matrix, respectively. The effect of left and right preconditioning by these matrices on the response matrix is shown in Figure 4.15.

The effect of this preconditioning is dramatic. In Figure 4.2, the condition number of the response matrix was shown as a function of X-ray and electron spectrum bin-width. This is reproduced in Figure 4.16 for different combinations of these left and right preconditioners. As with the case with no preconditioning, there is a set of binning schemes where the response matrix becomes singular. When preconditioning is applied, however, the sensitivity of the problem to the binning scheme chosen is reduced. This effect is strongest when both the left and right preconditioners are applied. Surprisingly, even though the preconditioners chosen are simple diagonal matrices formed based on intuition, for sensible binning schemes, the X-ray inversion problem becomes completely tractable. This is in contrast to the situation before, where even the coarsest binning schemes produced an inverse problem so ill-conditioned that only the strongest regularization methods could provide stability. As before, this reconditioning comes without introduced bias, and numerically, costs only calculating multiplications and inverses of the preconditioning matrices to transform back to physical units.

4.9 Constrained Optimization

The fact that physical spectra are everywhere positive can easily be applied as a constraint to reduce the solution space of the X-ray inversion problem. Direct inversion, truncated singular value decomposition, and Tikhonov regularization can all be framed as a least-squares optimization problem. Numerical constraints on these problems, especially positivity constraints, are well-

known and solves are included in most modern computer algebra systems. There are two primary reasons for imposing positivity constraints on solutions:

1. Since spectra with negative components are non-physical, the interpretation of the negative regions is unclear. Simple approaches, such as setting them to zero, or taking the absolute value, are almost always misleading.
2. Reducing the solution space improves the conditioning of the overall problem.

The first reason exposes a danger when applying preconditioning transformations. The response matrix has a simple physical interpretation, since it maps from the space of electron spectra to the space of X-ray spectra. This interpretation is lost after preconditioning transformations are applied. Although all the same information is contained in the better-conditioned, transformed, matrix, it is stretched and distorted to a degree which is difficult to visualize. This means that a “small” negative region in the preconditioned solution might be significant when mapped back to physical units.

Fast, constrained optimization is achievable through several techniques. For the particular case of non-negative least-squares optimization, the LAPACK FORTRAN routines have been wrapped and made available in most modern programming languages. These routines are fast, and since the X-ray inversion problem has a small (less than a million) entries in the matrices which encode it, the addition of a non-negativity constraint adds almost no computation time. Figure 4.45 shows an example of synthetic X-ray spectrum corresponding to an exponential electron distribution. The X-ray spectrum is generated at noise levels consistent with different electron fluxes (10^9 , 10^8 , and 10^7 total electrons), and an inversion is attempted in each case with the non-negativity constraint applied. The left and right preconditioners from the last section are used with second order Tikhonov regularization to compute the assumed electron spectrum from the data.

4.10 Evaluation on Synthetic Data

The techniques shown have taken a difficult inverse problem and transformed it into a more tractable version. Preconditioning through linear transformations reduces the condition number of the problem dramatically, and the subsequent application of second-order regularization provides the rest of the needed stability to the problem. The bias introduced by regularization techniques, which is difficult to quantify, has been kept to a minimum through preconditioning and cross-validation, and chosen so as to gradually relax to smooth solutions in scenarios with high measurement noise. The addition of a non-negativity constraint ensures that solutions are always physical, by construction. What remains is to generate a body of test data and demonstrate that the combination of these techniques generates models which describe the assumed spectra in the test data well. The approach taken will be two-fold. First, simulated data for a set of different plausible forms for precipitating electron spectra will be generated, and then iterated over their respective parameter spaces (such as folding factor, for exponential spectra.) Results will be shown in plots, and different properties, such as non-negativity constraints, will be turned off and on to see changes in relative performance. Additionally, a set of randomly generated spectra will be created, and then different inversion methods will be run and error metrics put in histograms.

There is another way to show the stabilizing effects of regularization on the X-ray inversion problem. Following the approach used in Xu and Marshall (2019), electron spectra are generated by weighting each energy bin with a random number. The resulting X-ray spectra are computed, and then the inverse problem is solved. Instead of comparing assumed electron spectra with the obtained solutions, the effects of different regularization orders can be understood through histograms of the accumulated errors in the cross-validation process. This is shown in Figure ?? for zeroth, first, and second order regularization evaluated over 100 random electron spectra. The distributions of errors have the smallest expectation value, and the smallest variance, for second-order regularization.

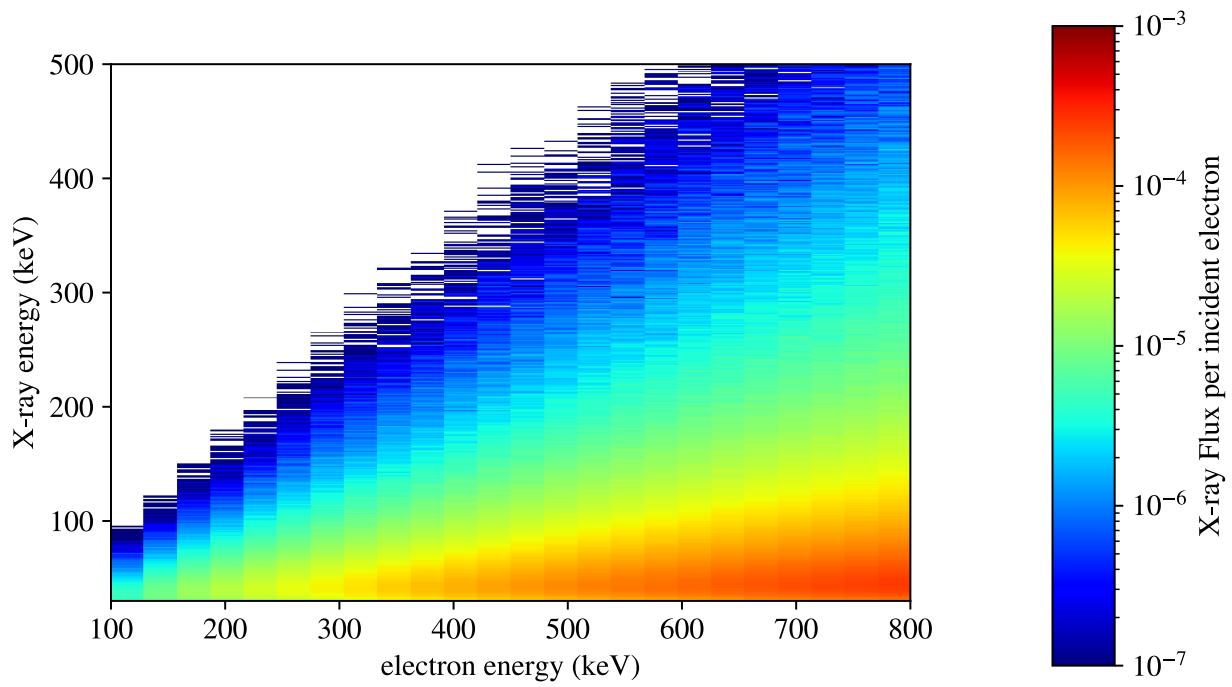


Figure 4.14: Response matrix for the X-ray inversion problem. Order of magnitude scale differences across both rows and columns of the matrix contribute to the high condition number. Units are normalized per incident electron.

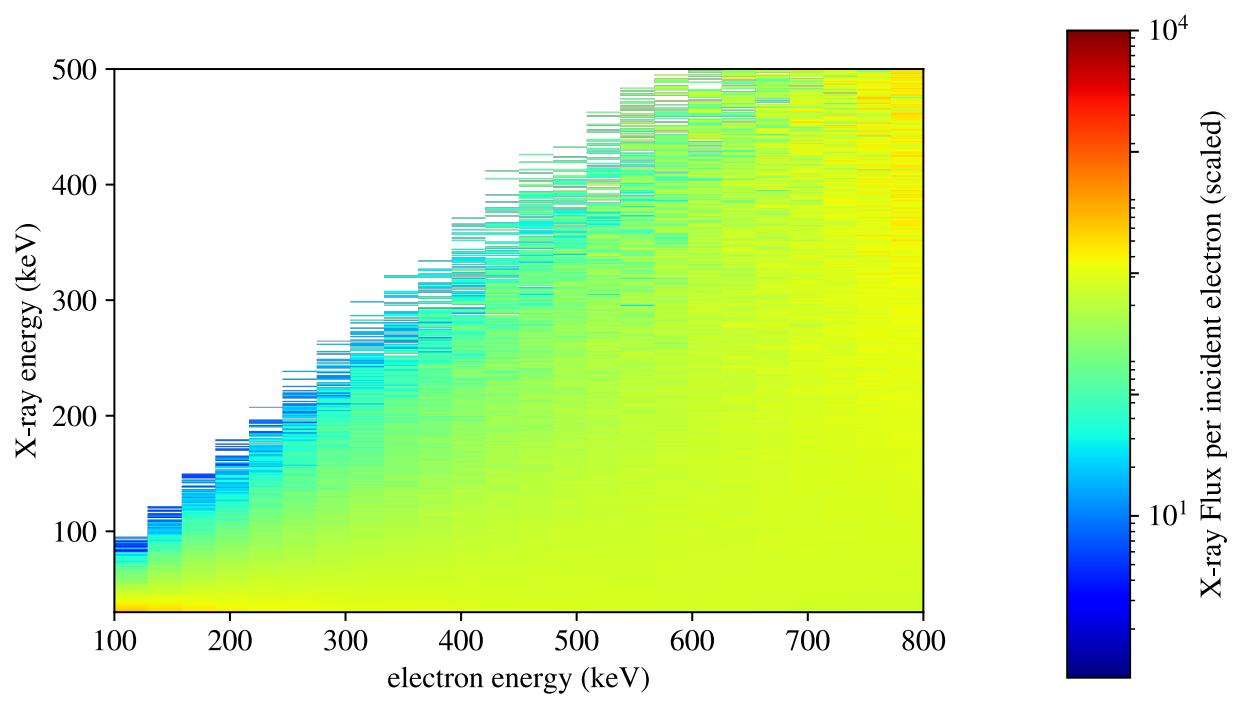


Figure 4.15: Response matrix for the X-ray inversion problem with left and right preconditioners applied. Scale differences across both rows and columns of the matrix have been reduced, improving the condition number. The units on the color axis are no longer physical, having been scaled by the preconditioning matrices.

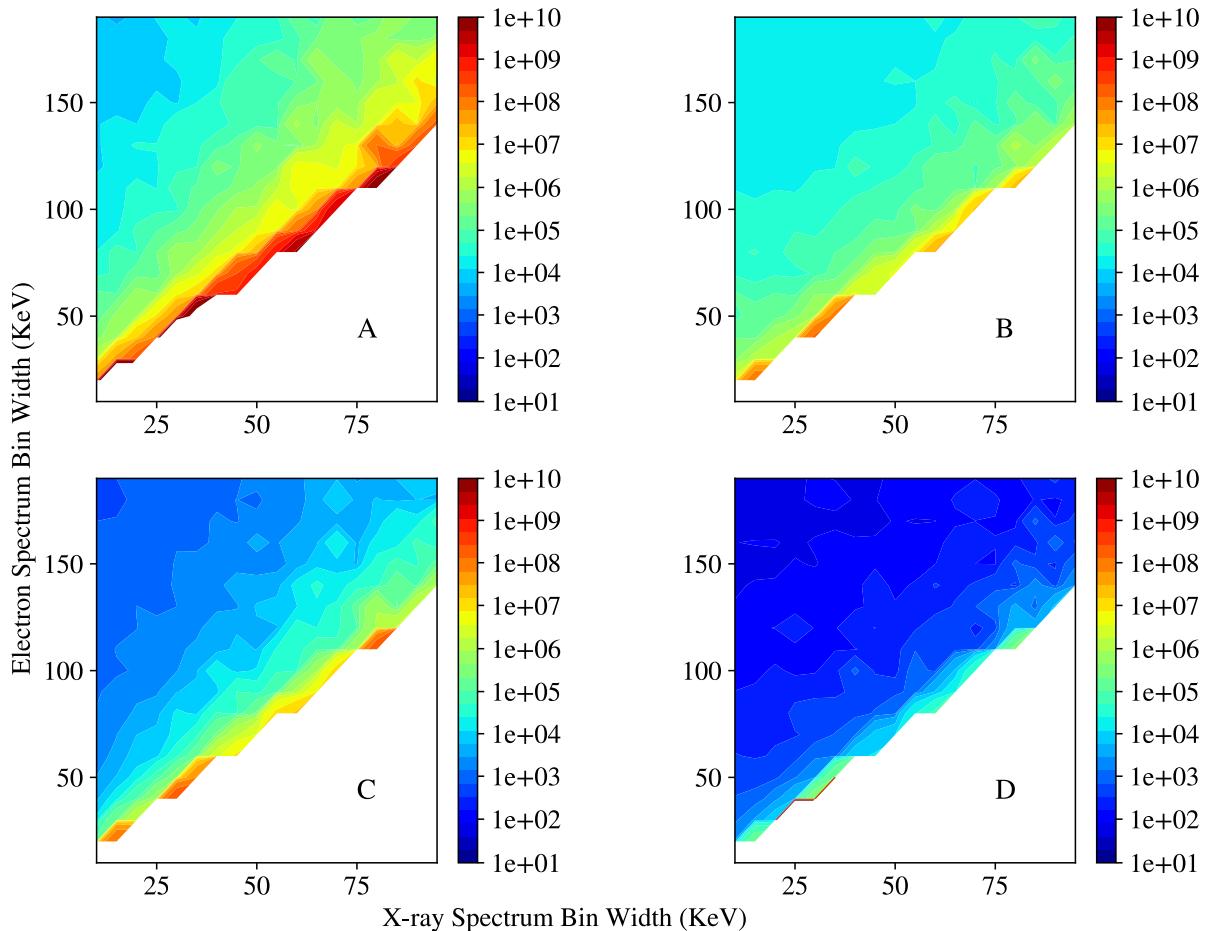


Figure 4.16: Response matrix condition numbers as a function of X-ray and electron spectrum bin widths for: (A) - no preconditioning, (B) - left scaling preconditioner only, (C) - right preconditioner only, (D) - both left and right preconditioners.

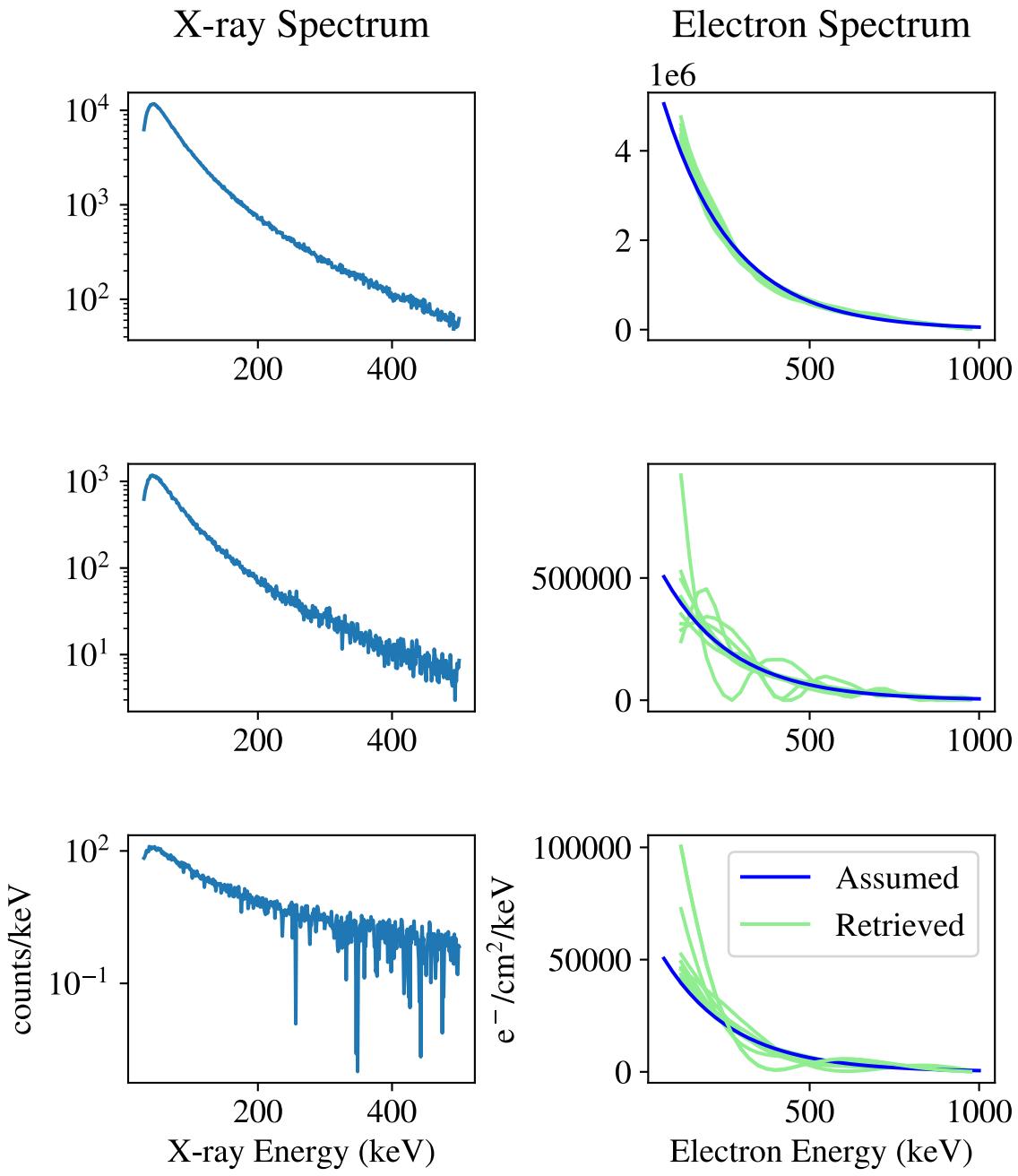


Figure 4.17: Retrieval of electron spectra from synthetic X-ray data across different simulated count rates. Left and right preconditioners, second-order regularization, and a non-negativity constraint are applied.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

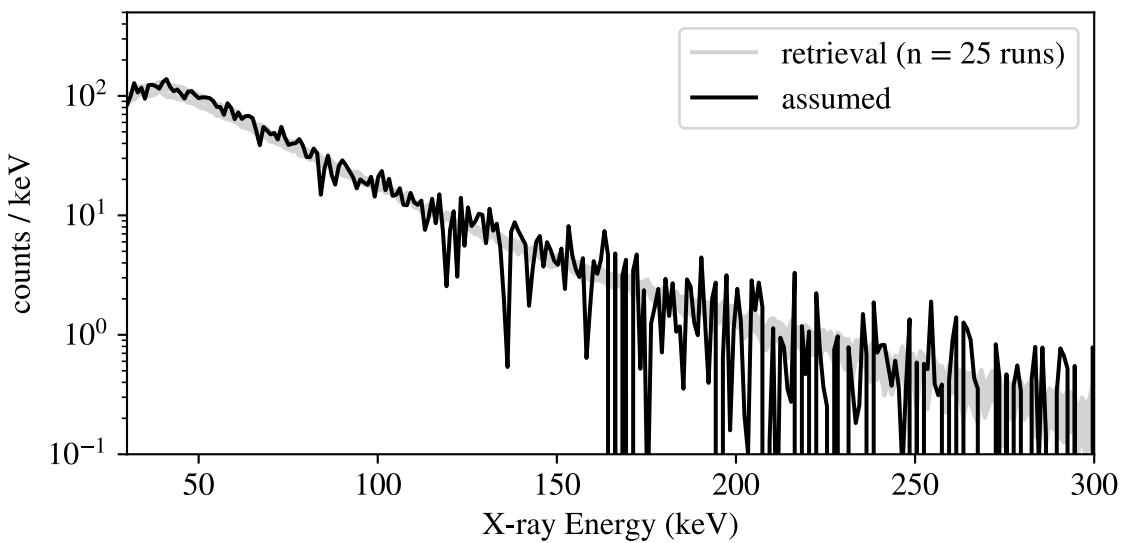
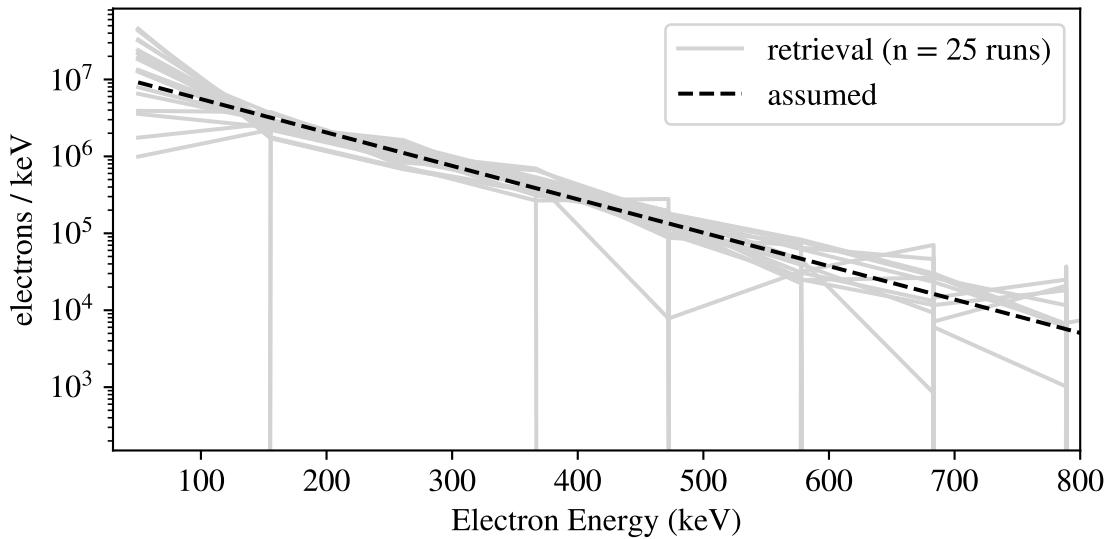


Figure 4.18: Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

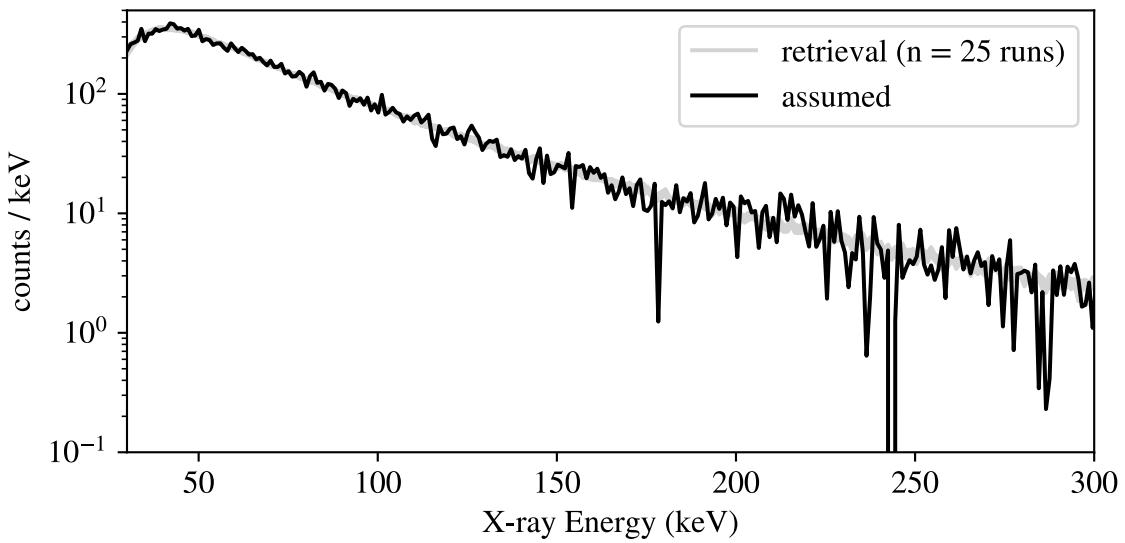
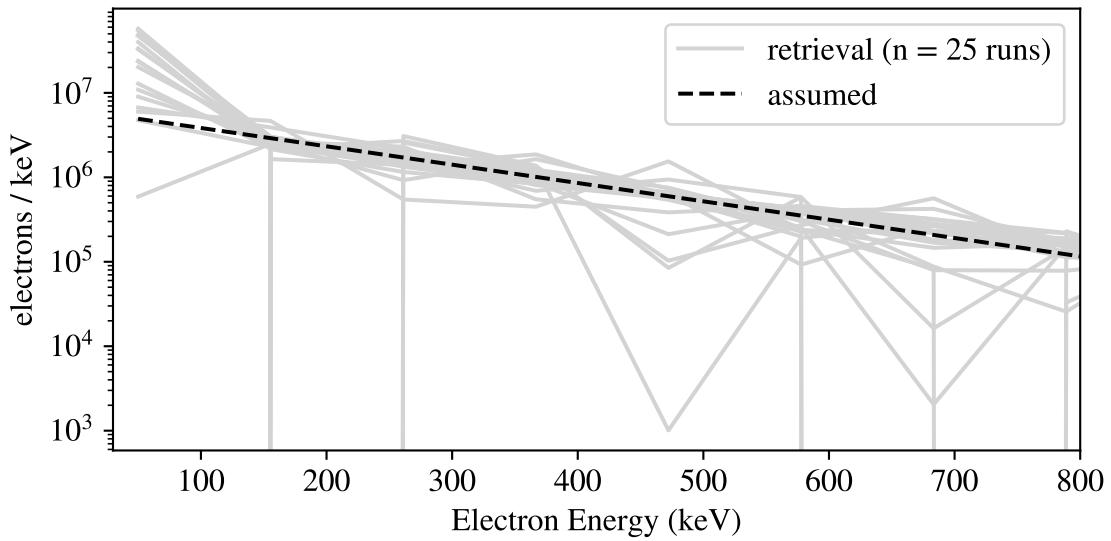


Figure 4.19: Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

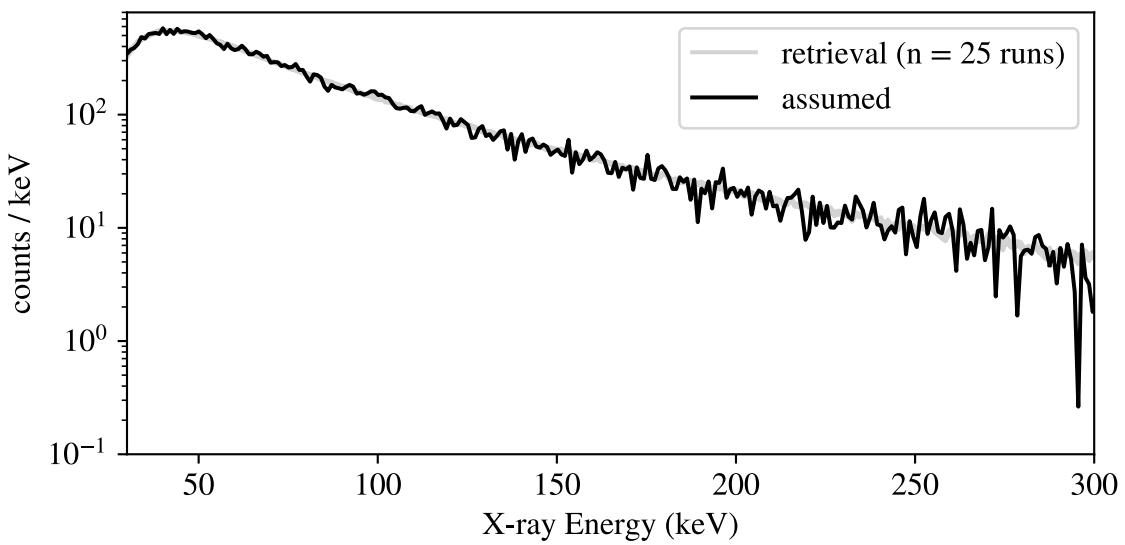
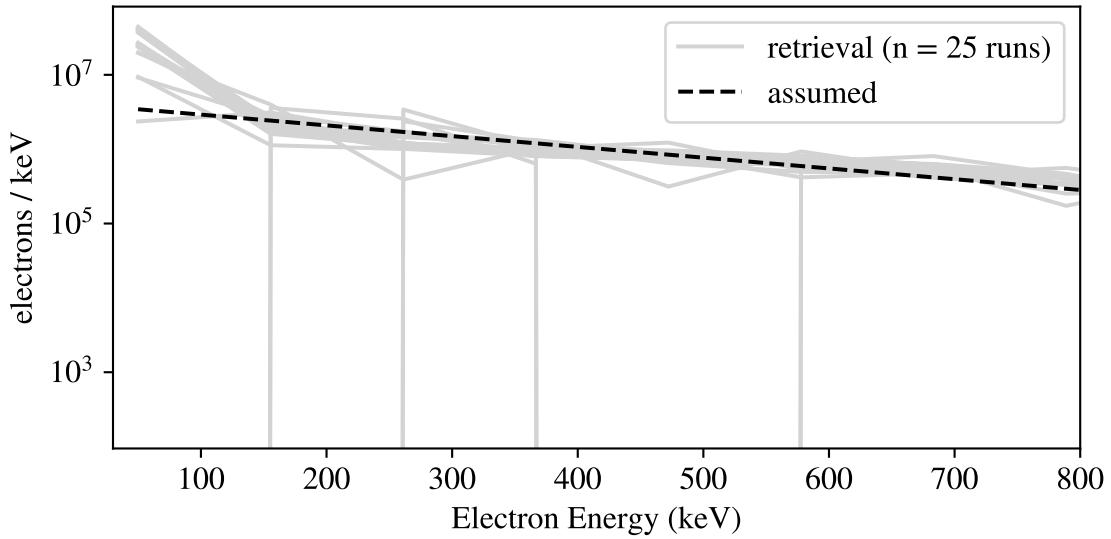


Figure 4.20: Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

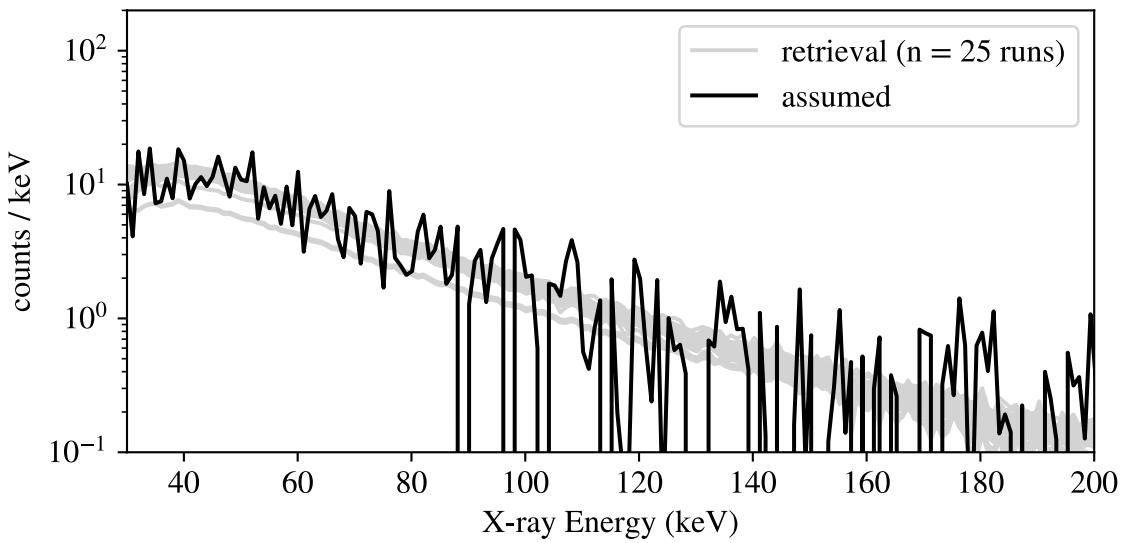
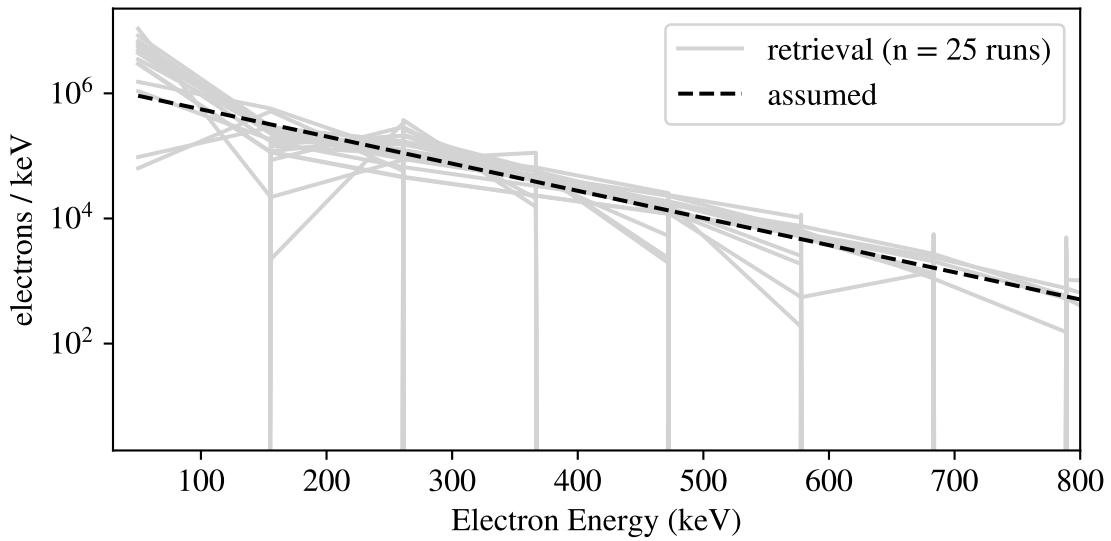


Figure 4.21: Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

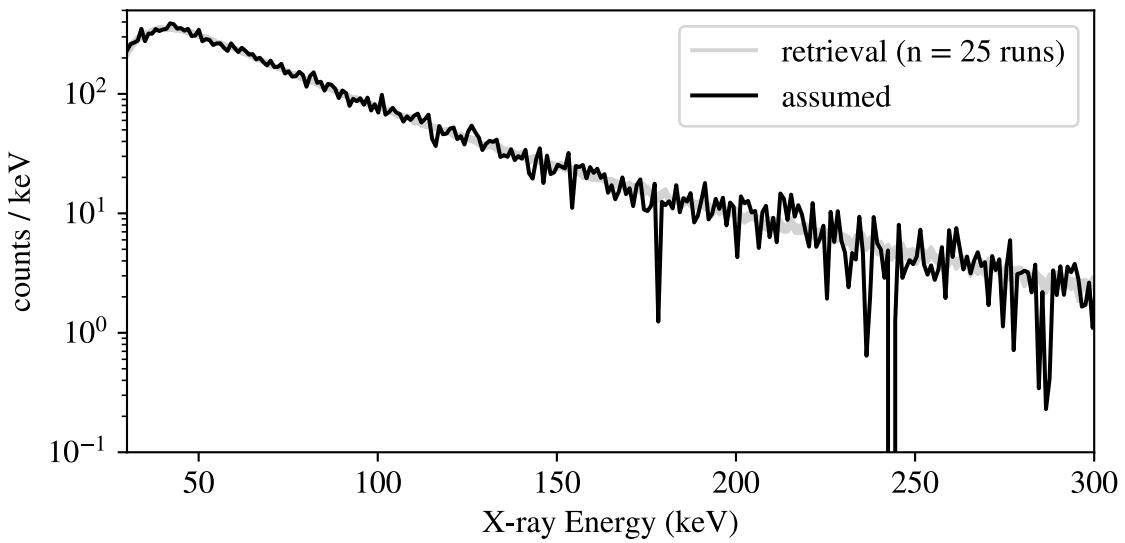
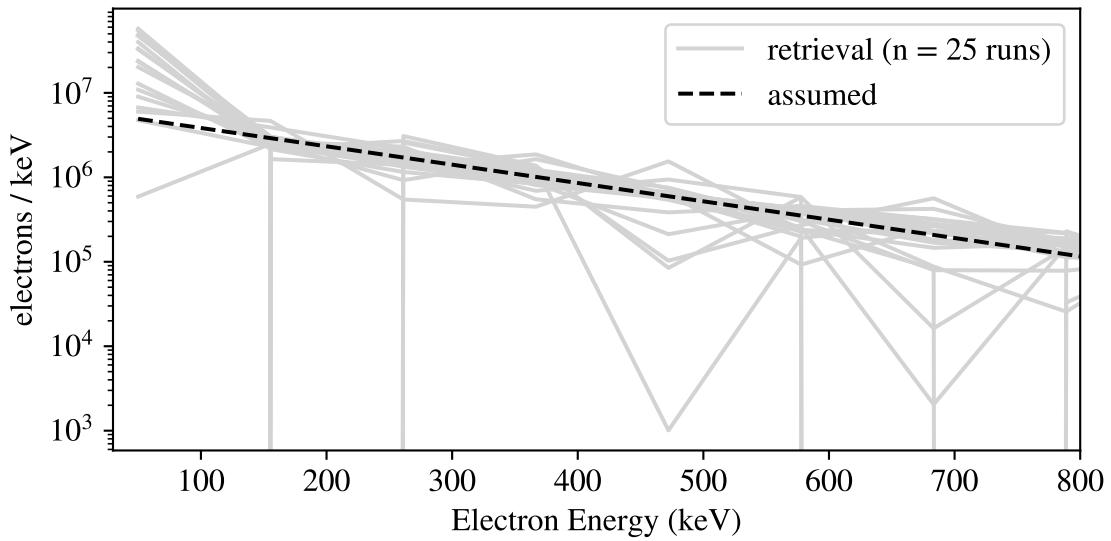


Figure 4.22: Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

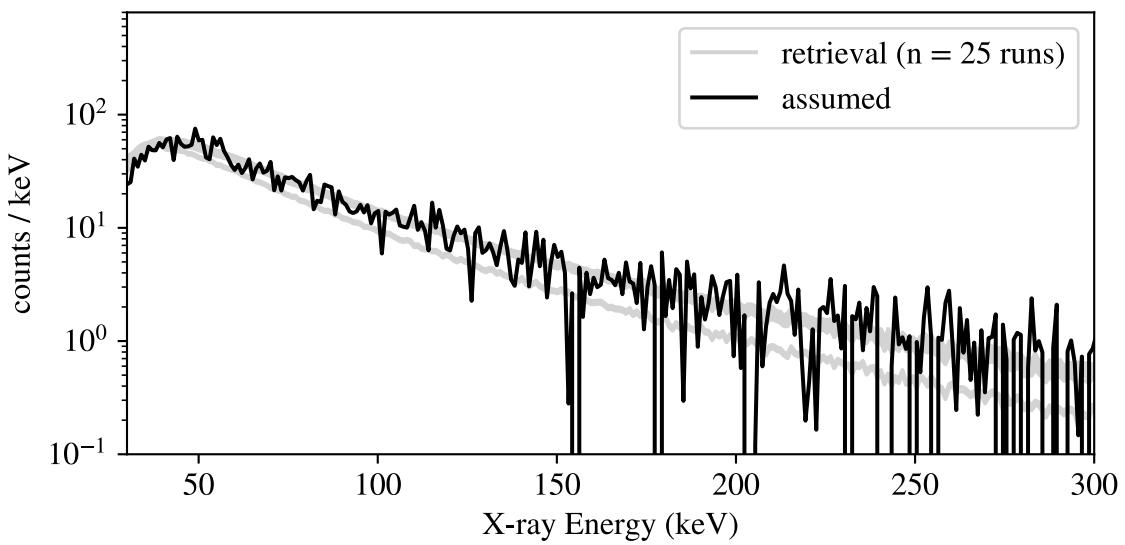
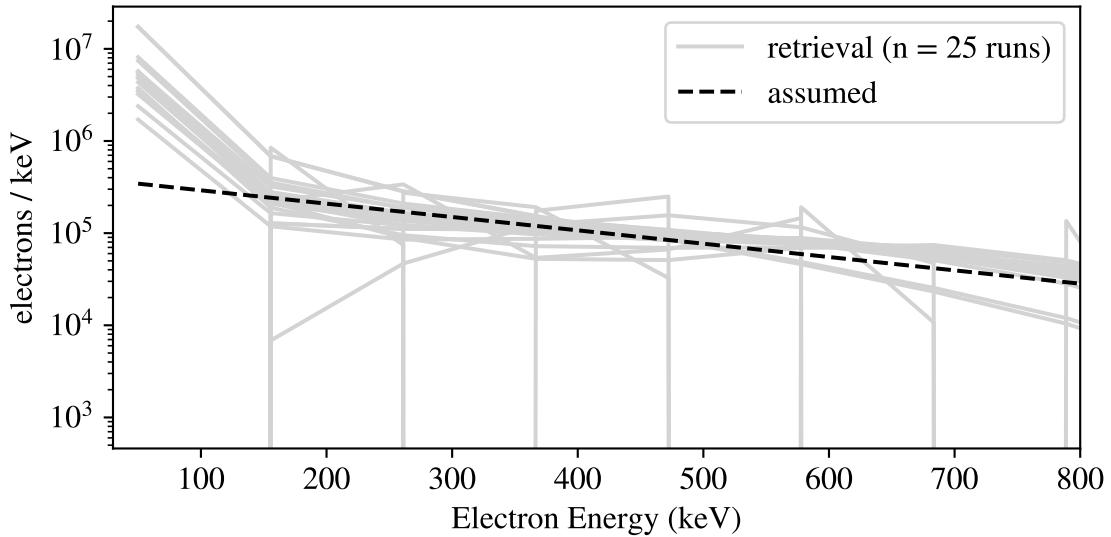


Figure 4.23: Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

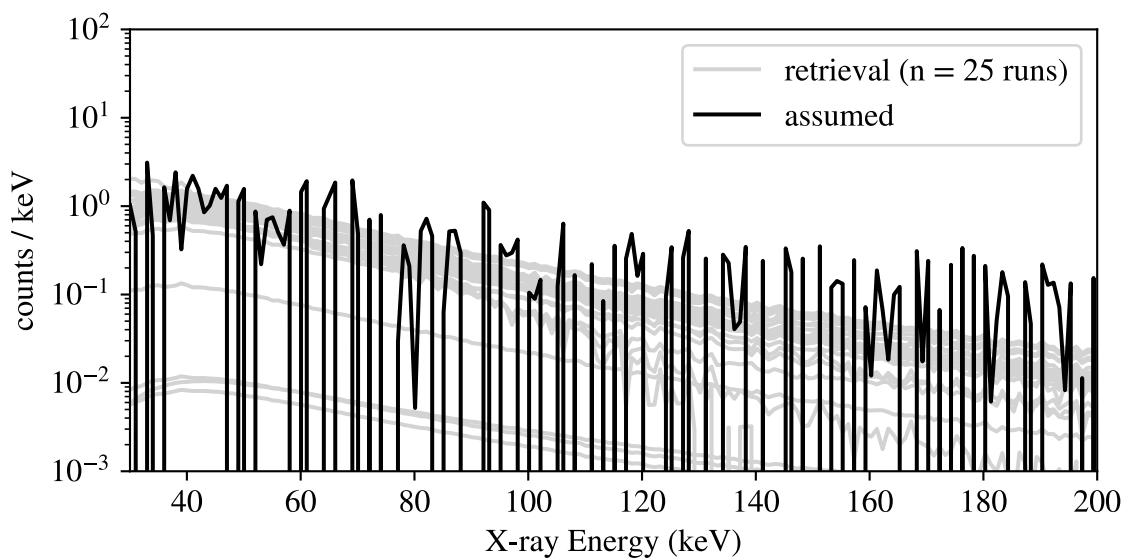
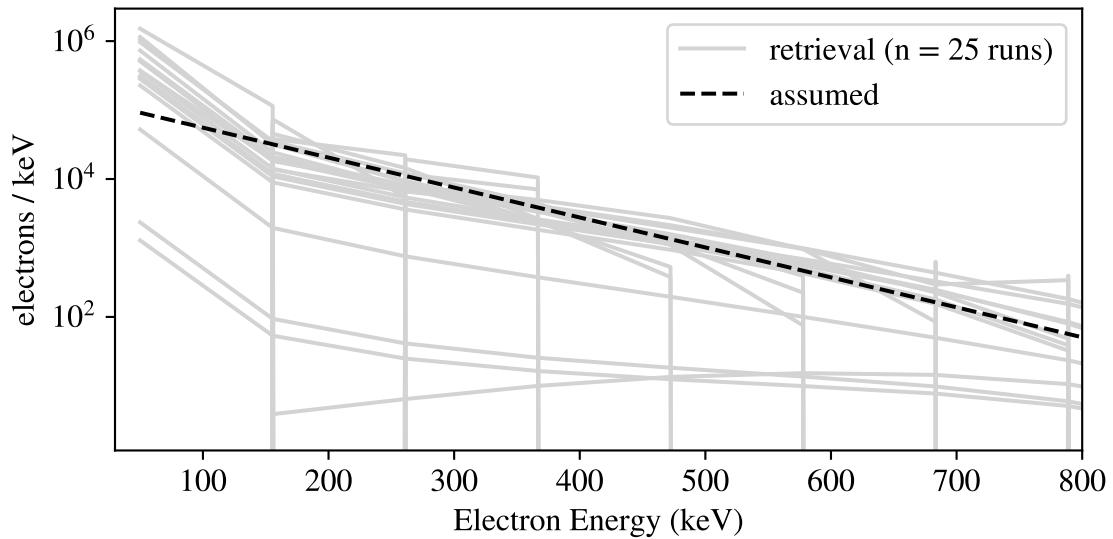


Figure 4.24: Retrieval of exponential electron spectrum (folding factor 100 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

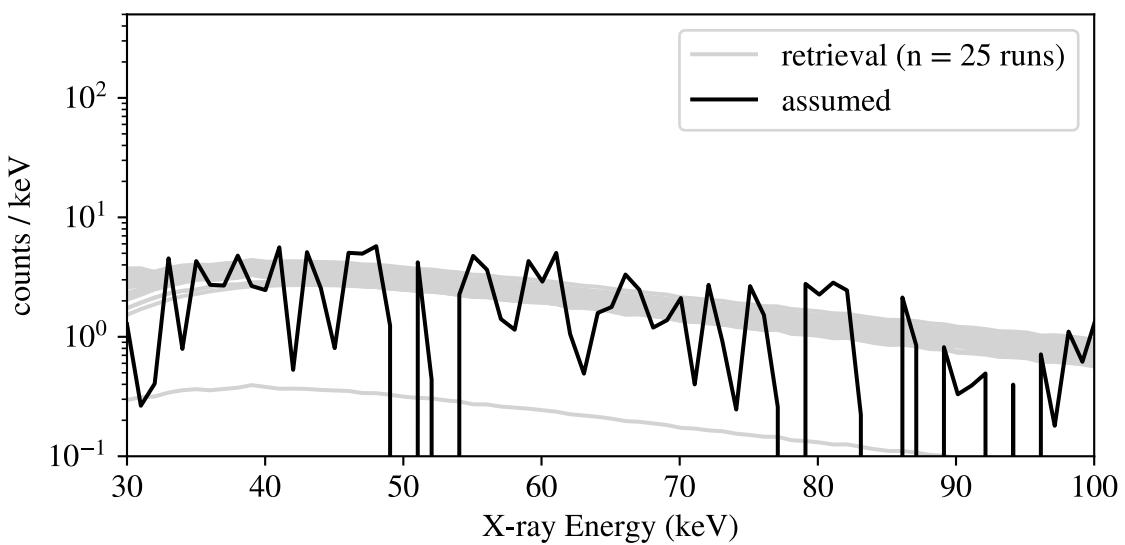
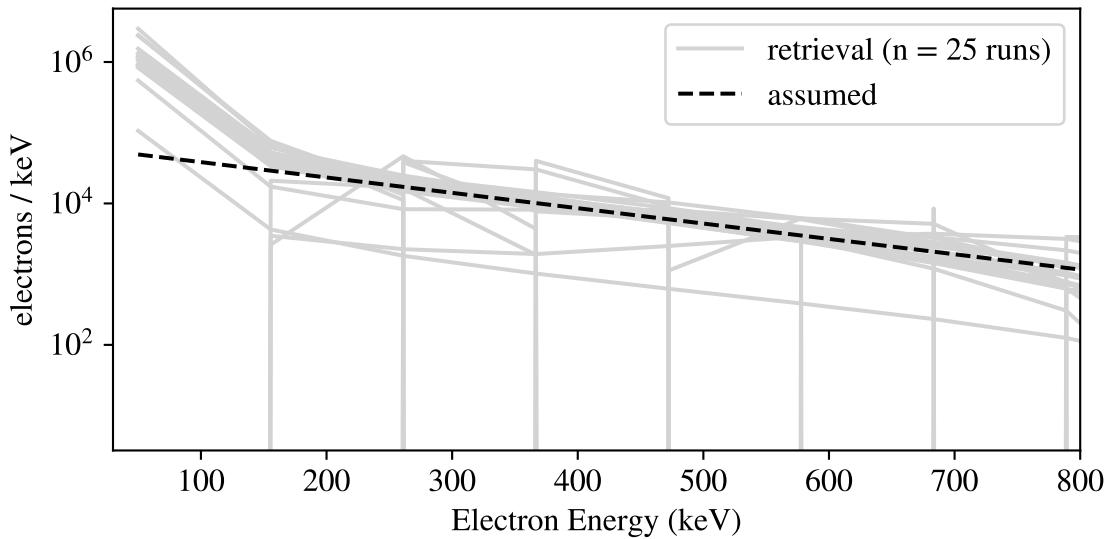


Figure 4.25: Retrieval of exponential electron spectrum (folding factor 200 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Exponential, Folding Factor 100 keV
 Positivity Constraint, Preconditioning Applied

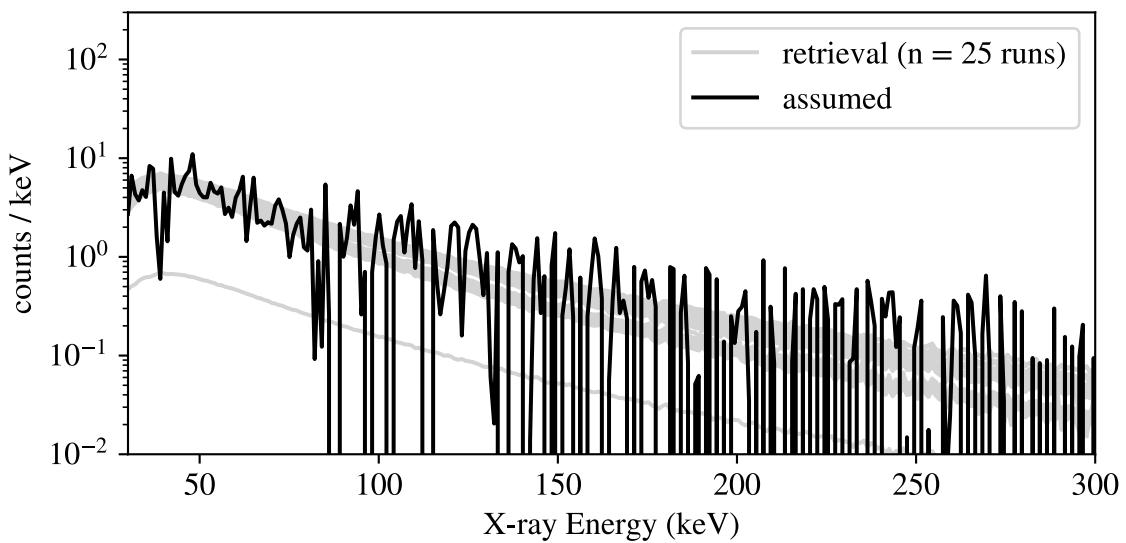
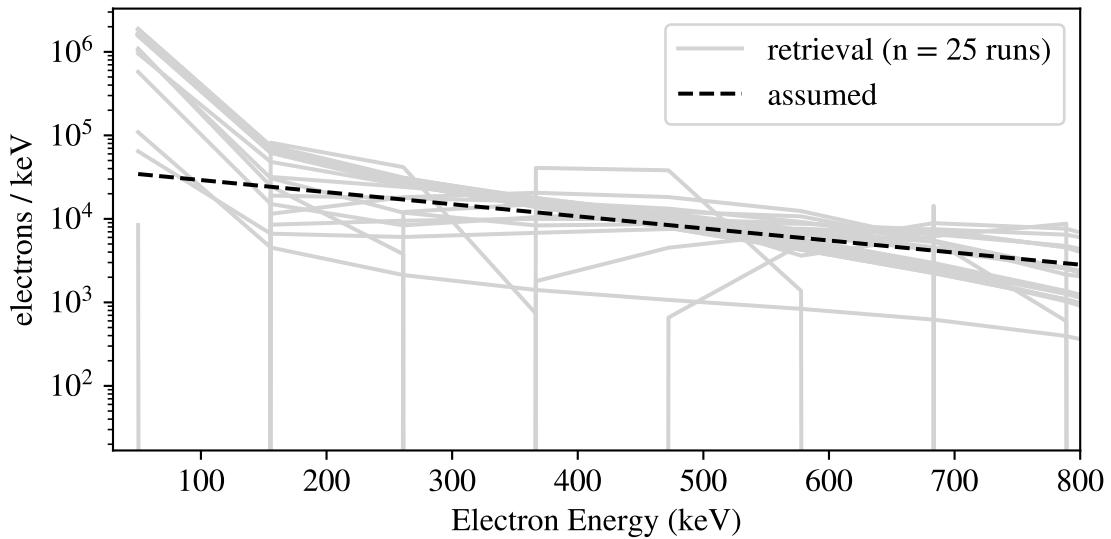


Figure 4.26: Retrieval of exponential electron spectrum (folding factor 300 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 200 keV
 Positivity Constraint, Preconditioning Applied

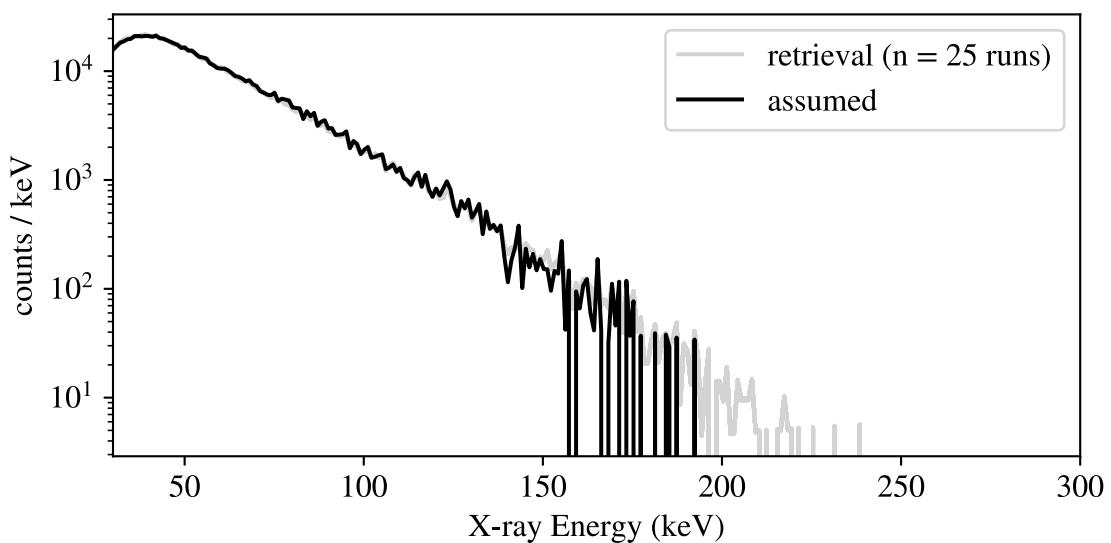
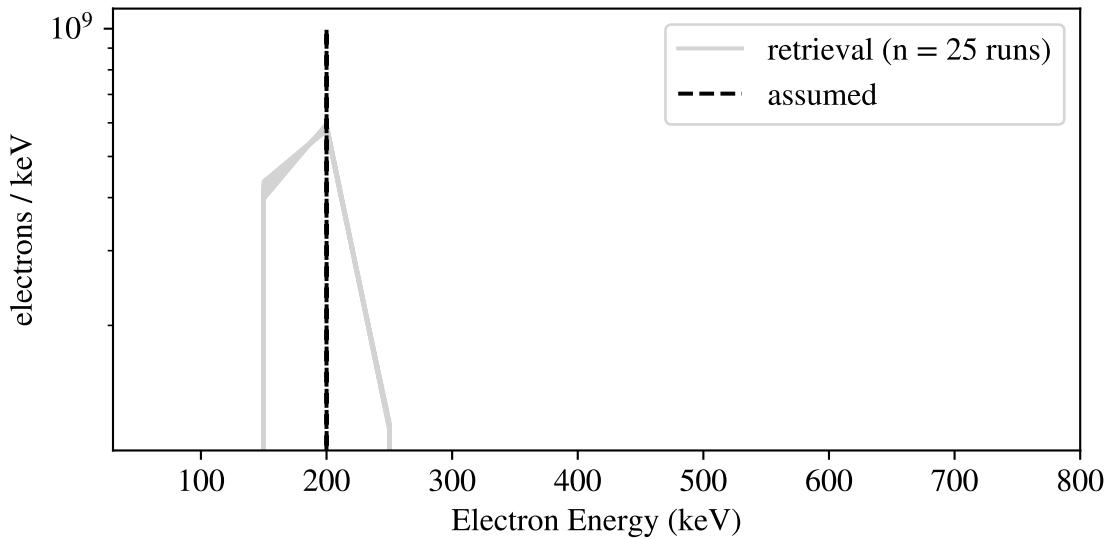


Figure 4.27: Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 400 keV
 Positivity Constraint, Preconditioning Applied

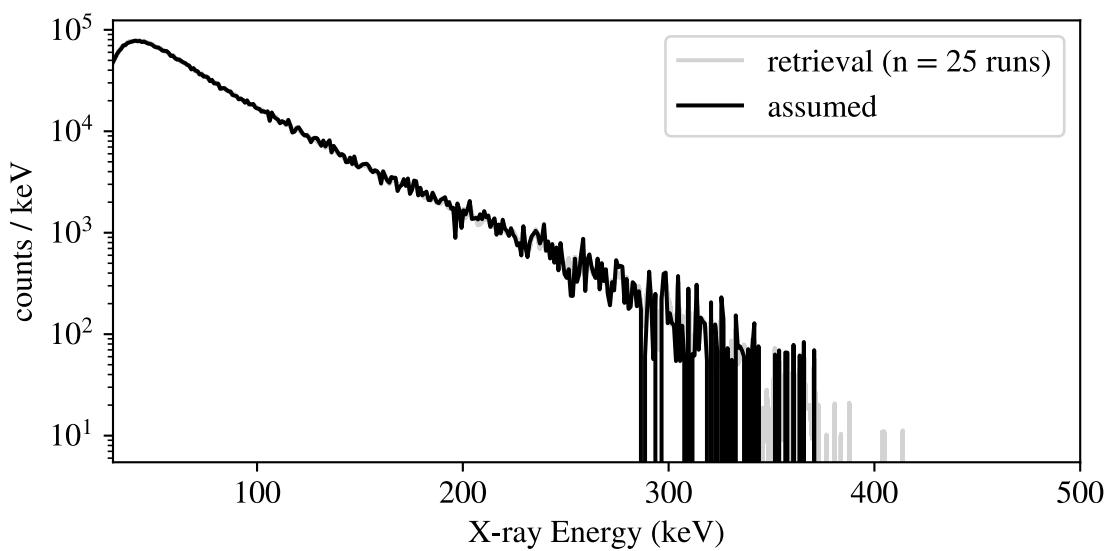
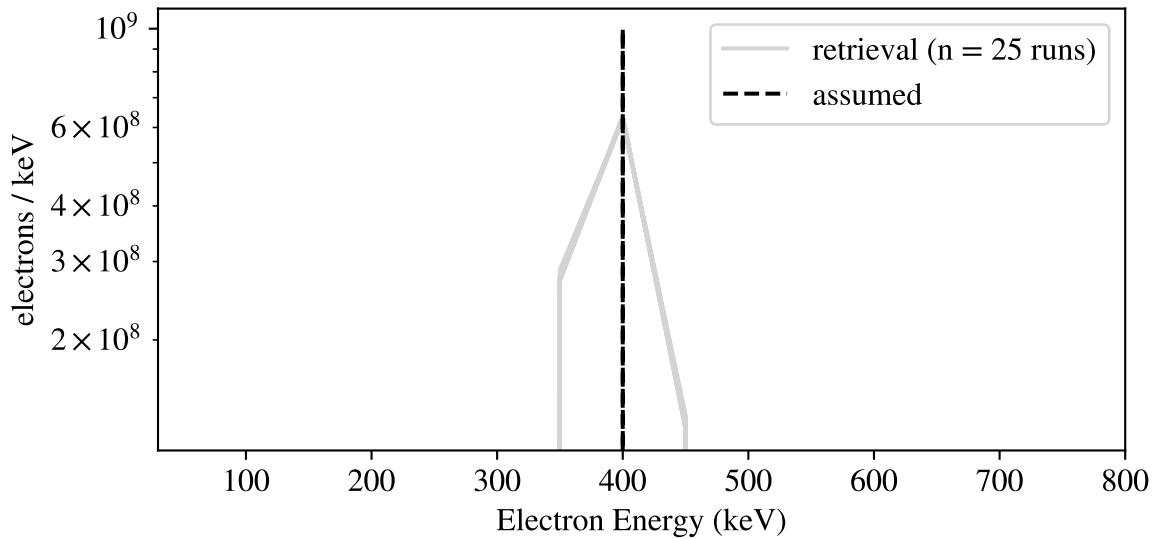


Figure 4.28: Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 600 keV
 Positivity Constraint, Preconditioning Applied

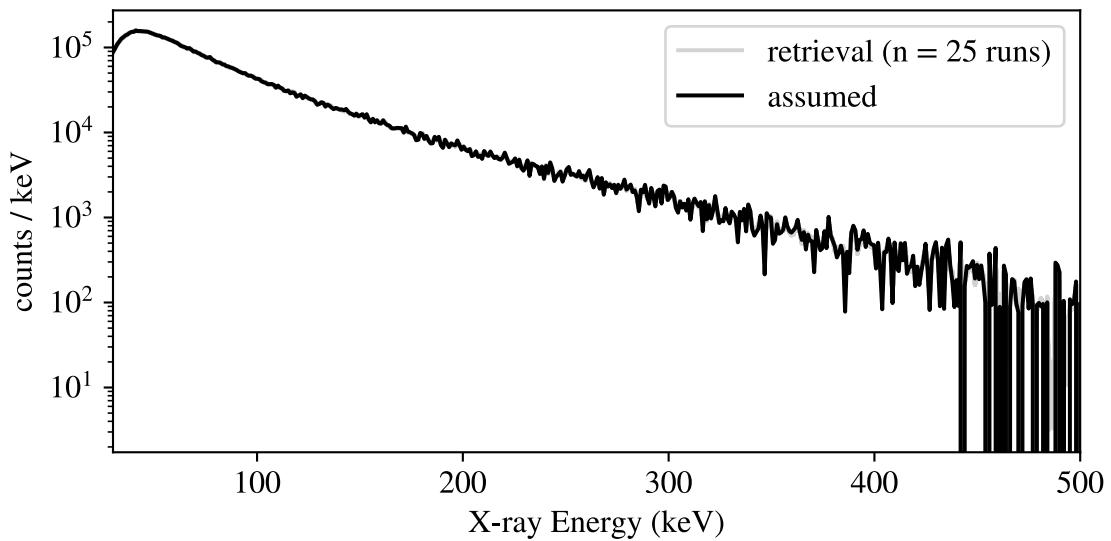
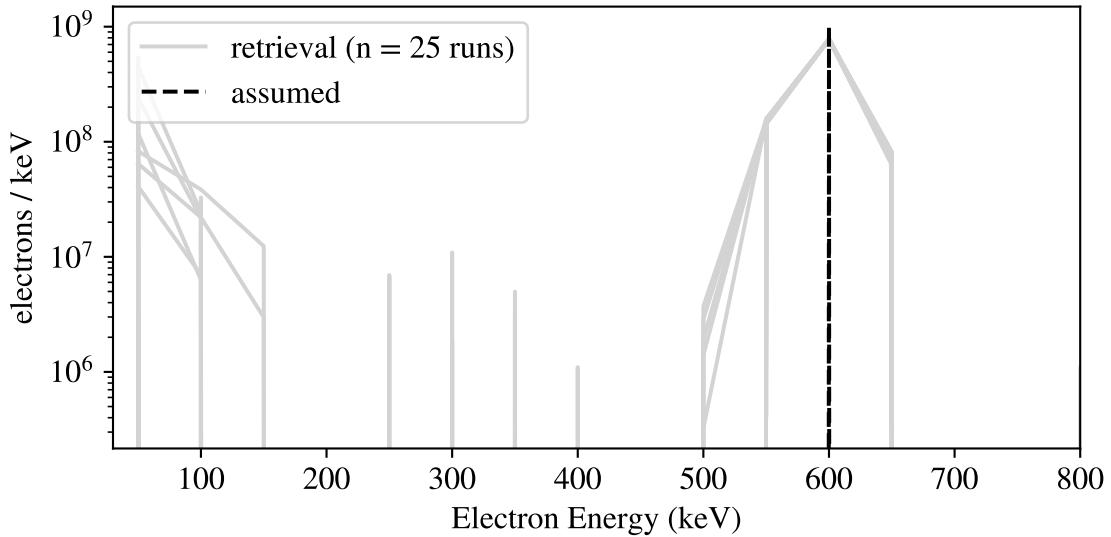


Figure 4.29: Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 200 keV
 Positivity Constraint, Preconditioning Applied

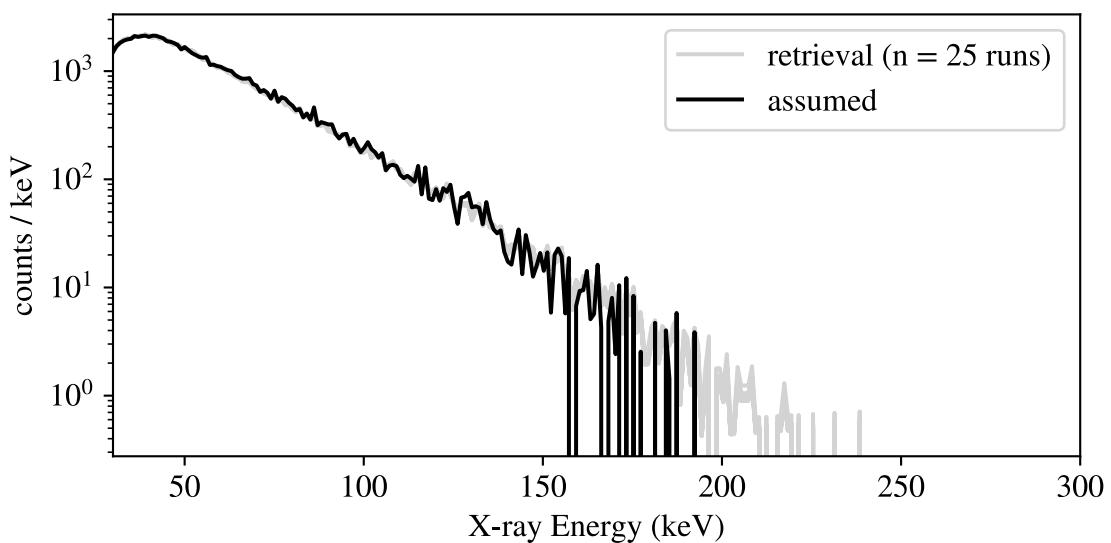
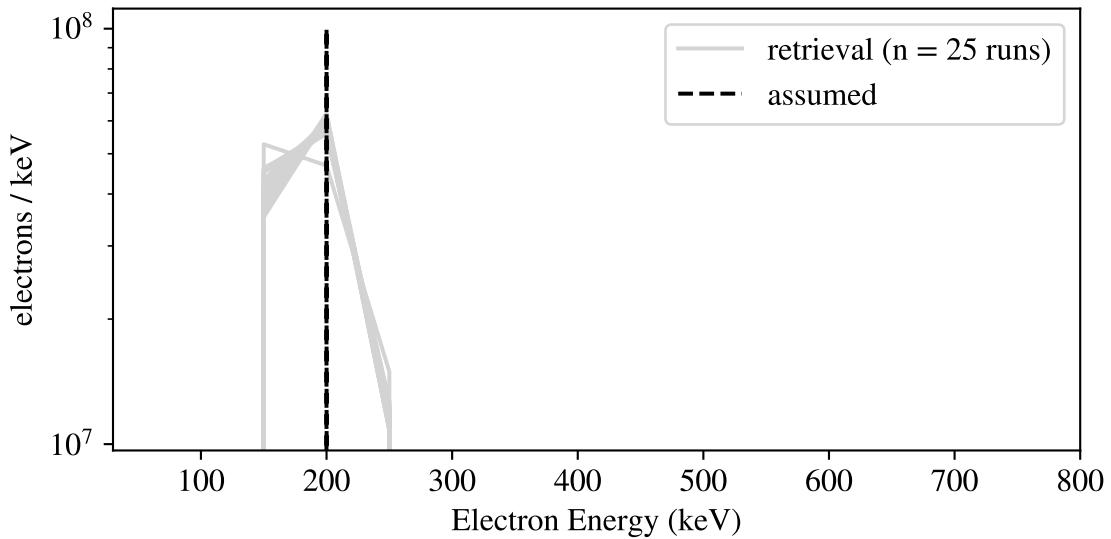


Figure 4.30: Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 400 keV
 Positivity Constraint, Preconditioning Applied

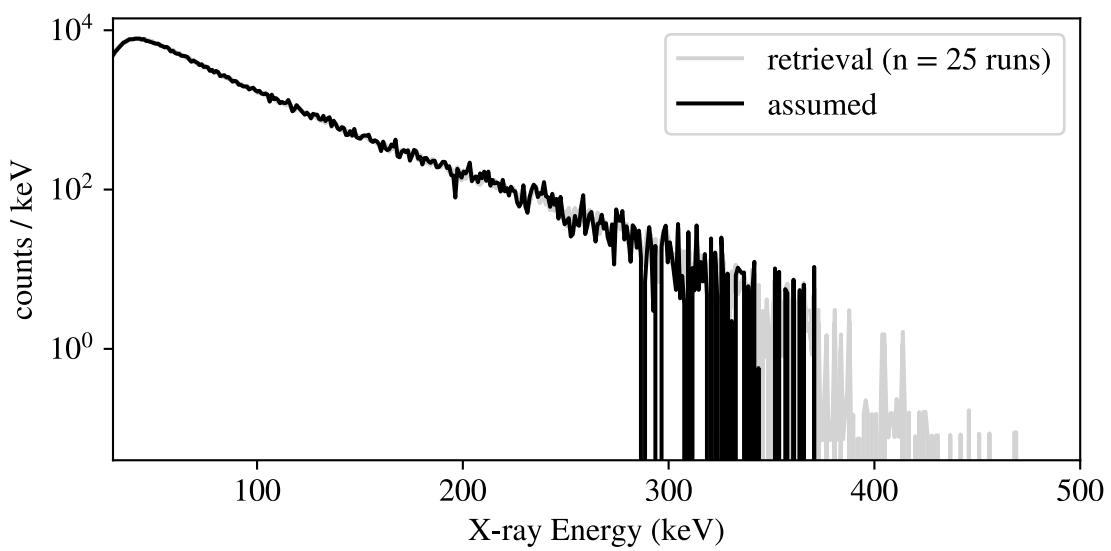
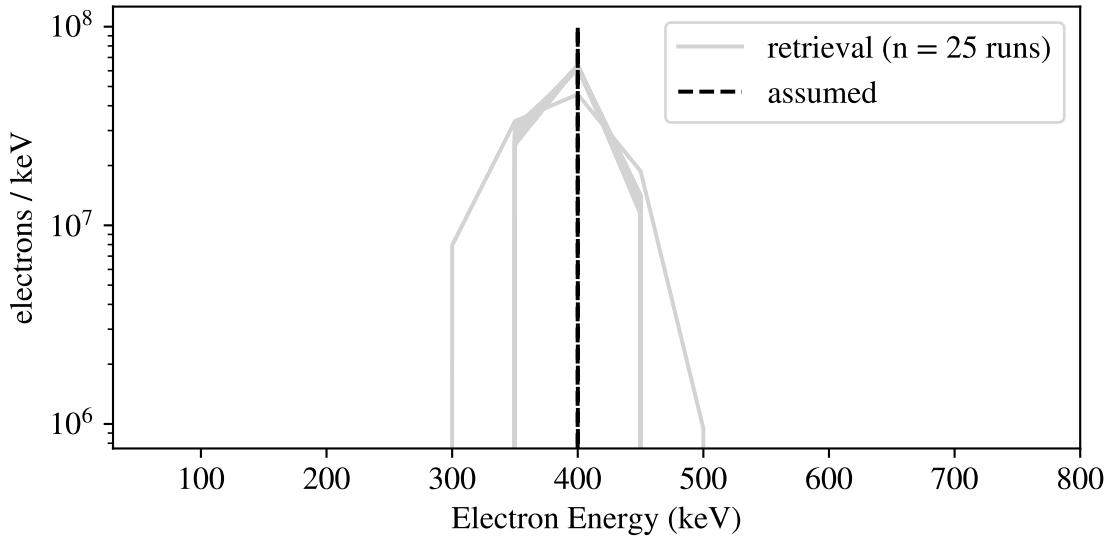


Figure 4.31: Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 600 keV
 Positivity Constraint, Preconditioning Applied

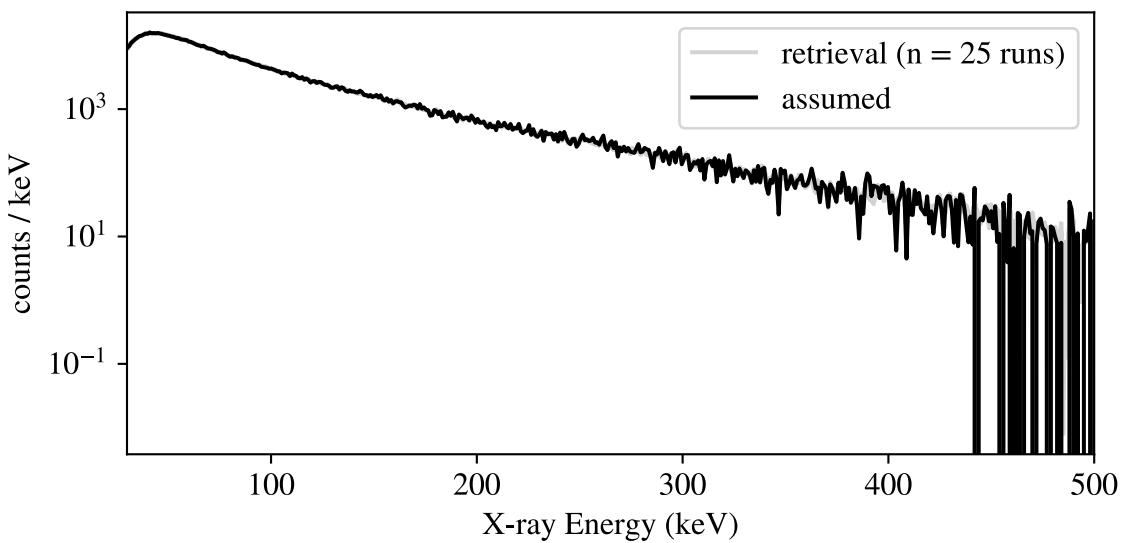
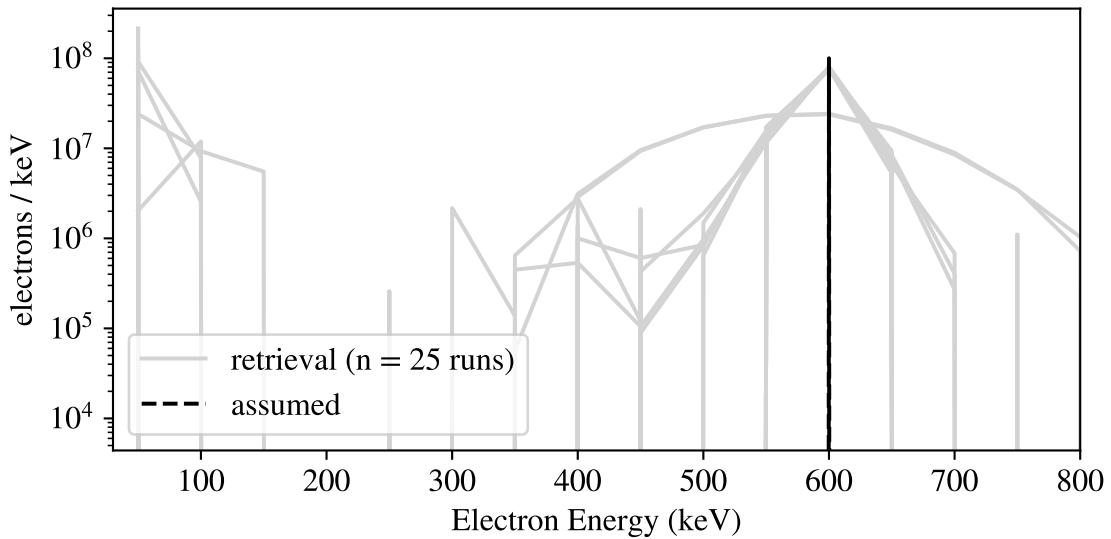


Figure 4.32: Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 200 keV
 Positivity Constraint, Preconditioning Applied

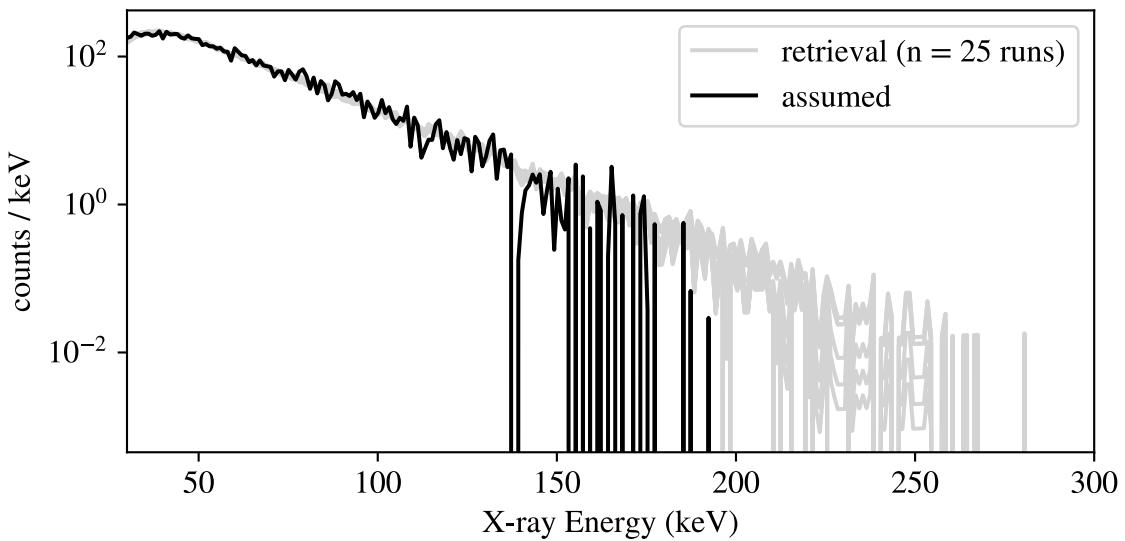
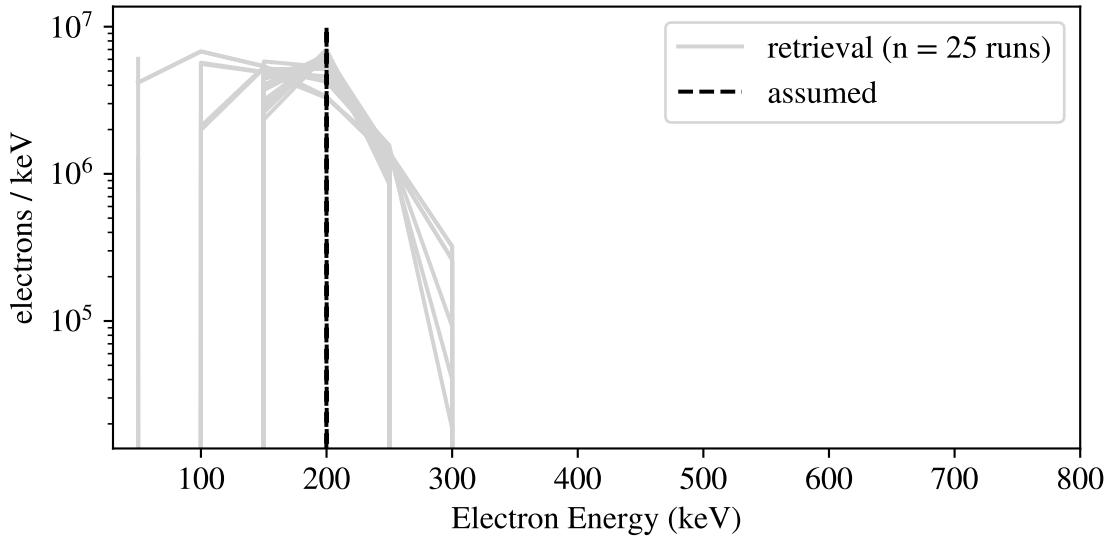


Figure 4.33: Retrieval of monoenergetic electron spectrum (energy 200 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 400 keV
 Positivity Constraint, Preconditioning Applied

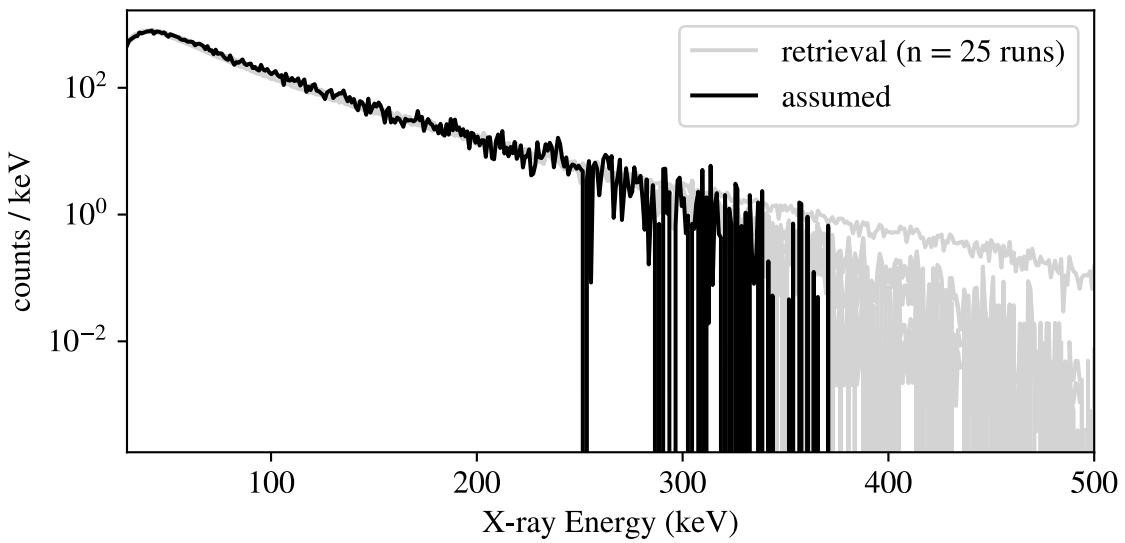
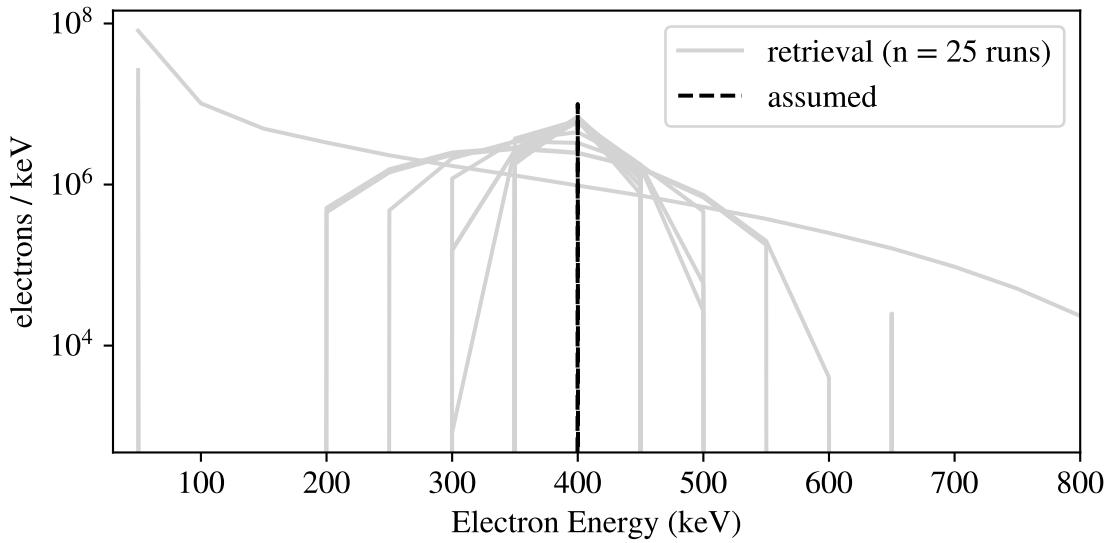


Figure 4.34: Retrieval of monoenergetic electron spectrum (energy 400 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Monoenergetic, Energy 600 keV
 Positivity Constraint, Preconditioning Applied

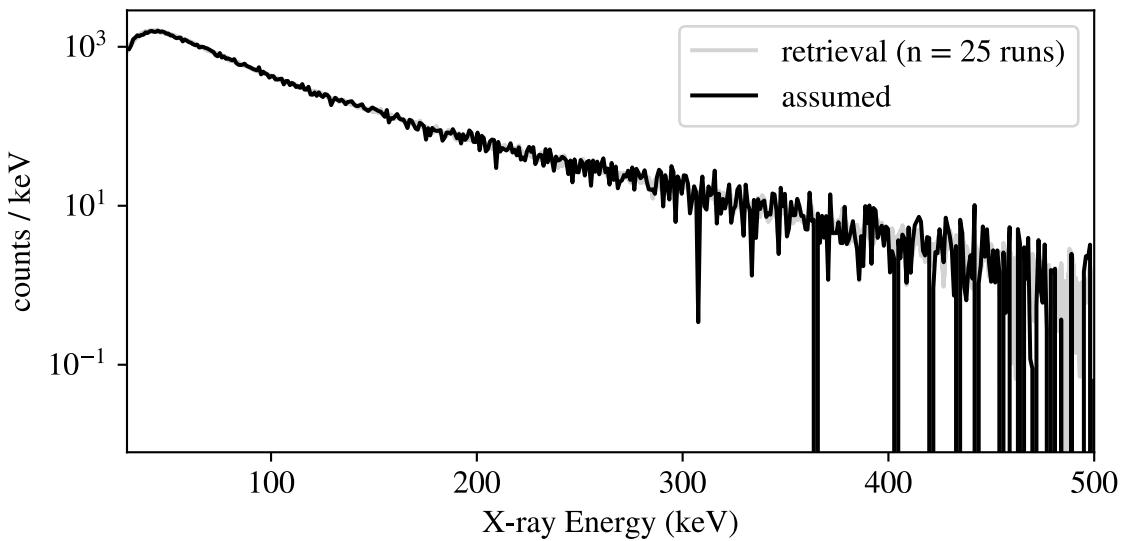
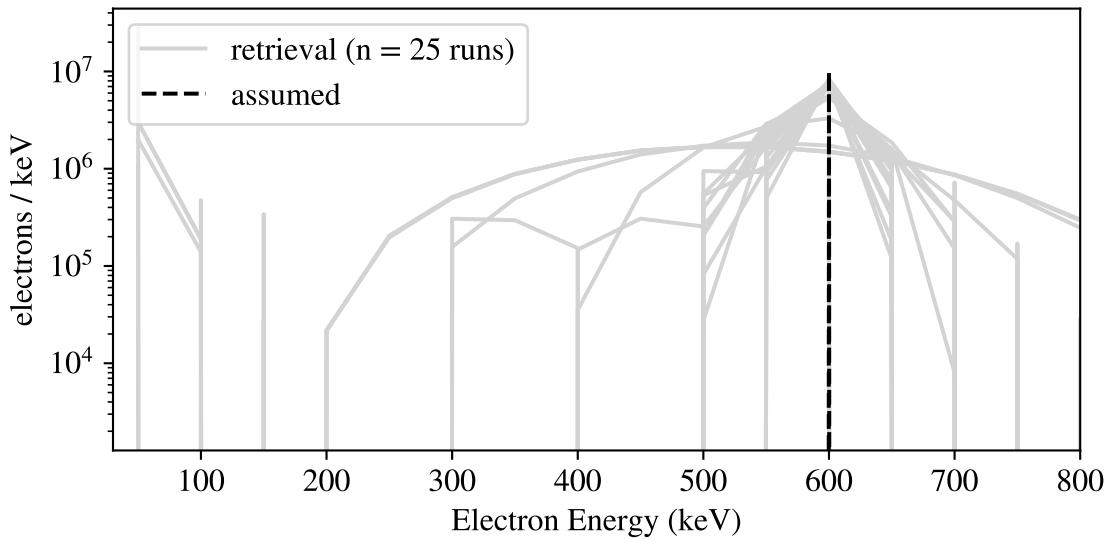


Figure 4.35: Retrieval of monoenergetic electron spectrum (energy 600 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 50 keV
Positivity Constraint, Preconditioning Applied

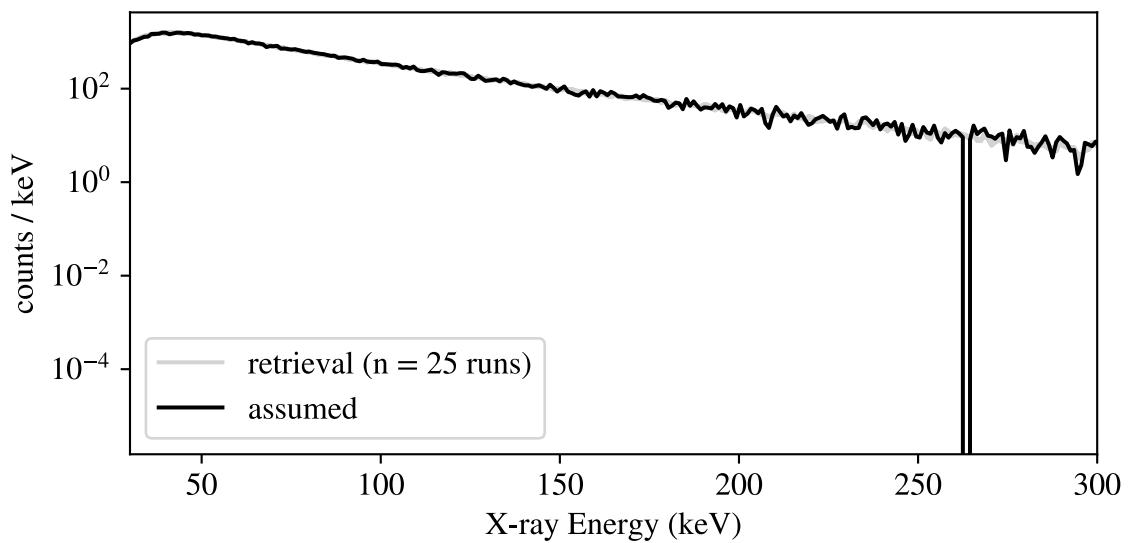
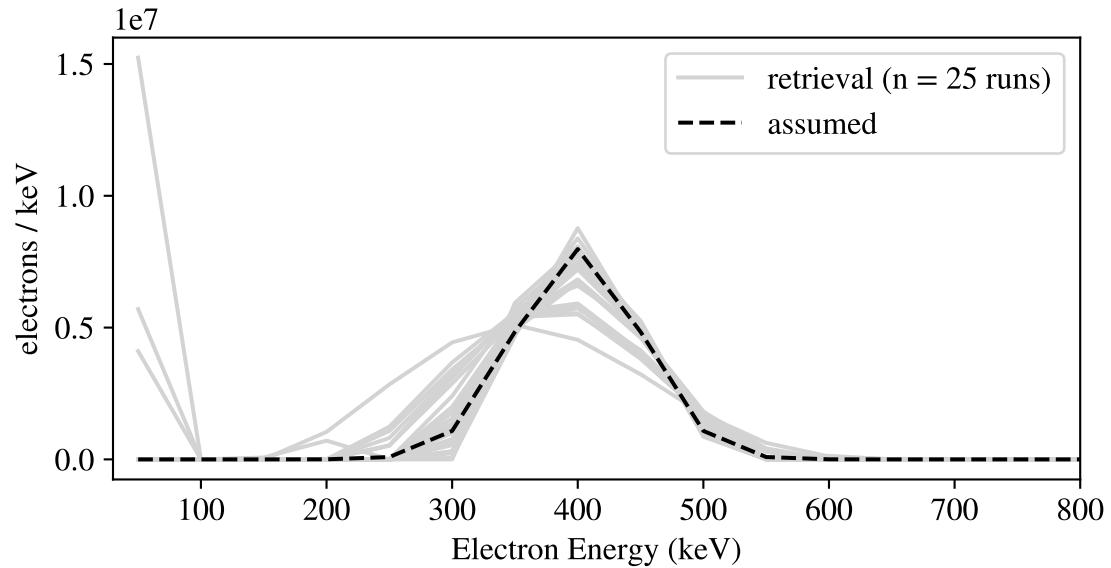


Figure 4.36: Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Gaussian, Sigma 100 keV
 Positivity Constraint, Preconditioning Applied

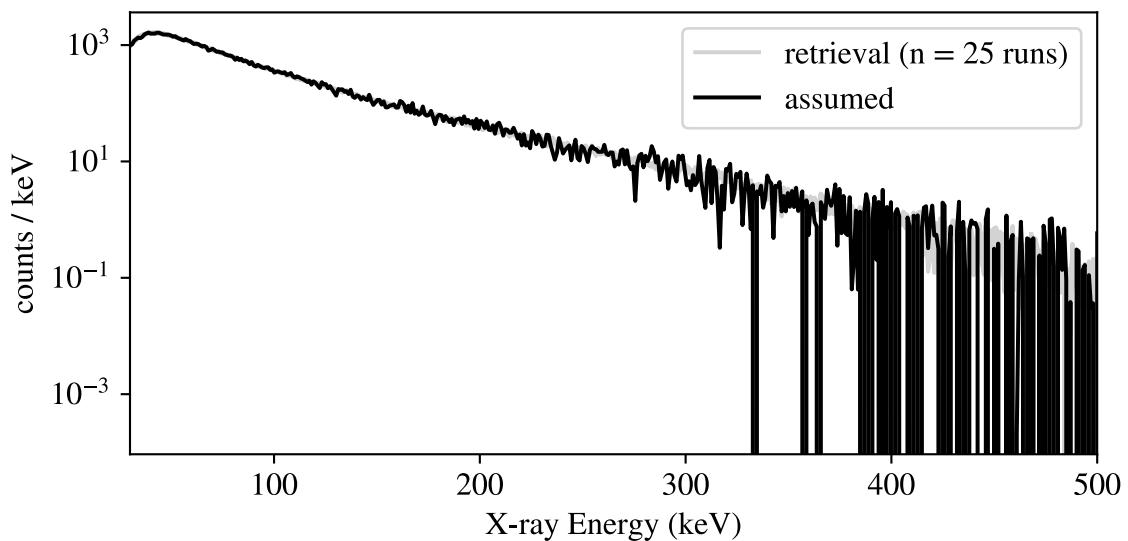
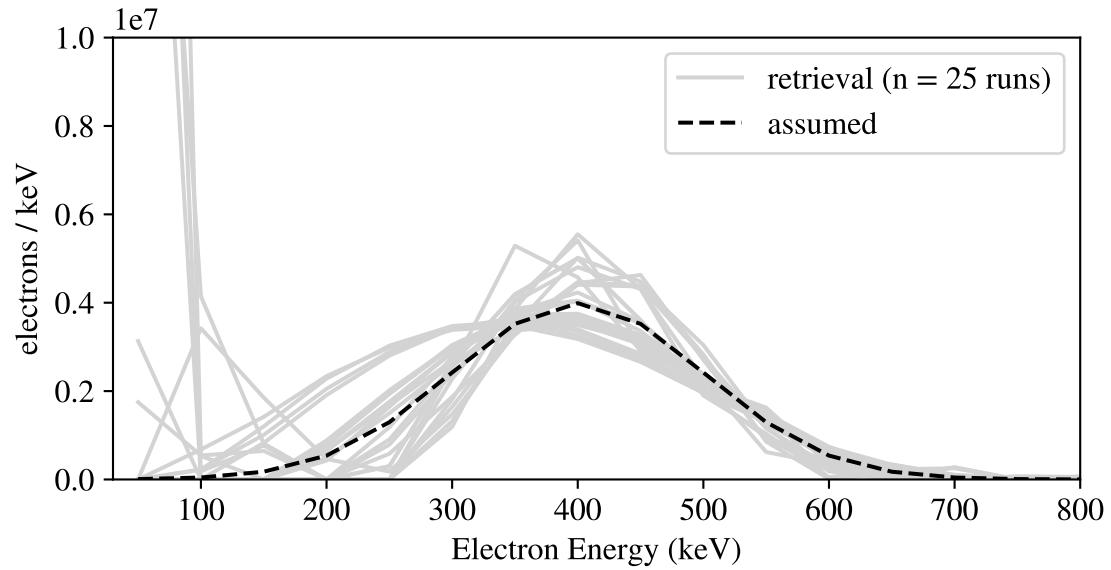


Figure 4.37: Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 150 keV
Positivity Constraint, Preconditioning Applied

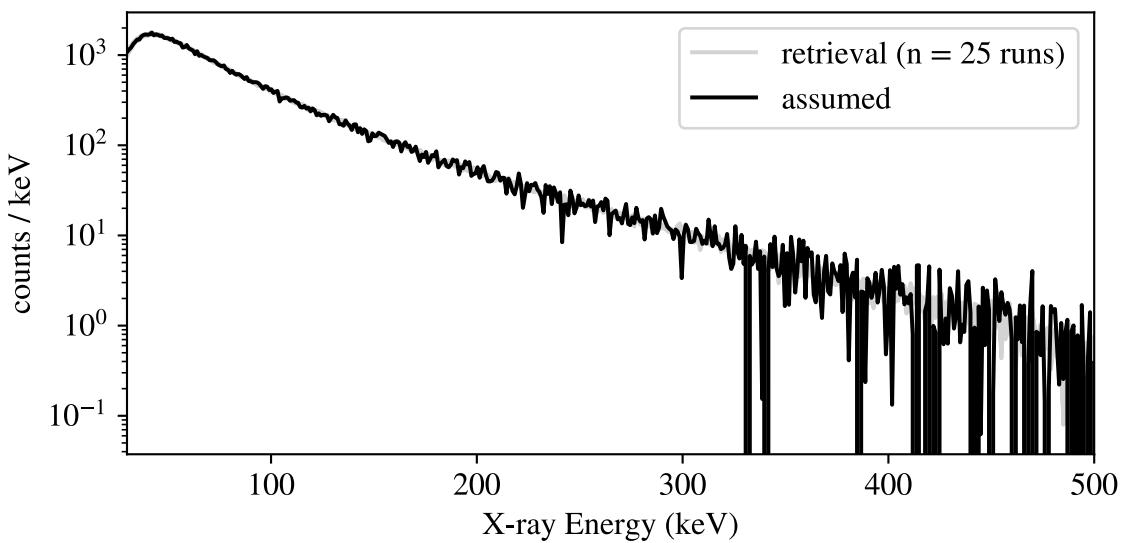
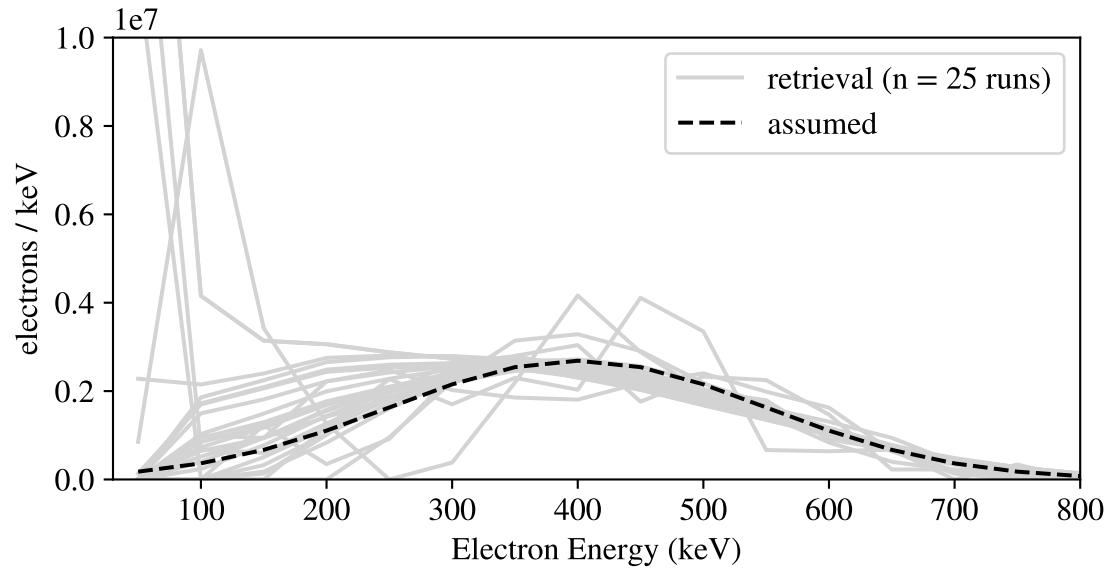


Figure 4.38: Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^9) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 50 keV
Positivity Constraint, Preconditioning Applied

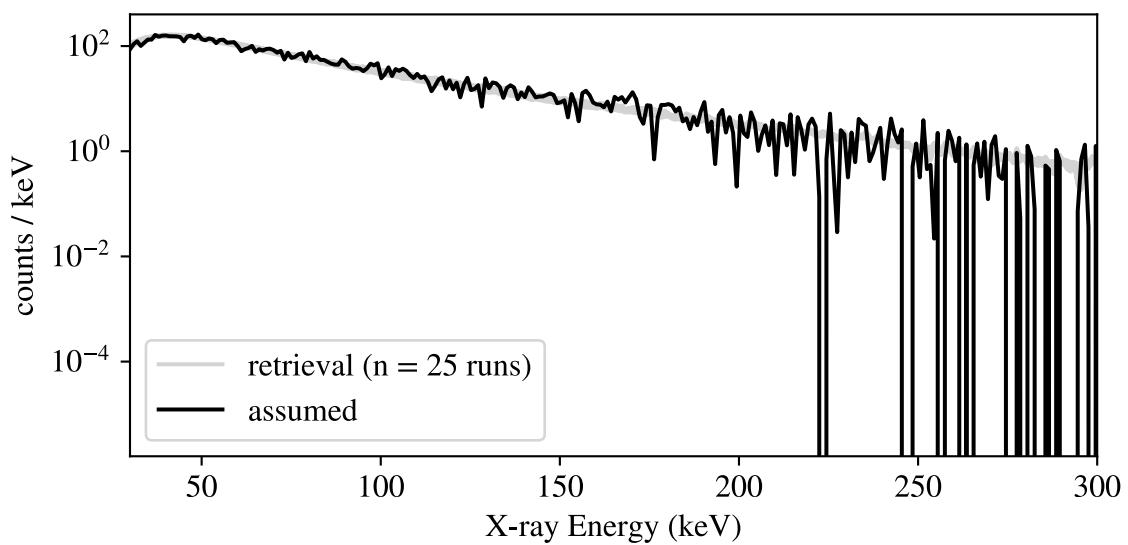
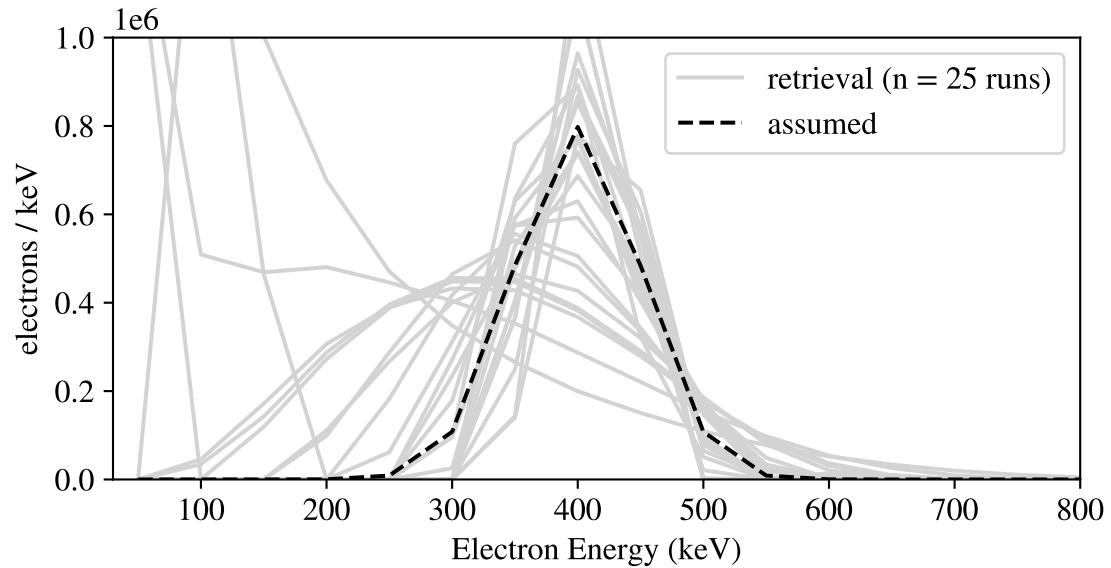


Figure 4.39: Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 100 keV
Positivity Constraint, Preconditioning Applied

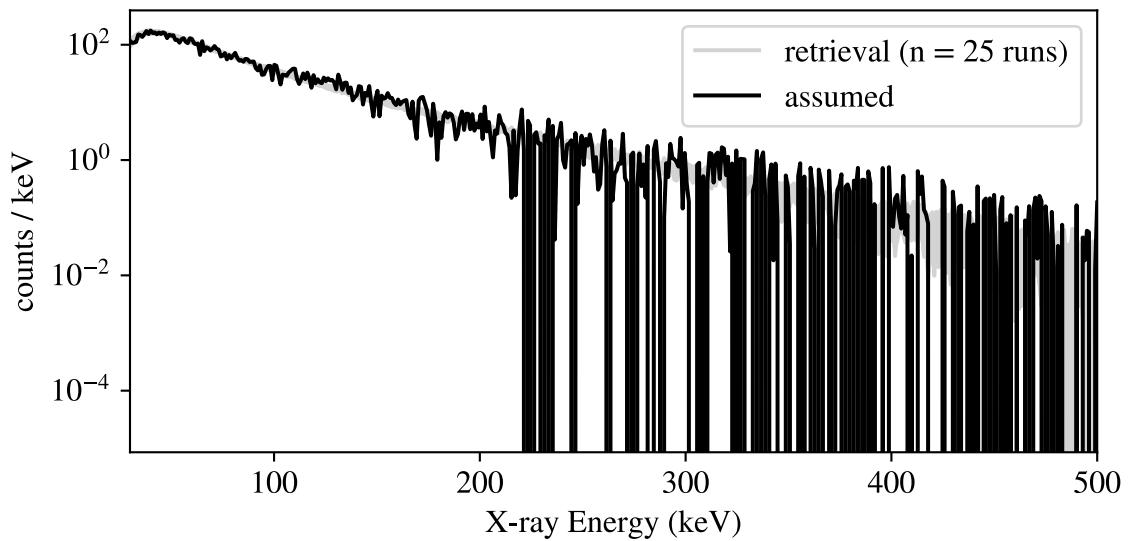
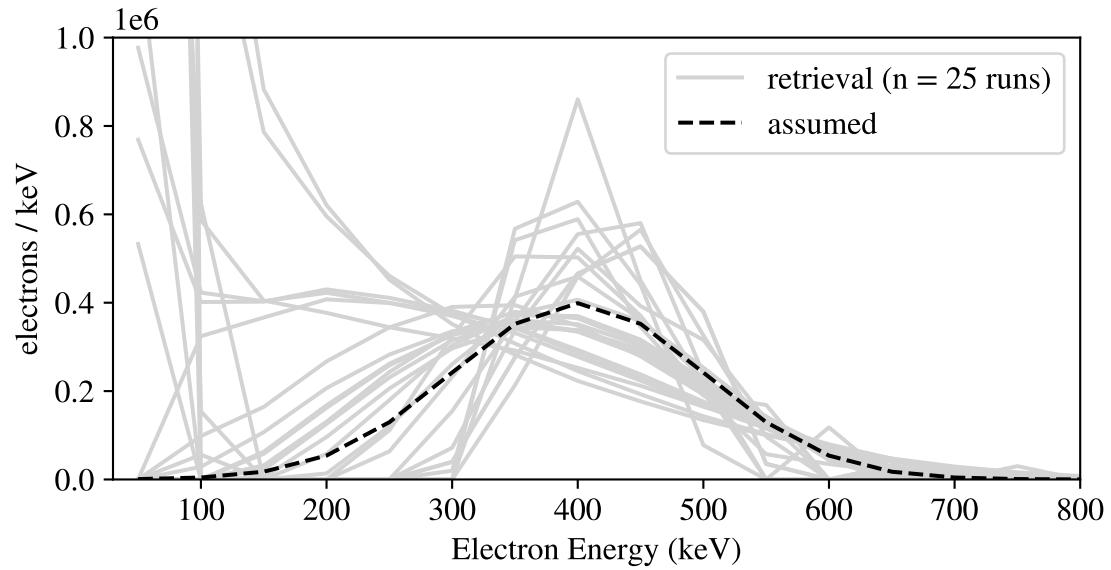


Figure 4.40: Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
 Gaussian, Sigma 150 keV
 Positivity Constraint, Preconditioning Applied

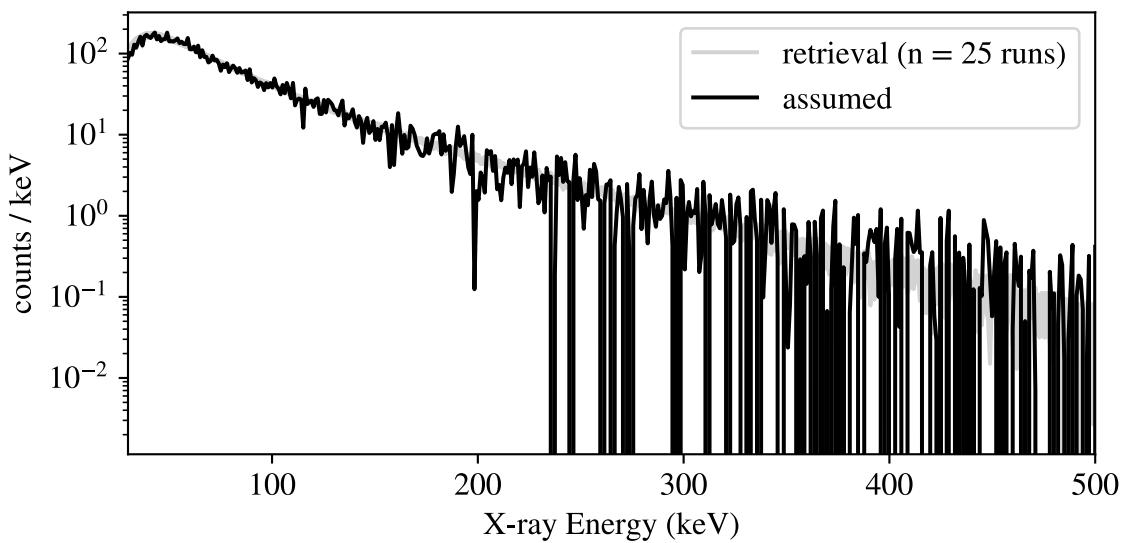
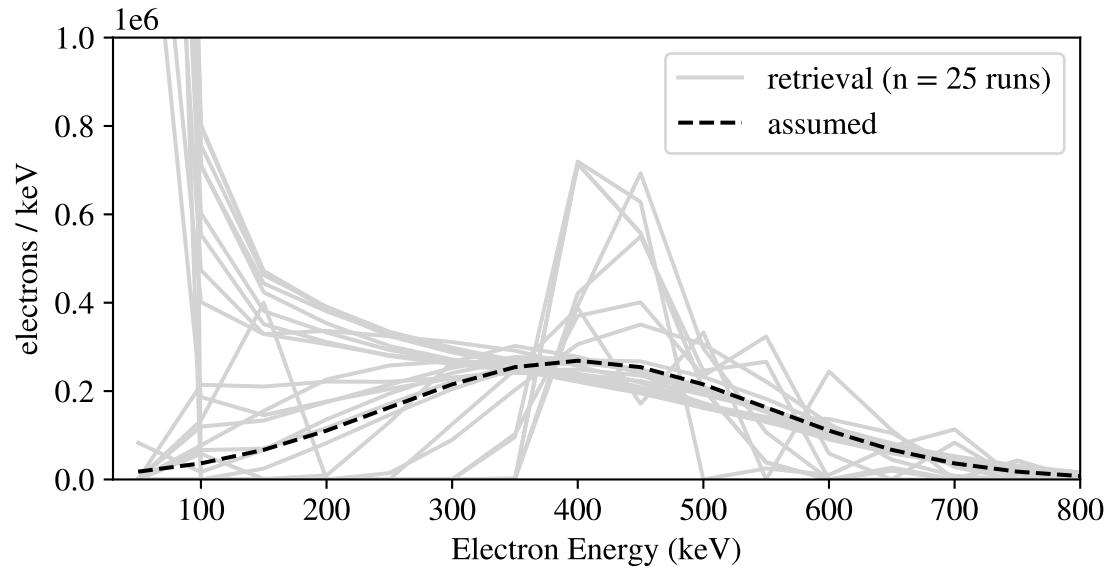


Figure 4.41: Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^8) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 50 keV
Positivity Constraint, Preconditioning Applied

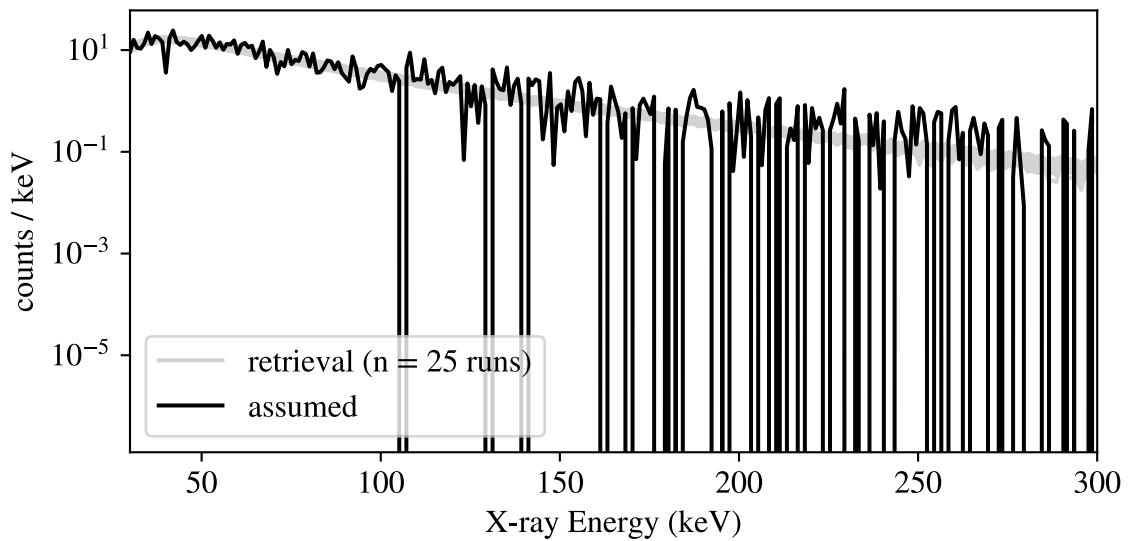
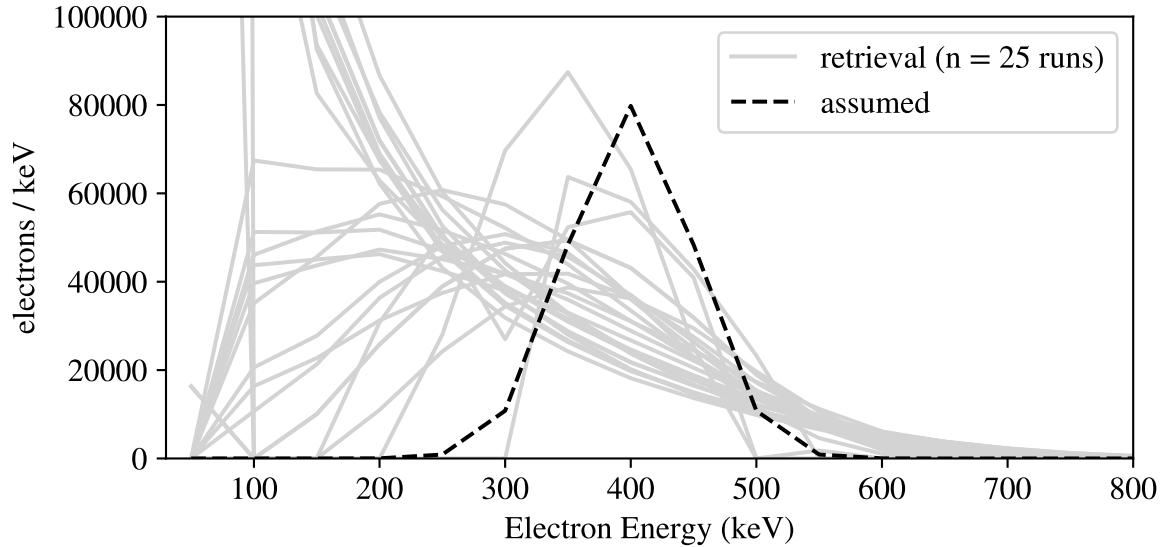


Figure 4.42: Retrieval of gaussian electron spectrum (sigma 50 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 100 keV
Positivity Constraint, Preconditioning Applied

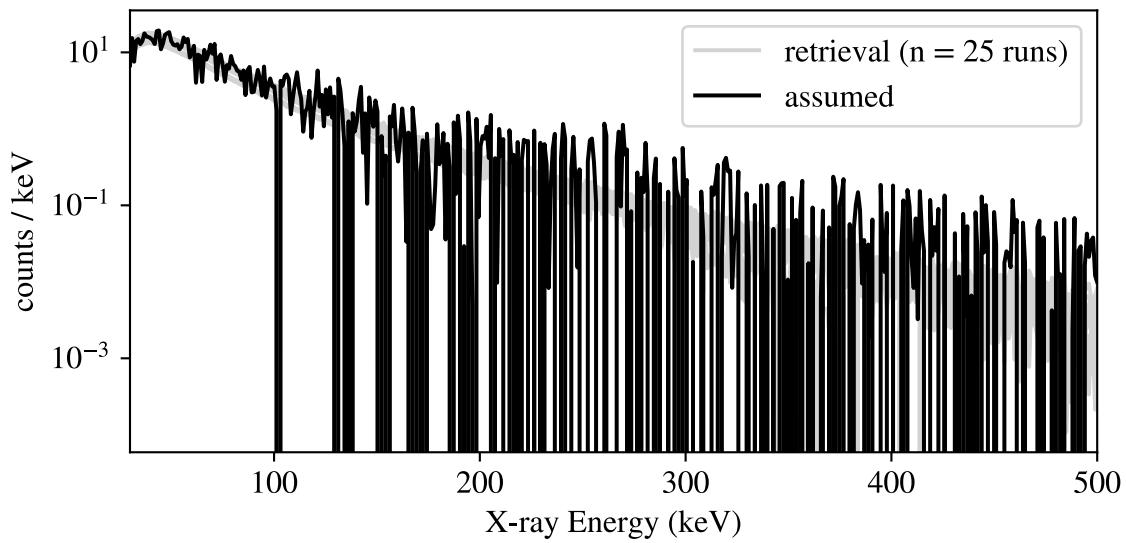
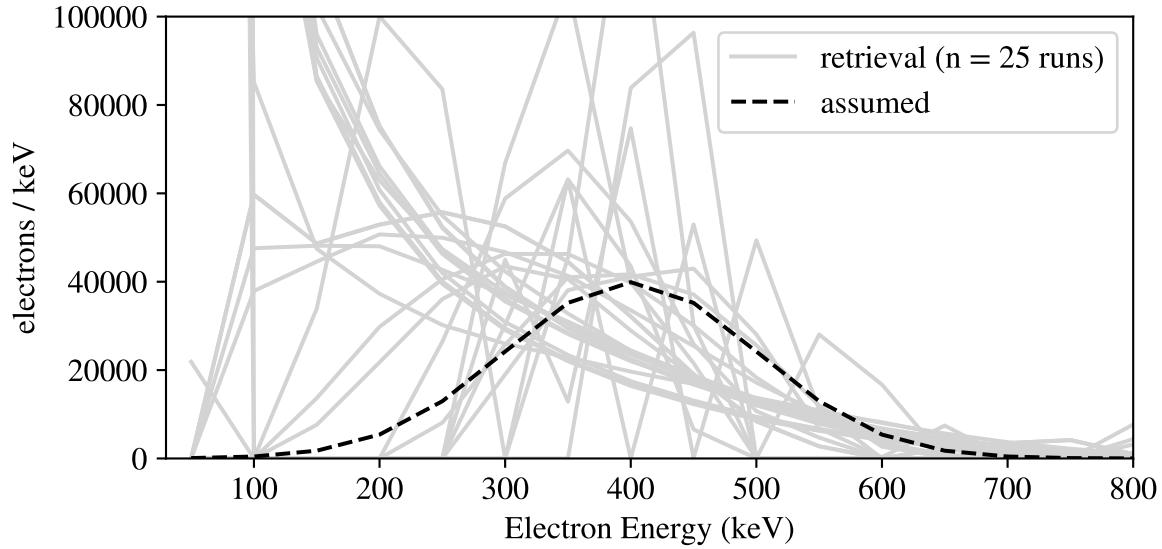


Figure 4.43: Retrieval of gaussian electron spectrum (sigma 100 keV, total electron count 1×10^7) from synthetic X-ray data.

Simulated and Retrieved Electron Spectra
Gaussian, Sigma 150 keV
Positivity Constraint, Preconditioning Applied

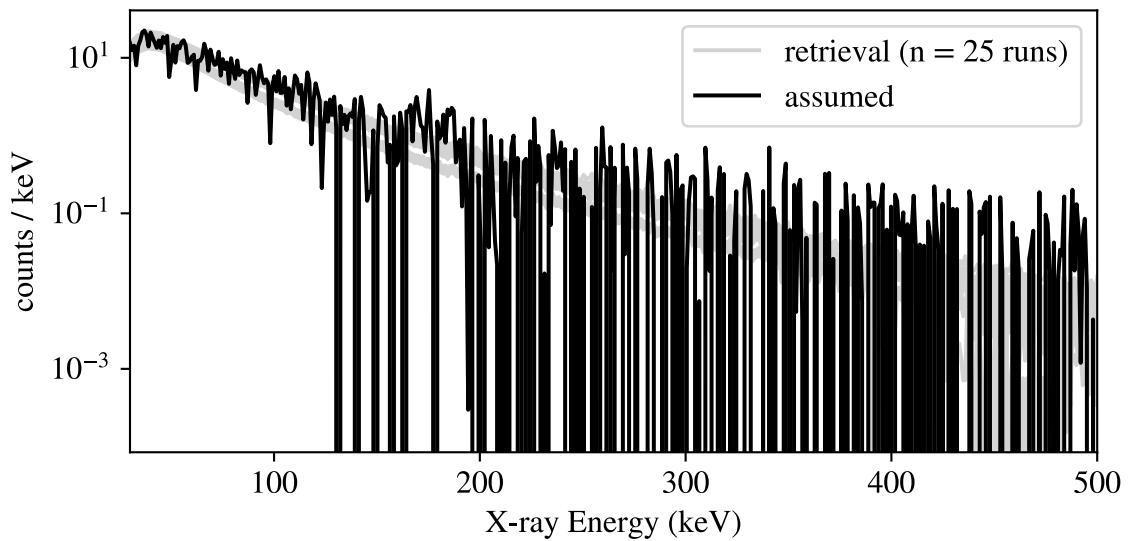
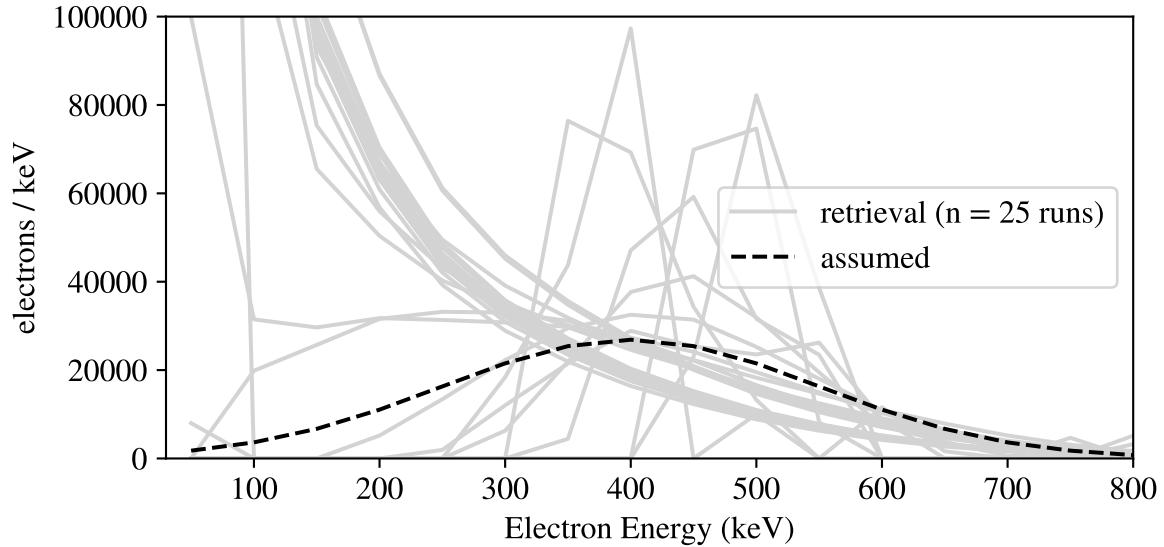


Figure 4.44: Retrieval of gaussian electron spectrum (sigma 150 keV, total electron count 1×10^7) from synthetic X-ray data.

LOOCV Residual Histograms ($n = 100$ random spectra)

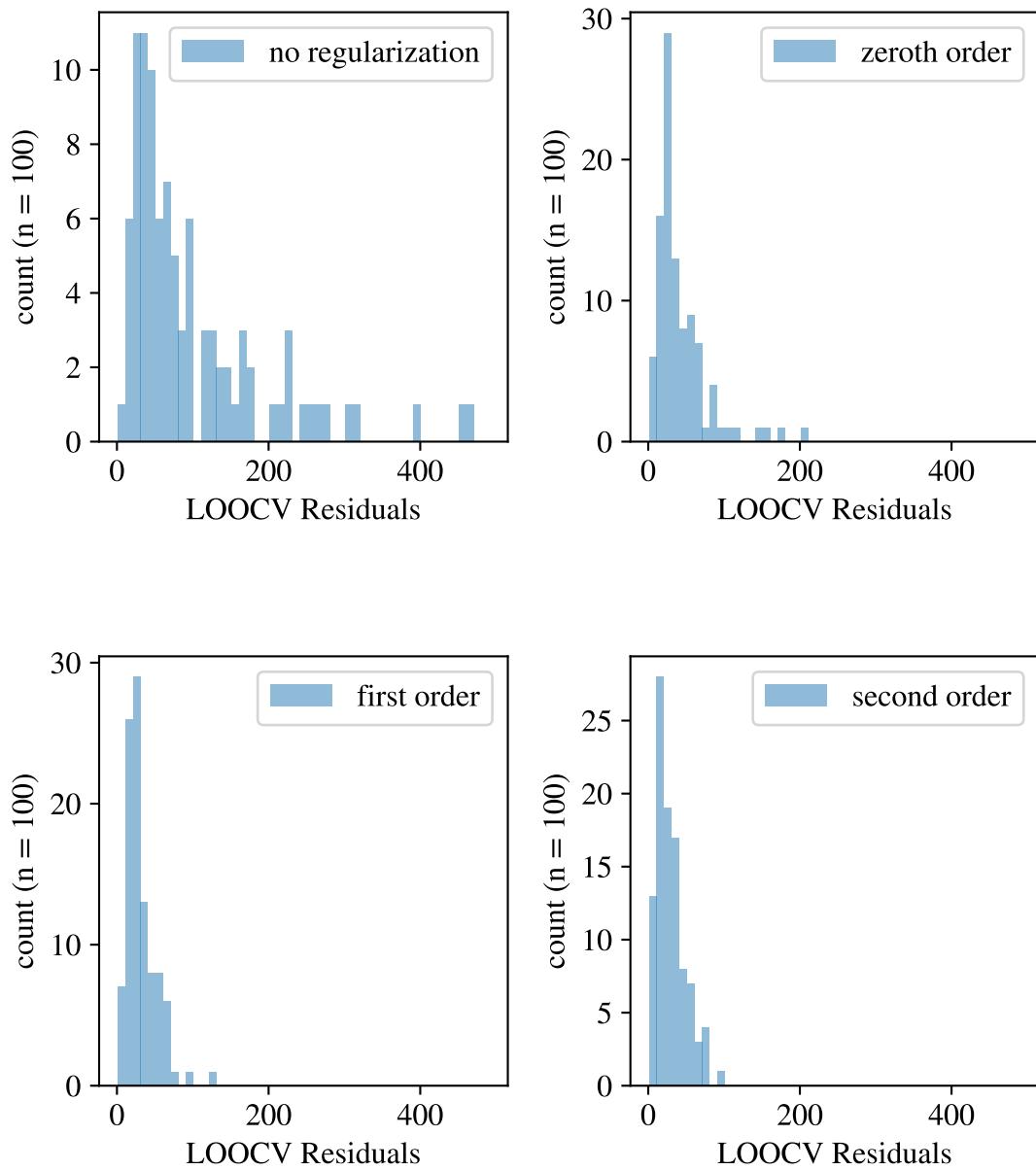


Figure 4.45: Distributions of accumulated errors in the cross-validation process used to select the regularization parameter for inversions of 100 randomly generated electron spectra.j

Chapter 5

Experiment Data and Analysis

5.1 Overview of The Balloon Experiments

High altitude balloons have an extensive heritage as platforms for experiments that take place in the upper atmosphere. In this chapter we will examine results from three balloon flight campaigns which carried X-ray spectrometers. The results of the previous chapters will be applied to determine as much information about the causative precipitating electrons as possible, while avoiding over-fitting the available data. The total precipitating flux and energy distribution of the electrons will be the main targets. The experiments on the balloon flights that we use in this chapter were all equipped with NaI scintillator detectors, and the analysis will be framed around the data that can be obtained from them. There will be questions that can be answered using detectors of a different design, particularly, those sensitive to the angular distribution of the incoming photons. These will be discussed at the end and left for future work.

The balloon flights used in this analysis were from the following campaigns.

Project Name	Location	Active	Primary Reference
ABOVE ²	Saskatoon, Saskatchewan	Summer 2016	None Current
BARREL	Antarctica	2013-2014	Millan et al. (2013)
EPEX	Fort MacMurray, Alberta	Summer 2018	None Current

The campaigns each consisted of a number of flights. Each had a somewhat different scientific target, which will be discussed in their analysis sections, however, all of them carried NaI scintillation detectors of a similar design, which we can use with the results from the previous two chapters. We will begin with a discussion of the detector design and operating principles.

5.2 Detector Design

The sodium iodide scintillation detector carried by flights on all three balloon campaigns was designed by Dr. Michael McCarthy at the University of Washington, Seattle campus. The instrument design has a heritage extending back to Winckler et al. (1958). The instrument specifications for the BARREL campaign flights follow, reproduced from Millan et al. (2013).

Attribute	Value	Comments
Mass	2.8 Kg	includes insulation and harness
Energy Range	20 keV to 10 MeV	
Electronics Resolution	2.4 keV / channel	reduced by binning
System Resolution	7 percent at 662 keV	
Effective Area	16 cm ² at 1 MeV	
Dead time Per Event	52 μ s	
Operating Temperature	-10 to 40 deg C	efficiency decreases below 15 deg C
Voltage Requirement	+/- 5 V DC	
Current Requirement	40 mA +, 10 mA -	depends on count rate

Simulations completed with GEANT3 (the predecessor to GEANT4) indicate that the detector behaves as essentially uncollimated, though the view through the atmosphere results in it having a field of view of approximately 60 degrees Millan et al. (2013).

A picture of the detector as flown on the ABOVE² campaign is shown in Figure 5.1. The unit is cylindrical, standing on a plexiglass mount and standing approximately 30 cm tall. The cylinder includes the NaI scintillation crystal, photomultiplier tube, and optical coupling. X-rays deposit their energy in the NaI crystal, and a fraction of the resulting light pulse is converted to a voltage pulse by the photomultiplier tube Millan et al. (2013). This voltage pulse is then converted to a digital value proportional to the deposited energy by the DPU and output on a serial port for processing and storage by the rest of the flight system. The disconnected wire harnesses visible in Figure 5.1 connect the detector to the various support hardware included with the flight system.

Temperature sensors and heaters were used to maintain the detector temperature at 15 deg. C during the flight. Power was provided by an array of lithium-ion rechargeable batteries, which are contained in the rectangular foam box behind the detector in Figure 5.1.

The detector sends measurements to a serial port as digital data. The outputs, reproduced from Millan et al. (2013), follow.

Product	Cadence	Energy Range (keV)	No. of Energy Channels
Rate Counters	4	-	-
Fast Spectra	0.05	20-1500	4
Medium Spectra	4	100-4000	48
Slow Spectra	32	20-10000	256

For the EPEX and ABOVE² campaigns, the instruments were designed to be recovered, and the data were recorded to onboard storage, with some telemetered results as a backup. For the BARREL campaign, the logistics involved in recovering all the payloads are difficult. While a subset were recovered, most were not, and the data were telemetered using an iridium satellite modem. For these flights, the analysis uses the data products published at NASA CDAWEB. The primary citation for this data set is Millan et al. (2013).

The functional layout of the detector is shown in the block diagram of Figure 5.2, which is reproduced from Millan et al. (2013).

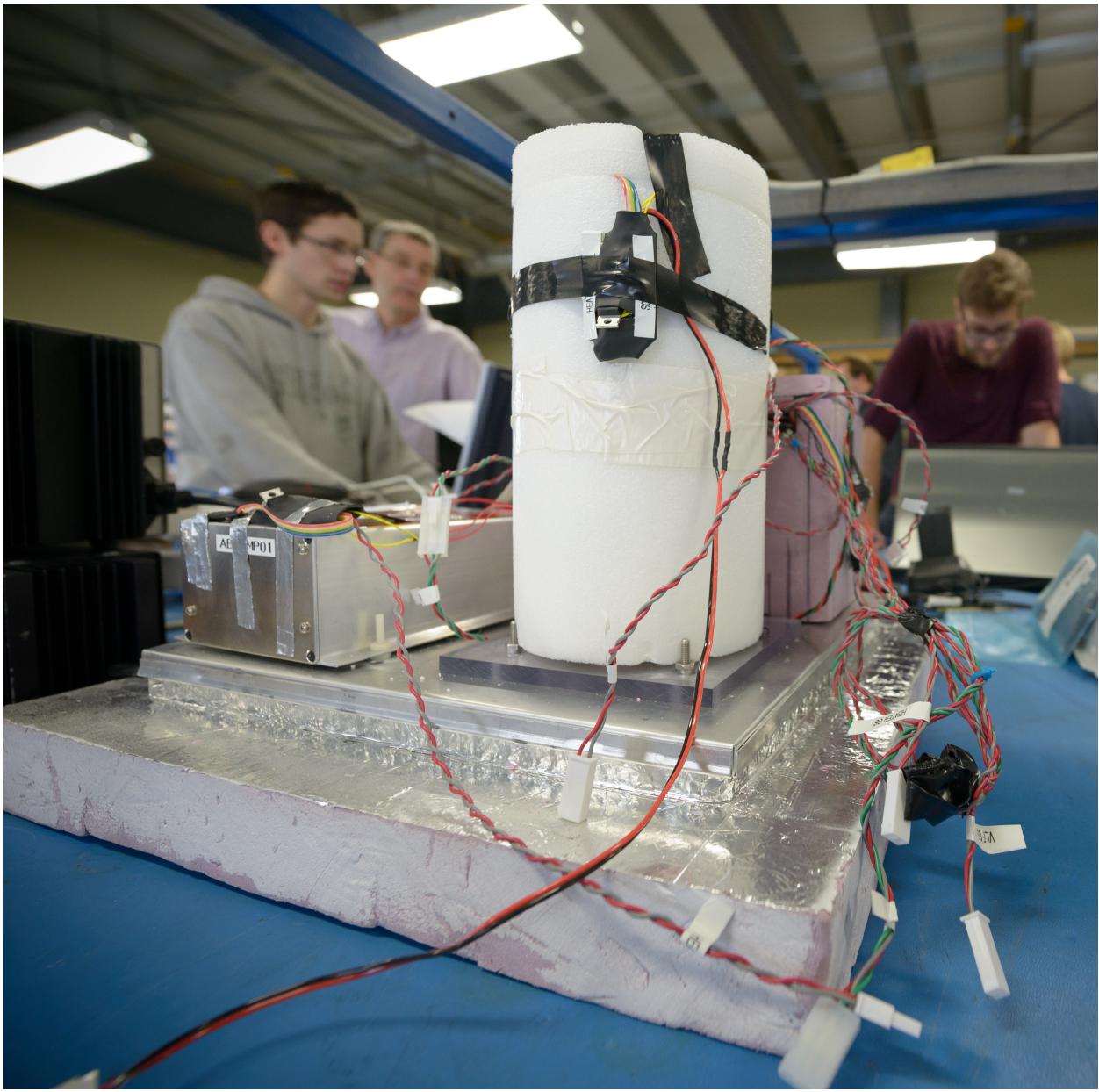


Figure 5.1: X-ray spectrometer (white cylinder) encased in insulation foam and installed with flight systems in balloon payload. The DPU (digital processing unit) is contained in the metal rectangular enclosure on the left. The lead collimator assembly is hidden under the white styrofoam insulation.

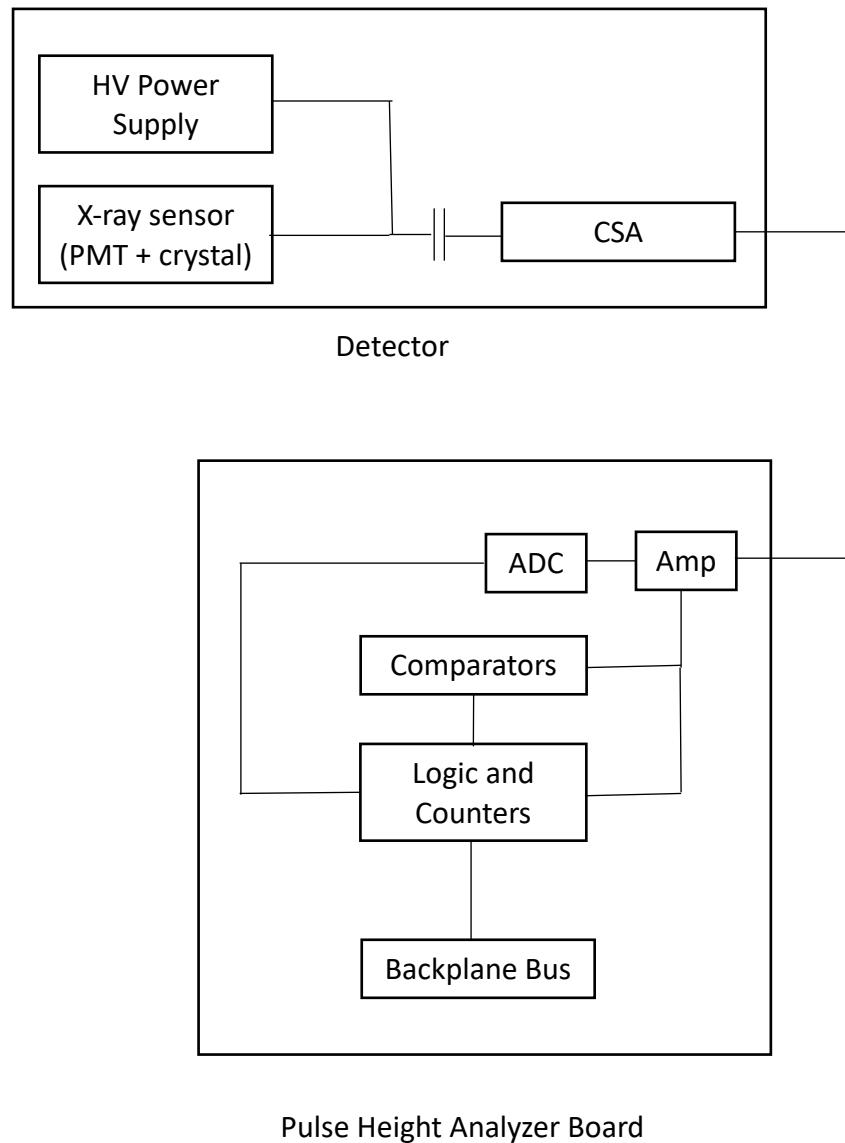


Figure 5.2: Block diagram of X-ray scintillation detector system

5.3 The ABOVE² Campaign

The ABOVE² balloon campaign was flown during the Summer of 2016. There were three flights, each using zero-pressure balloons manufactured by Near-Space Corporation, and designed for a float altitude of greater than 33 kilometers. The first flight was done to validate the operational aspects of the balloon system and telemetry. Flights 2 and 3 were equipped with the X-ray spectrometer. Flight 2 did not see any X-rays besides the expected background, however flight 3 observed a precipitation event that lasted several minutes. Unfortunately, flight 3 had to be flown using the X-ray spectrometer recovered from flight 2, due to an error during installation which damaged the flight 3 spectrometer. This normally wouldn't be a problem, except that the landing under parachute at the end of flight 2 created a shock which was later found to have damaged the crystal, and potentially the optical coupling between the crystal and photomultiplier tube. The effect of these defects is difficult to fully quantify post-flight, except to note that the energy calibration of the detector certainly changes. Further, damage to the optical coupling between the crystal and the photomultiplier tube will result in fewer photons created in the crystal reaching the photomultiplier tube, which will effectively reduce the detector sensitivity. These effects degrade the accuracy and precision of the results from the third ABOVE² flight, an effect which will be amplified during the application of the inversion techniques discussed in Chapter 4. Despite these problems, there is still value in examining the data set. The location, timing, and relative amplitudes of the X-ray event are intact, and a comparison between the observed X-ray flux and measurements from a nearby VLF (very low frequency) receiver site shows a strong, though not unexpected, correlation. Figure 5.3 shows ABOVE² science flight 2 being prepared for launch.

Figure 5.4 shows the total counts recorded by the X-ray spectrometer on science flight 2 as a function of time. After power-on and while the detector is still on the ground, the total counts are roughly constant. Once the balloon is launched (approximately 180 minutes since power on), the detector is removed from naturally occurring radioactive materials in the ground, and the count rate decreases. As altitude increases during the ascent, the count rate starts to increase, because the



Figure 5.3: ABOVE² balloon being filled and prepared for launch in Saskatoon, Saskatchewan, August 2016. The payload and flight train are laid out on the ground, extending approximately 10 meters to the right. Image credit: Joel Kesler / RMD Engineering.

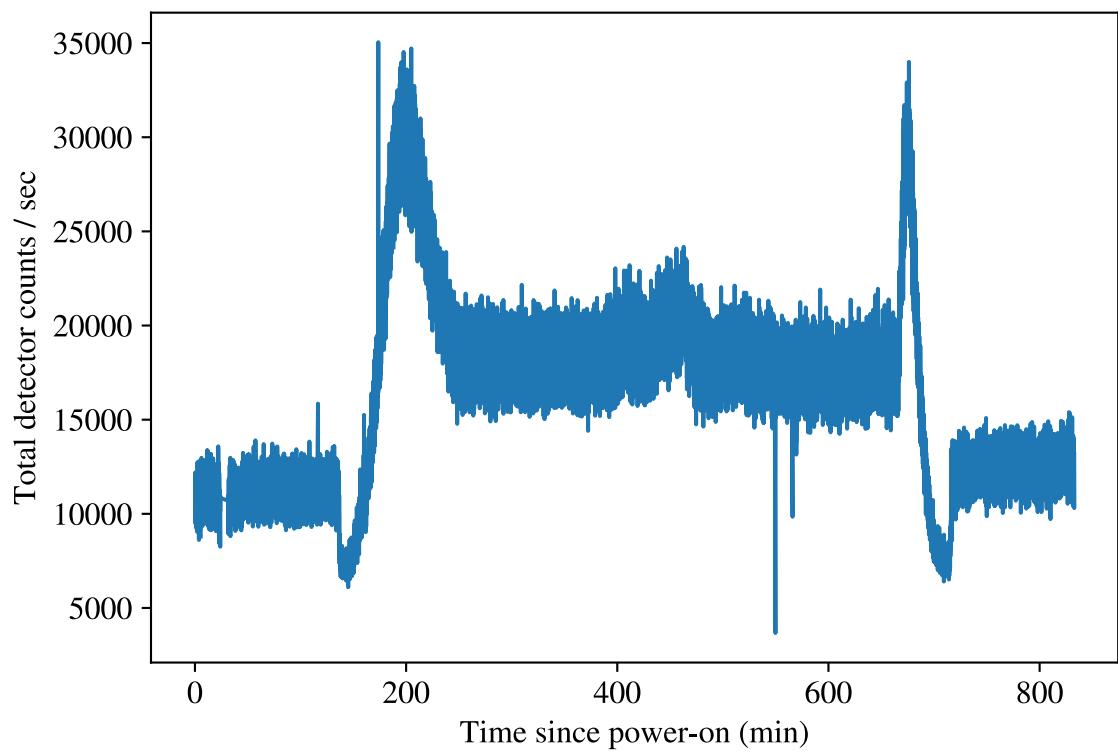


Figure 5.4: ABOVE² X-ray detector counts per second, summed across all energy channels, as a function of time since power-on of the system.

atmosphere, which acts as a shield from natural radiation from space, begins to decrease in density. After the balloon reaches equilibrium with the atmosphere, the count rates are again roughly constant. On descent, the reverse of this process occurs. The precipitation event observed by science flight 2 happens at approximately 400 minutes from power-on, and lasts for approximately 20 minutes. This is a very weak event.

A background-subtracted spectrum for the precipitation event is shown in Figure 5.5. The background was subtracted by taking an average spectrum of the recorded data before the event happened. Above approximately 200 keV, the background subtracted spectrum becomes ragged because it is not distinguishable from the background spectrum. Second-order Tikhonov regularization with preconditioning was applied to the background subtracted spectrum from 30 keV to 200 keV to obtain a model spectrum using a regularization parameter of $\alpha = 10^5$, determined using cross-validation. The solution is shown in Figure 5.6. There is not much structure in the solution, due to the application of a high degree of regularization ($\alpha = 10^5$).

Figure 5.6 shows that the best model generated by regularization fails to describe the data in the low energy region. The high value of α found through cross-validation reflects this fact and reduces the amount of detail in the model accordingly. The model spectrum has an average energy of approximately 150 to 200 keV. The width of the distribution is not able to be effectively determined for this data set owing to the high value of α .

The regularized solution fails to qualitatively describe the measured X-ray spectrum. An attempt to fit exponential and mono-energetic spectra to the event have the same problem. Fitting to the available data in high energy ranges (more than 150 keV) results in model spectra that “overshoot” the measured data in lower energy ranges. This indicates a problem with the data rather than the fitting procedures.

It is believed that the detector was damaged on a previous flight, resulting in measured data that cannot be described by any model.

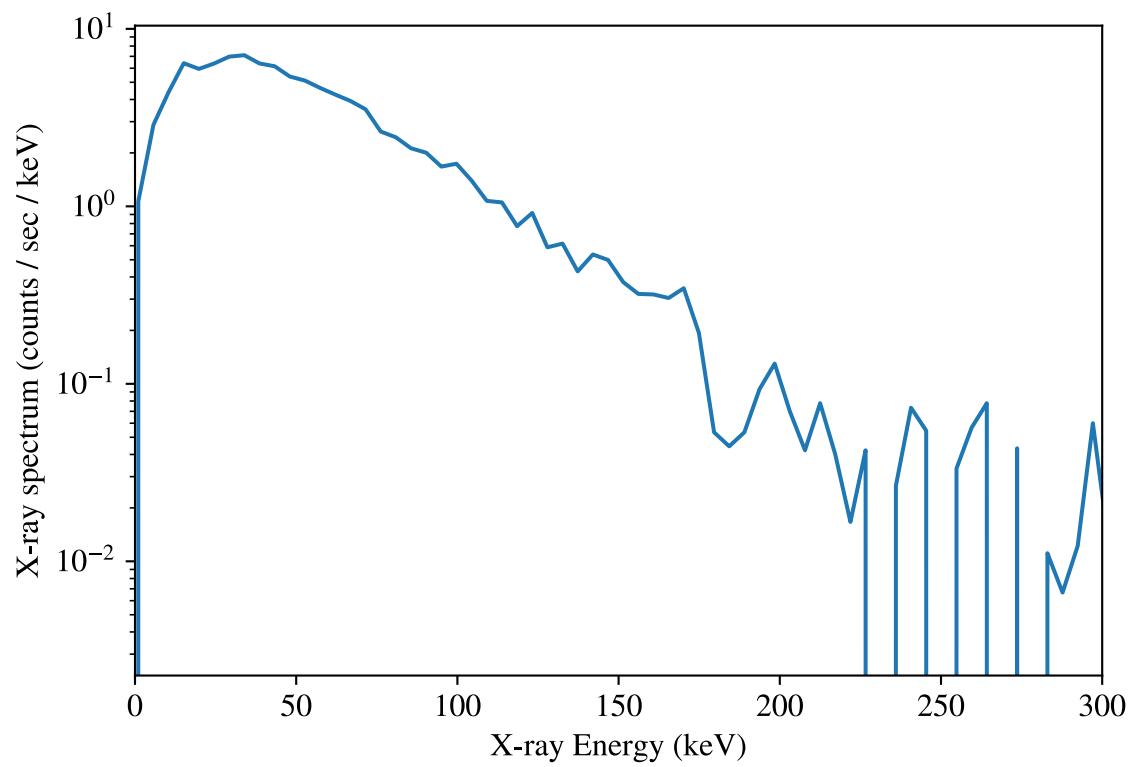


Figure 5.5: Background-subtracted X-ray energy spectrum of the precipitation event observed by ABOVE² science flight 2. The integration time is 900 seconds, and the bin width is 1 keV.

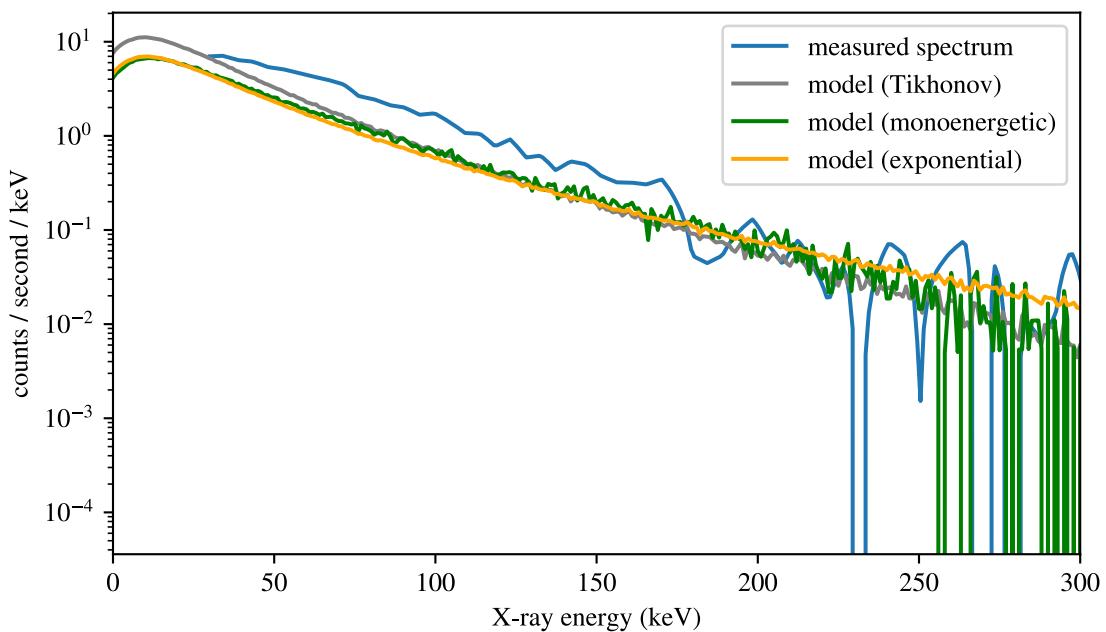


Figure 5.6: Application of Tikhonov regularization to the background-subtracted precipitation event measured by ABOVE² science flight 2. The algorithm applied is as described in Chapter 4. Second-order regularization with left and right preconditioning is applied together with a positivity constraint to generate the solution.

5.4 The EPEX Campaign

The EPEX balloon campaign was flown in August, 2018 at Fort McMurray, Alberta, Canada. The campaign had similarities to the ABOVE² flights, but had the primary goal of testing a new generation of X-ray detector which used a coded-aperture approach to generate images of the X-ray distributions caused by electron precipitation. My role in the EPEX project consisted of the following:

- Design and construction of the flight train and mechanical harnessing for the payload.
- Design and construction of the thermal and electrical support systems for the X-ray instruments.
- Implementation of a dual-mode radio tracking and telemetry system consisting of both VLOS (visual line of sight) and BVLOS (beyond visual line of sight) satellite and point to point radio modems.
- Instrument testing after integration to the payload

The analysis of data from the new detector is outside the scope of this thesis. Some of the EPEX flights carried an instrument which was functionally identical to the one flown on ABOVE². One of the flights which carried this instrument observed an intense precipitation event while it was at altitude. The event is shown in spectrogram form in Figure 5.7. Immediately apparent is the temporal structure and how it differs from the ABOVE² event. This event started at a sharply defined time (6:40 UT, see Figure 5.7) and quickly reached its peak before decaying. The fact that the intensity was so high makes this a good target for the application of the inversion techniques used in Chapter 4.

A background-subtracted spectrum of the precipitation event is shown, along with the reconstruction using the techniques of Chapter 4, in Figure 5.8. The retrieved precipitating electron

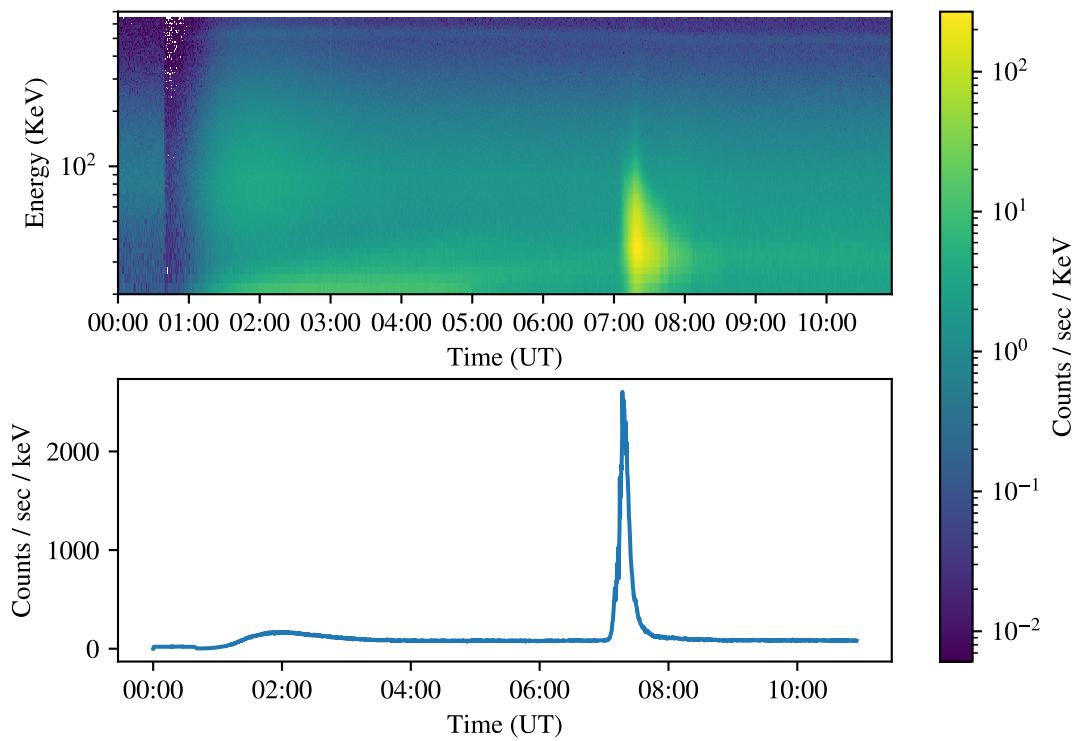


Figure 5.7: EPEX X-ray spectrometer data in spectrogram form. The precipitation event was observed at 07:00 UTC. The total counts per second are shown below the spectrogram as a function of time.

spectrum is shown in Figure 5.9. It was generated using second-order Tikhonov regularization with left and right diagonal preconditioning and a positivity constraint. The X-ray spectrum data from 70 to 500 keV were fit. The retrieved electron spectrum has a very mono-energetic structure at approximately 80 keV. The model agrees well with the data, unlike the result from the ABOVE² event, because this detector was flown in a refurbished and undamaged state. The total precipitating electron flux is estimated by the model as 4.8×10^8 electrons cm² / second. This was a very bright event which was well-defined in time and energy. The low value of $\alpha = 3.5 \times 10^3$ (compared to the ABOVE² event) suggested by cross-validation indicates that a good fit to the data was found by the reconstruction algorithms without requiring intense reconditioning of the problem. This is consistent with the fact that this bright event has a very high signal to noise ratio in the background subtracted data. This event can serve as an example of the usefulness of the regularization techniques in analyzing X-ray data. The regularized model in Figure 5.8 shows better agreement with the data than either the best exponential or monoenergetic fits. The reduced χ^2 statistics for the monoenergetic and exponential fits are 0.53 and 58.0, respectively. The reduced χ^2 for the regularized model is 0.75. This indicates that the exponential fit is a bad description of the X-ray data set for this event. The monoenergetic fit is better, but the regularized solution is best, having a reduced χ^2 statistic closest to 1.0. The strength of the regularization methods, combined with cross-validation, is that, with mild assumptions, we can a model-independent characteristic energy and width of the electron distribution.

5.5 The BARREL Campaign

I examine one flight from the NASA BARREL balloon campaign, which occurred during 2014. This flight was launched with an X-ray spectrometer similar to those flown on the EPEX and ABOVE² campaigns, and recorded an electron precipitation event while magnetically local to an orbiting spacecraft (Van Allen Probe B), which provides conjugate measurements of energetic electrons via the MAGEIS instrument. This flight was analyzed by Halford et al. (2015), who de-

termined that the precipitation event occurred in the aftermath of an ICME (interplanetary coronal mass ejection) shock. Since the BARREL balloon was airborne and recoding data while the space-craft was on the same approximate L-shell, Halford et al. (2015) could complete an “end-to-end” study of the recorded precipitation event, from creation, to detection by an RBSP spacecraft, to the precipitation in the atmosphere and resulting creation of X-ray radiation detected by the balloon. We will analyze the event measured by the balloon in this section.

A spectrogram of the event observed by Halford et al. (2015) is shown in Figure 5.10. This is a very low intensity event, but it occurs over a time period of tens of minutes. We will use an average spectrum taken over the event time range. The background-subtracted spectrum of the event, and the best-fit model spectrum is shown in Figure ???. Finally, the retrieved electron spectrum is shown in Figure 5.12. The retrieved electron spectrum has two main components, a distribution at around 250 keV, and a smaller high energy component at around 600 to 700 keV. The significance of the high energy component is not clear, since the X-ray spectrum has a noise floor which begins at approximately 200 keV.

5.5.1 Energy Selective Scattering

As a validation of the inversion method, the electron spectrum can be compared with in-situ measurements of the corresponding electron flux measured by the satellite in the magnetosphere. The in-situ data provides a limit on the precipitating electron flux, which is the product of the available flux and the scattering efficiency. A subset of the flux measured by the satellite will precipitate into the atmosphere.

Spacecraft measurements of energetic electrons do not generally resolve the population within the loss cone. We show that the techniques developed in Chapter 4 can be applied to infer this energy range from balloon measurements, by building on an analysis published by Halford et al. (2015). Halford et al. (2015) reported on an energetic electron precipitation event that happened while one of the Van Allen Probe B was at a similar L-value and close in MLT with a BARREL balloon 2L that measured the resulting X-ray spectrum. Their analysis provides an end-to-end

picture that describes the scattering processes which drove the event, and the measurable impact it made on the atmosphere.

The Van Allen Probes are a pair of orbiting spacecraft which carry instruments to understand electron acceleration in the radiation belts. The Magnetic Electron Ion Spectrometer (MAGEIS) instruments are magnetic spectrometers that provide in-situ measurements of the energies, angular distributions, and differential fluxes of electrons in the radiation belts Blake et al. (2013); Spence et al. (2014). There are four magnetic spectrometers on each spacecraft, each designed for a specific energy range. Electrons enter each instrument through a collimator and are deflected to a pixelated detector surface by a uniform magnetic field. We will use the pitch angle resolved, background-subtracted differential energy fluxes provided by the Energetic Particle, Composition, and Thermal Plasma (RBSP-ECT) investigation Spence et al. (2013).

A key assumption that must be made in the analysis of the precipitation event from Halford et al. (2015) is that the available MAGEIS data in the lowest (approx. 8 degree) pitch angle bin represents the energy distribution of electrons that are scattered into the loss cone. This is because MAGEIS does not resolve fluxes within the loss cone. The assumption is that the electron spectrum in the loss cone is the same as the electron spectrum in the lowest pitch angle bin, except that it has undergone a scattering process which selects for a range of electron energies. but leaves the rest of the spectrum unaltered. ? uses quasi-linear theory and satellite data to show that chorus waves lead to rapid pitch angle scattering over a narrow pitch angle range near the loss cone for electrons from 100 eV to a few keV.

Halford et al. (2015) apply exponential and mono energetic fits to the measured X-ray data, and also compared with the energy spectra observed near the loss cone from Van-Allen Probe B MagEIS data which was at a similar L-value and close in MLT to BARREL payload 2L (see Figure 4 in Halford et al. (2015)). The best fit to the observed X-ray spectra was found when using the observed energy spectra from Van Allen Probe B. The agreement reveals that the measured event does contain enough information to distinguish between electron beams with zero width (mono-

energetic), and beams which extend to the highest modelled energy (exponential). We use this event as a test of the regularization method to see how much more information about the scattered electron spectrum it can extract. In particular, under the assumption that the scattering processes responsible for the observed precipitation select an energy range from the magnetically conjugate Van-Allen Probe measured electron spectrum, we evaluate if that energy range can be determined from the X-ray data.

The comparison of mono-energetic and exponential fits to the balloon X-ray spectrum done by Halford et al. (2015) is reproduced in Figure 5.13. We apply the constrained second-order regularization technique to obtain a model electron spectrum. The X-ray spectrum predicted to result from this retrieved spectrum is plotted with the measured X-ray data, and shows closer agreement with the BARREL spectrum than the best fits that can be produced by assuming the mono-energetic or exponential forms.

Data from the RBSP MAGEIS instrument used in the analysis of the Halford et al. (2015) event is shown in Figure 5.14. Under the assumption of strong scattering, with a full loss cone. The topside electron flux from a spacecraft in magnetic conjunction with a given location on Earth is given as $F \approx (B_i/B_{top})J_{top}\Delta\Omega$ Kasahara et al. (2018). B_i and B_{top} are the magnetic field strengths at the precipitation point and spacecraft, and J_{top} is the electron flux at the spacecraft. $\Delta\Omega$ is the width of the loss cone.

This estimate requires the assumption that the electrons measured in the lowest pitch angle bin of the MAGEIS data (approximately 8 degrees) represent the energy spectrum of electrons in the loss cone. If both of these assumptions are true, then we can follow the technique of ?, who relates We use $\Delta\Omega \approx 2.5^\circ$ as estimated for this event. Applying the apparent extent of the electron energies in the model spectrum, which mostly range from 100 to 350 KeV, to the spectrum in the lowest pitch angle bin measured by MAGEIS estimates a topside electron flux of $8.3 \times 10^6 \text{ e}^-/\text{cm}^{-2}\text{s}^{-1}$. This can be compared with the precipitation rate of $1.6 \times 10^6 \text{ e}^-/\text{cm}^{-2}\text{s}^{-1}$ obtained using the count rate and energy spectrum measured by the balloon spectrometer and the model energy spec-

trum. Under the assumptions we made, this indicates that the regularization method can provide useful estimates of the effective energy range over which the scattering processes for a given event operate.

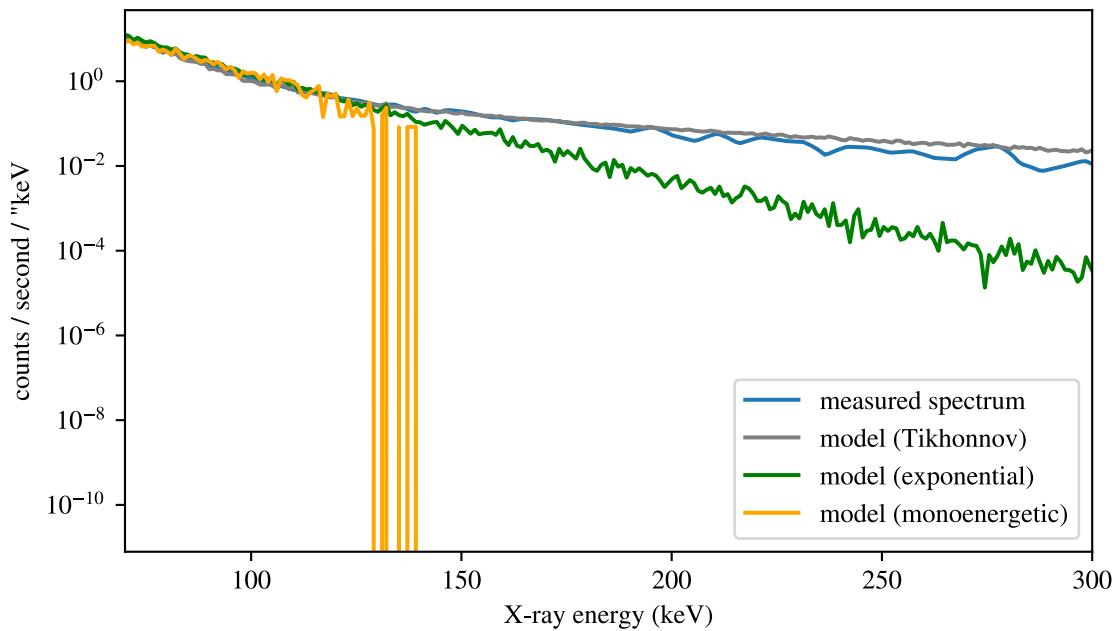


Figure 5.8: Background-subtracted X-ray event measured by the EPEX spectrometer. The model fit using Tikhonov regularization is shown overlaid.

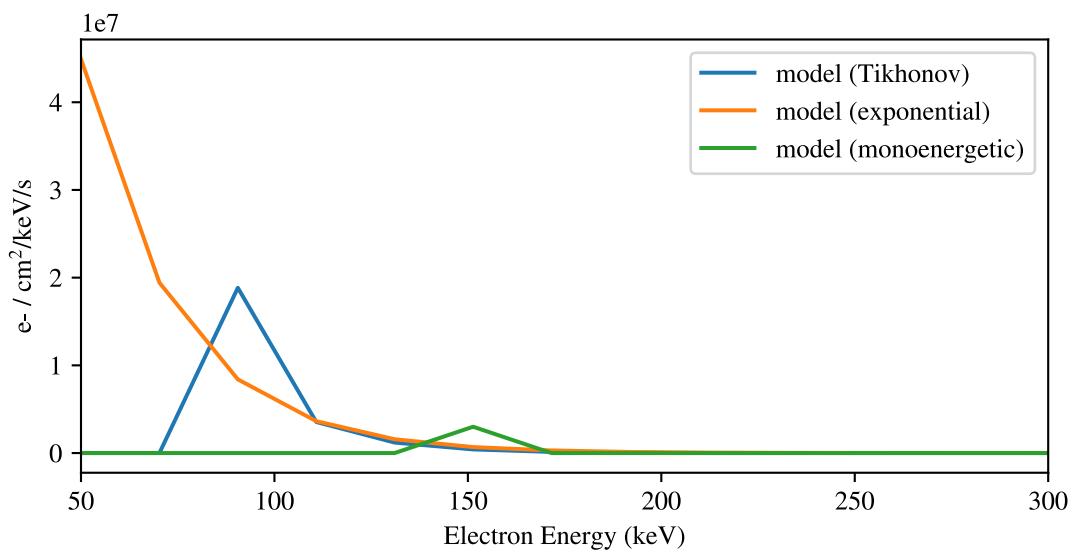
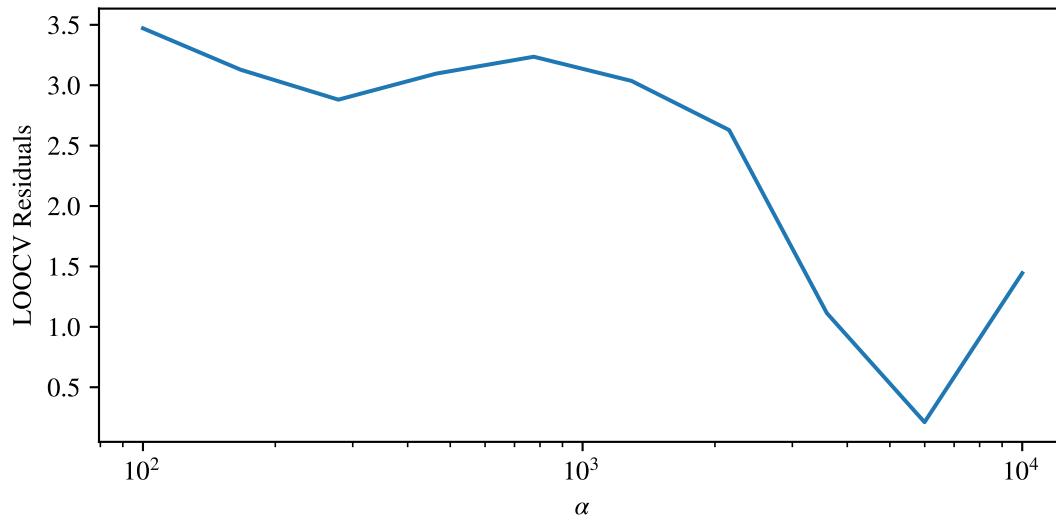


Figure 5.9: Precipitating electron spectrum retrieved from EPEX X-ray data using the techniques of Chapter 4. The required value for α is low and well-defined through cross-validation.

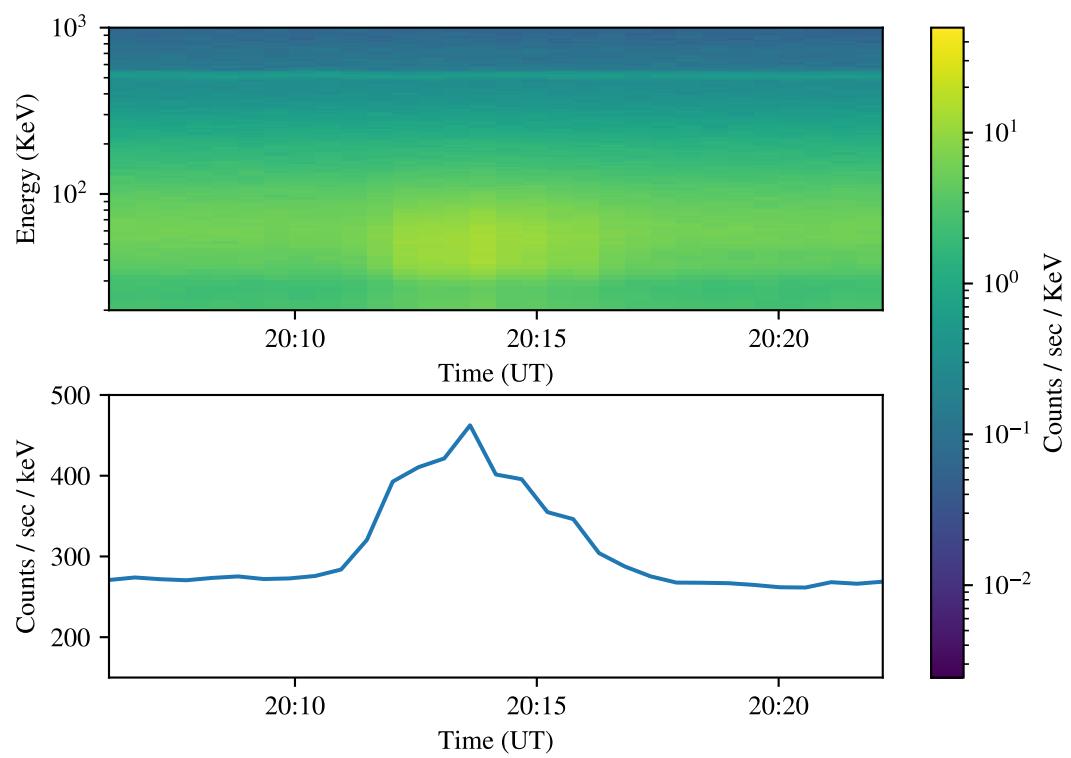


Figure 5.10: Spectrogram and raw count rates of the event measured by BARREL and studied by Halford et al. (2015).

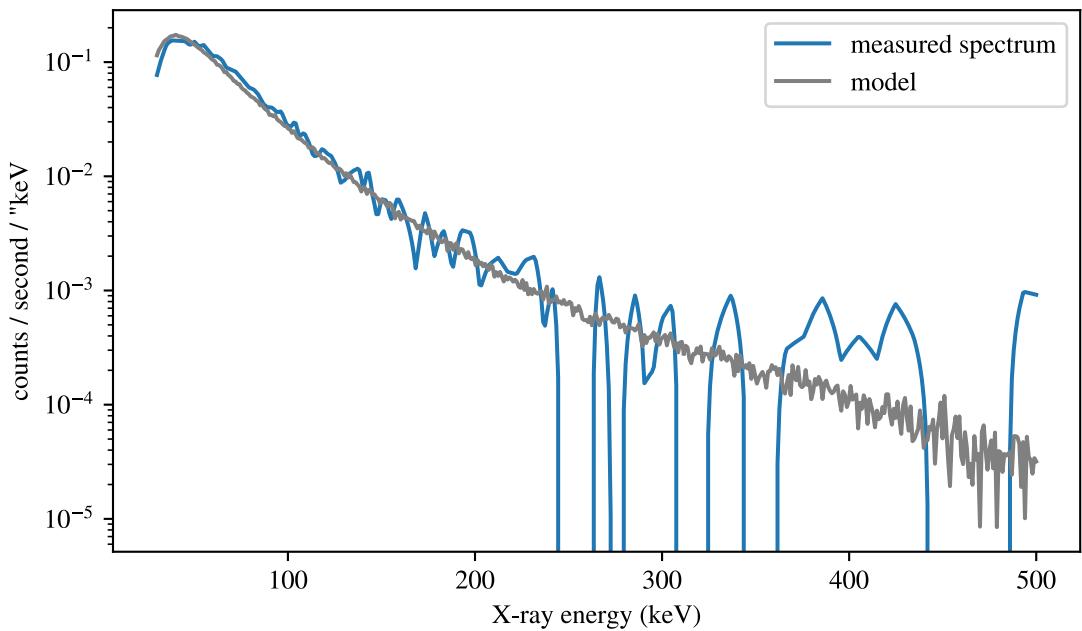


Figure 5.11: Background-subtracted X-ray event measured by the BARREL spectrometer. The model fit using Tikhonov regularization is shown overlaid.

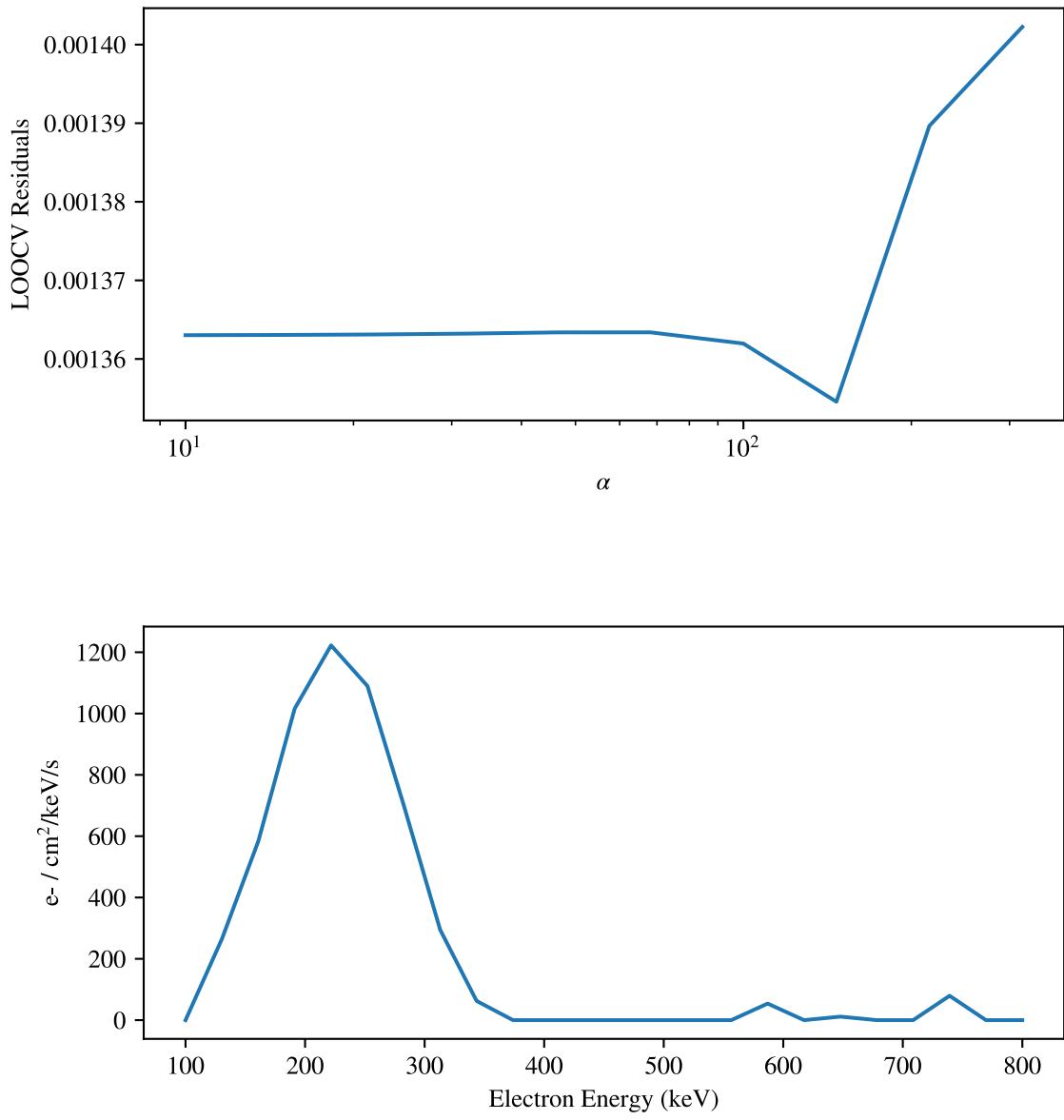


Figure 5.12: Precipitating electron spectrum retrieved from BARREL X-ray data using the tfech-
niques of Chapter 4. The required value for α is approximately 5×10^2 .

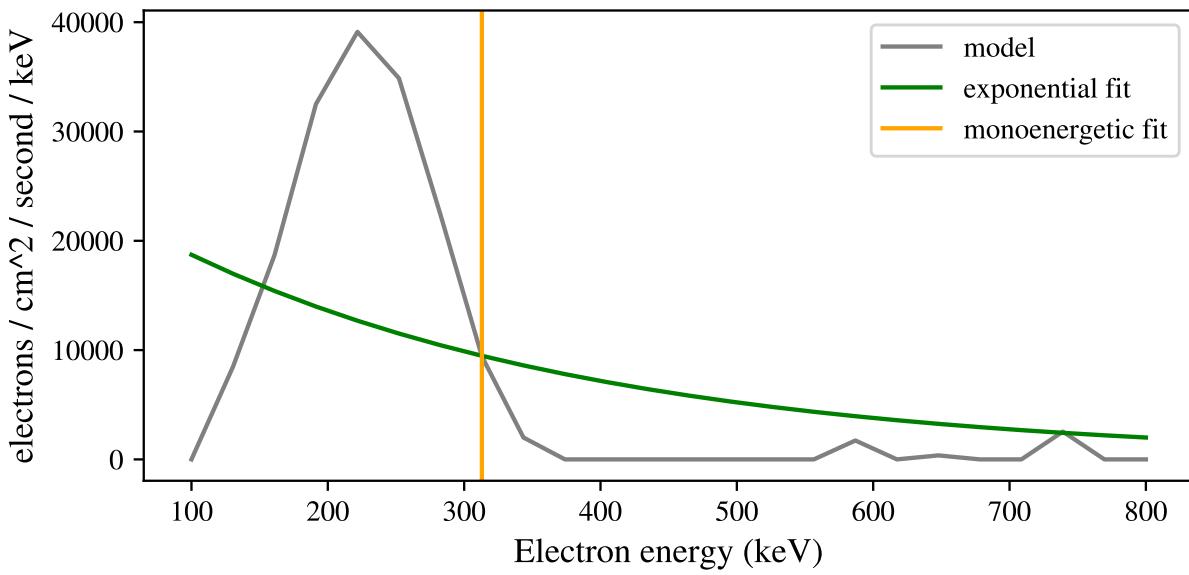
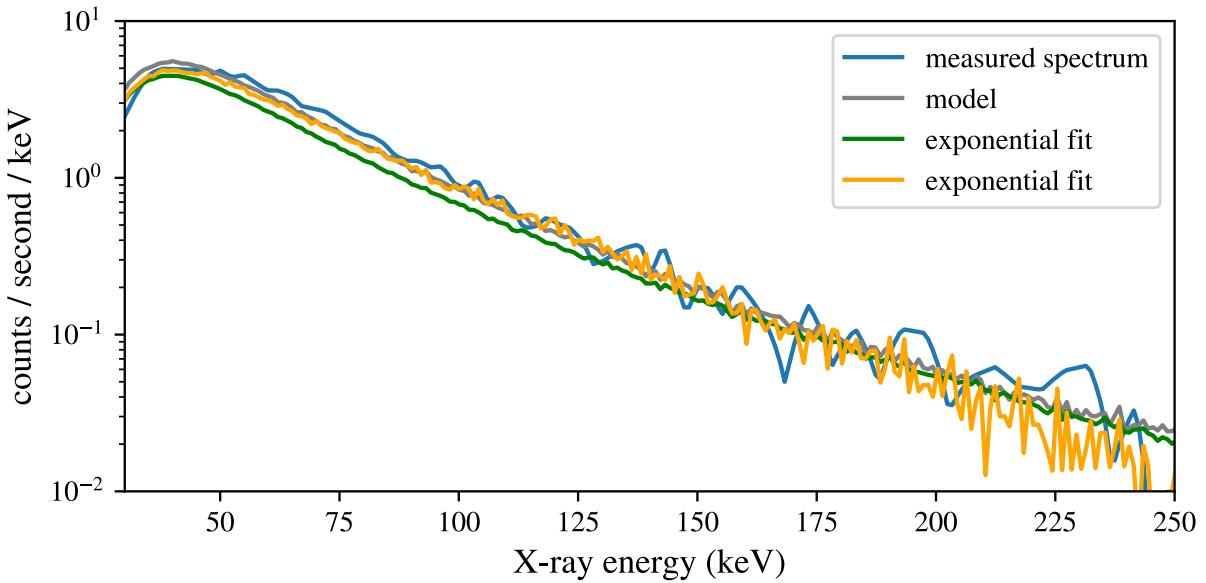


Figure 5.13: Mono-energetic and exponential fits to the BARREL event from Halford et al. (2015), with the fit obtained by applying the techniques developed in Chapter 4 (preconditioned, constrained second-order Tikhonov regularization using cross-validation). The total electron flux for the regularized model is found to be 1.6×10^6 electrons/cm²/sec.

MAGEIS Differential Electron Energy Flux
Average over 600 seconds, 2014-01-09:20:10:00 UTC

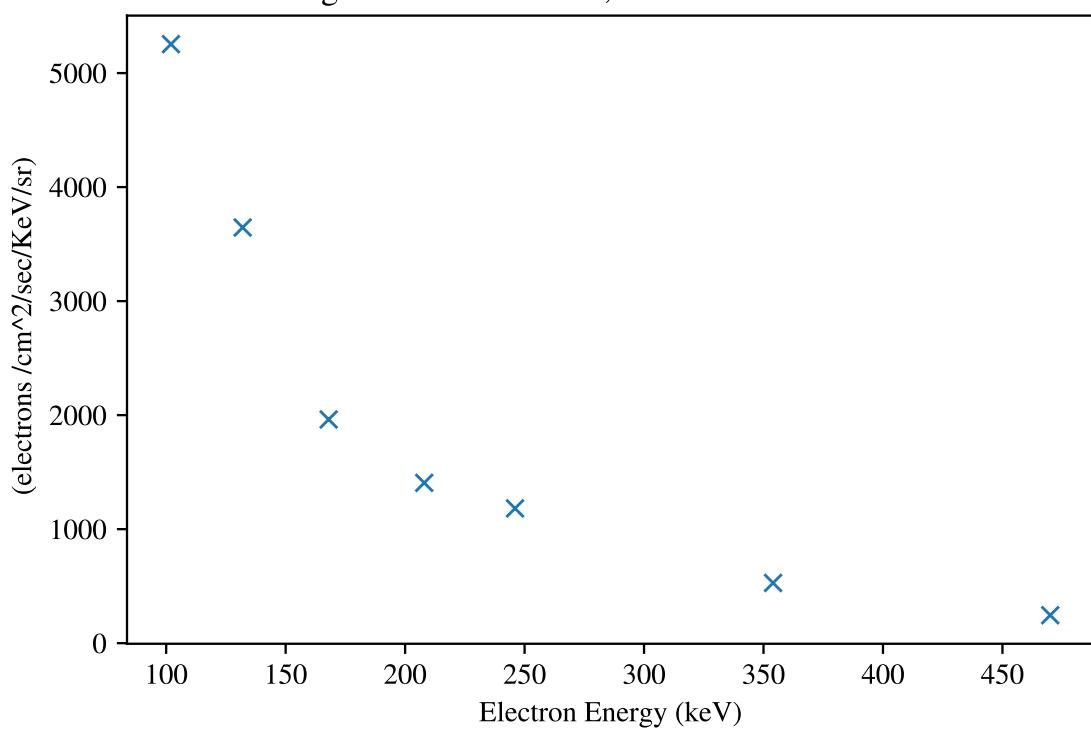


Figure 5.14: 600 second average differential electron energy flux measured by the RBSP-B MAGEIS instrument and used in the analysis by Halford et al. (2015).

Chapter 6

Conclusions

6.1 The Inverse Problem

The primary question addressed by this thesis is how much information can reasonably be expected to be obtained from balloon measurements of X-ray spectra. The key results for the analysis methods are as follows:

- When accompanied by an accurate forward model of the expected instrument response given a precipitating electron spectrum, the electron spectrum can be recovered provided that the measurement noise is low enough.
- The required statistical purity in the measured X-ray spectrum depends on the shape of the particular spectrum being recovered, and the amount of detail sought in the solution.
- Preconditioning techniques such as diagonal scaling improve the condition of the inverse problem with no additional assumptions or bias.
- Biased estimators, such as truncated singular value decomposition, or Tikhonov regularization, can be applied to the inverse problem to make it tractable in realistic simulations and actual experiments.
- The application of cross-validation methods provides a way to balance the required stability of the solution against the inherent ill-conditioning of the problem, and is capable of selecting regularization parameters which represent the available information in the data without requiring heuristic arguments or detailed assumptions on the form of the output.

These conclusions, when taken together, form a picture of the X-ray inversion problem that is quite different than what is being currently used in experiments. As discussed in Chapter 4, the ill-conditioned inverse problem needs to be addressed to obtain meaningful results. The conventional approach of assuming a particular functional form for the model electron spectrum solves the ill-conditioning by reducing the parameter space of the model. The analysis subsequently uses the well-known χ^2 tests to select an optimal fit and produce an error estimate for that fit. This approach, favoured for its relative simplicity and ability to produce an error estimate, necessarily gives a misleading solution. The reason is that the uncertainty presented by the ill-conditioning of the problem, which results in the vast majority of the uncertainty in the solution, is effectively “swept under the rug” and contained in the assumption on the model function chosen for the solution. The remaining error analysis is trivially based on counting statistics and cannot be used to identify between different model spectra, except when significant a-priori information about the precipitating spectrum is known.

This perspective assumes that very little useful information can be obtained from X-ray measurements. This thesis offers an alternative picture: instead of reducing the information extracted from the data to fit the analysis methods, we adapt the analysis based on what can be determined from the data in a given scenario. This picture is much more optimistic, but critically, it isn’t unrealistic. The linearity of the forward problem, combined with the widespread availability of sophisticated particle transport software, makes constructing an accurate representation of the balloon experiments possible. Then, different frameworks for solving the inverse problem can be tried and compared using simulated data. We have found that some relatively simple linear algebra can be used to relieve most of the ill-conditioning of the problem, without loss of generality, but that the transformed problem still requires the use of a biased estimator to give useful, general, results in realistic situations.

Tikhonov regularization, which is a specific form of the general techniques which modify the singular value spectra of the response matrix of the forward problem, provides a way to continu-

ously adjust the effective dimensionality of the solution, and therefore, the amount of information solved for. Combining this parameterization with cross-validation allows for the optimized selection of the amount of information contained in the solution, so that it neither underfits, no overfits, the available data. In particular, second-order Tikhonov regularization effectively blurs the detail in the represented solution to match the data, and requires no general assumptions on the shape of the model spectrum. Tikhonov regularization can be done using constraints which further reduce the parameter space of the problem at no additional cost. Particularly, we have found that a simple positivity constraint reduces the dispersion in model spectra. This ensemble of methods is modular - in the sense that additional constraints may be added, without additional loss of generality, if any a-priori information about the precipitating spectrum is known, for example, if magnetically local spacecraft measurements are available.

The primary cost of the use of biased estimators is that they force the mismatch between the information in the input and output of the ill-conditioned problem out in the open - we now have to handle it, and its consequences. This can be taken as a more honest view of the inverse problem in general. While the techniques in Chapter 4 provide strong results when applied to synthetic data, what they cannot do, is provide standard error bars on the shape of the output spectrum. This is because poisson statistics and their simple interpretation no longer apply in the same sense under biased estimators. Statisticians have said that the determination of meaningful error estimates for this kind of problem is one of the frontiers of their research. This places it well outside the scope of this thesis to address, and at the time of writing, only preliminary results exist, and none in the less abstract form needed for practical application to this particular problem.

The key point is this: it is better to have the fundamental lack of information that causes the ill-conditioning of the inverse problem accounted for in some way, than to conceal it behind an assumption which cannot be justified in most experiments. By honestly attempting to, at least heuristically, account for the amount of information available in a given dataset and in a particular situation, we can learn more than by the misapplication of simple, but tractable, analysis methods.

A secondary conclusion comes from the necessity of a forward model of the precipitation process. The inverse problem, owing to its ill-conditioning, is highly sensitive to both noise, and bias, in the input data. It follows that the accuracy and precision of the forward model used to represent the problem must be high enough not to significantly affect tests on simulated data. The exact instrument response plays a key role in the forward model, and is something that cannot be adequately determined through simulation alone. Lab measurements against known X-ray spectra are required for every different experiment. These are not always easy to accomplish, owing to the somewhat random, and discrete, X-ray peaks produced by commonly available radioactive sources. Further, the instrument response needs to be well-approximated as linear with regards to energy, in order for it to be applied to the linear framework developed to address the inverse problem. The critical simplification made in Chapter 3 was that the instrument response in counts to a monochromatic beam of X-ray photons could be represented using a Green's function approach. Additionally, a large simplification of the forward problem was obtained by essentially ignoring the angular distribution of the precipitating electrons. In this sense it made the problem separable, and reduced the inverse problem to one dimension in energy only.

The environment surrounding the particular detector will need to be simulated in the forward model if it changes significantly between lab testing and installation in a balloon. Balloon experiments are hosted in insulated containers, which are usually surrounded by other flight components, some of which can interact with natural radiation in significant ways. None of this is insurmountable, but needs to be made explicit in light of the primary conclusions. All of the work in Chapter 4 and the subsequent analysis of experimental data applies because the forward model of the experiment was available and verifiable against lab data. Future experiments, even ones flown with changes only to the support systems and flight hardware, rather than instrument itself, will need to re-evaluate the conclusions of Chapter 4, based on the altered forward model of the problem. This could be seen as a lack of generality. It is instead an acknowledgement that the inverse problem is fragile to small changes in the input data and response matrix. This is not as significant of a

problem when artificially suppressed by strong assumptions on the output.

6.2 Application to Experimental Data

When applied to the available body of experimental data (U of C balloon flights and the BARREL campaign), application of the analysis techniques we have developed support the following conclusions:

- X-ray data obtained from sodium iodide scintillation detectors is sufficient to determine useful information about the precipitating electron spectrum, including the overall electron flux and approximate energy distribution.
- In the case of the precipitation event recorded by a balloon when magnetically local spacecraft measurements of the electron spectrum were available, the data can be used to infer a spectrum compatible with the spacecraft measurements, in the sense that the retrieved energy spectrum produces approximately the precipitating electron flux that would be expected based on the spacecraft measurements under the strong scattering assumption with a full loss cone.
- When the instrument is not performing according to the forward model, as in the U of C balloon flight where it was damaged, our analysis of the data shows a clear problem with the fit.

The first conclusion has been known since the earliest balloon experiments equipped with X-ray detectors were flown. The questions that have persisted over the subsequent decades have all hinged on the interpretation of the measured data, and the amount of information that can be obtained from them. The primary contribution of this thesis to the problem is to demonstrate that the amount of information available is well-represented as a continuous function of the bias and noise level of the particular experiment flown. This function can be parameterized as a one dimensional optimization, which prevents the introduction of experimenter bias.

The second conclusion relies on available magnetically connected satellite measurements to the balloon measuring the precipitating spectrum. These occurrences are subset to the available balloon data, and owing to the logistical challenges of timing balloon release against spacecraft orbits, and the expense of balloon flights in general, are not common. The event analysed in Chapter 5 from Halford et al. (2015) provides a picture that extends from the trapped population in the radiation belts, through scattering and finally precipitation to the atmosphere and detection by the balloon spectrometer. The measured spectrum at the spacecraft and the spectrum measured by the balloon are not the same - in general, the precipitating population will be subset to the population measured by the spacecraft, and the population will undergo a largely unknown scattering process prior to precipitation. Having measurements of the population from the spacecraft and the balloon does, however, permit a more detailed picture of the end-to-end process to be created. Under the assumption of strong scattering and a full loss cone, and the assumption that the electron population in the lowest pitch angle bin resolved by the spacecraft represents the population in the loss cone, then the precipitating electrons should be subset. Under the additional assumption that the scattering processes responsible for the precipitation are energy selective (which is reasonable, based on the wave-particle dynamics explored in Chapter 2), the balloon spectrum represents the population that was scattered into the loss cone. Since the analysis techniques of Chapter 4 allow the determination of the energy spectrum and total flux of the precipitating population, we have an end-to-end picture of the scattering and precipitation process. Given an estimate of the energy range of the precipitating electrons, the work done by Kasahara et al. (2018) allows the estimation of a precipitating electron flux from the spacecraft data. The results were consistent. Although this is not a direct comparison between two independent sources of data, the overlapping data available between the balloon and the satellite are consistent, which suggests that, given more balloon-satellite conjunctions and measurements, we could show experimentally that the balloon data are sufficient to quantify the precipitating spectrum.

6.3 Future Work

6.3.1 Simulations and Models

To be useful to balloon experiments, the analysis in chapters 3 and 4 needs to be expanded to include different detector designs and specifications. In this thesis, we have demonstrated the effectiveness of one particular detector design, in one experimental setting. Next generation solid state and plastic scintillators have different properties than the sodium iodide crystals used in this work, and their analysis will depend on an accurate representation and simulation. The question of whether a different detector design could be more effective at providing data that solves the inverse problem is so far, unanswered. If differently designed detectors have more complex responses to X-ray photons, which require a different mathematical representation, then much of the work done in Chapter 4 will need to change to account for this. It is not known whether the preconditioning and regularization techniques will still provide useful results given a different detector model.

The analysis techniques developed in Chapter 4 assume an even sampling of energy spectra, using constant-width rectangular bins. This is likely not optimal, and it may be beneficial to increase the effective bin width, and lower the resolution of represented spectra, with increasing energy. There is no reason that this alteration of energy resolution needs to be linear. Accounting for the fact that the lower recorded counts at high energies should have a lower encoded resolution is standard in detector design, and it should be explored for potential benefit in the analysis as well.

The main piece of knowledge that the analysis we have developed does not provide is an estimate on the error in the developed solution. This is a current topic of research in applied statistics, and there are no simple ways to add this function to the analysis method. There are theoretical methods to provide estimates of the spread of confidence intervals over subsequent measurements of the same spectrum, but this is information that is generally not available experimentally. Once a precipitation event occurs, recording the same event multiple times isn't possible. Recent work on the generation of meaningful confidence intervals for biased estimators is ongoing. It may be that a practical method is found, and can then applied to the X-ray inversion problem.

One actionable method which could be applied today, is the computation through the forward model of many different prototype functions, and their subsequent inversion under specific assumptions on the noise and bias of the instrument in use. Given a rough form for the retrieved spectrum, using the techniques in Chapter 4, the behaviour of changes to counting statistics on the input data could be roughly quantified through monte-carlo simulations. This will depend on specific knowledge of the detector being used for a given experiment. There are no general results at present.

The simplification obtained by ignoring the angular spectrum of the precipitating electrons also results in a loss of information that can be obtained from the data. Future detectors which have angular resolution could, in principle, be used to obtain data on the location, size, and shape of patches of precipitating electrons. This analysis would require moving from a 1 dimensional representation of the inverse problem to a 2 dimensional model. Based on the computation required to generate the forward model in Chapter 3, which was done for one assumed angular distribution, the required computational resources will be large, and possibly impractical without a significant reparameterization of the model. Based on the current analysis, three dimensional histograms would need to be generated for every possible input. Hundreds of hours of computer time on significantly parallel machines was required to form the model used in this thesis, so such an approach would likely not be successful. What may work is the refactoring of the forward model into a set of different basis functions, such as spherical harmonics, or perhaps others, which are orthogonal, or sufficiently orthogonal to not make the inverse problem impossible. The design of detectors which can resolve angular distributions in incoming X-rays from balloons will need to take into account the ill-conditioning of the one dimensional problem. Their output will need to be simulated, and tests on synthetic data carried out, to see how much information can be obtained from them.

6.3.2 Experimental Data

Despite successful campaigns of zero-pressure balloons flown over months, and multiple flights of opportunity taken when satellite conjunctions are available and space weather is favorable to precipitation events, there is still a great need for more data which can be compared between balloons and spacecraft. Tests based on synthetic data alone are not sufficient to fully validate the method. For this to happen, we need pitch-angle resolved measurements of precipitating electrons in space, which are magnetically conjugate to the balloon measurements. If an expected precipitating spectrum could be determined, free from strong assumptions on the scattering processes responsible, then a more direct comparison with the balloon data would be possible. The techniques developed in this thesis also permit the analysis of X-ray data for comparatively low intensity precipitation events when the counting statistics available in the recorded data are not very pure. Given the large dataset of the BARREL balloon flights, there are certainly events available which can be analysed, that have not yet.

The next generation of instrumented balloons may greatly increase the depth of the available data. Work done at the University of Calgary to develop rapidly deployable, low cost, long duration balloons has shown promising initial results. When combined with lightweight solid-state detectors, and flights which permit recovery operations, it is hoped that flights of opportunity can be accomplished as active space weather conditions occur. The reduced cost and risk of this type of flight may make it an attractive platform on which to run X-ray measurement experiments.

Bibliography

- S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. J. G. Cadenas, I. Gonzlez, G. G. Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampn, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. M. de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. D. Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. S. Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wanner, H. Weber, J. P. Wellisch, T. Wenaus, D. C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche. Geant4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506: 250–303, 2003. ISSN 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- J. M. Albert and J. Bortnik. Nonlinear interaction of radiation belt electrons with electromagnetic ion cyclotron waves. *Geophysical Research Letters*, 36, 6 2009. ISSN 0094-8276. doi: <https://doi.org/10.1029/2009GL038904>. URL <https://doi.org/10.1029/2009GL038904>.

<https://doi.org/10.1029/2009GL038904>.

J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, G. A. P. Cirrone, G. Cooperman, G. Cosmo, G. Cuttone, G. G. Daquino, M. Donszelmann, M. Dressel, G. Folger, F. Foppiano, J. Generowicz, V. Grichine, S. Guatelli, P. Gumplinger, A. Heikkinen, I. Hrivnacova, A. Howard, S. Incerti, V. Ivanchenko, T. Johnson, F. Jones, T. Koi, R. Kokoulin, M. Kossov, H. Kurashige, V. Lara, S. Larsson, F. Lei, O. Link, F. Longo, M. Maire, A. Mantero, B. Mascialino, I. McLaren, P. M. Lorenzo, K. Minamimoto, K. Murakami, P. Nieminen, L. Pandola, S. Parlati, L. Peralta, J. Perl, A. Pfeiffer, M. G. Pia, A. Ribon, P. Rodrigues, G. Russo, S. Sadilov, G. Santin, T. Sasaki, D. Smith, N. Starkov, S. Tanaka, E. Tcherniaev, B. Tome, A. Trindade, P. Truscott, L. Urban, M. Verderi, A. Walkden, J. P. Wellisch, D. C. Williams, D. Wright, and H. Yoshida. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53:270–278, 2006. ISSN 1558-1578 VO - 53. doi: 10.1109/TNS.2006.869826.

J. Allison, K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso, E. Bagli, A. Bagulya, S. Banerjee, G. Barrand, B. R. Beck, A. G. Bogdanov, D. Brandt, J. M. C. Brown, H. Burkhardt, P. Canal, D. Cano-Ott, S. Chauvie, K. Cho, G. A. P. Cirrone, G. Cooperman, M. A. Corts-Giraldo, G. Cosmo, G. Cuttone, G. Depaola, L. Desorgher, X. Dong, A. Dotti, V. D. Elvira, G. Folger, Z. Francis, A. Galoyan, L. Garnier, M. Gayer, K. L. Genser, V. M. Grichine, S. Guatelli, P. Guye, P. Gumplinger, A. S. Howard, I. Hivnov, S. Hwang, S. Incerti, A. Ivanchenko, V. N. Ivanchenko, F. W. Jones, S. Y. Jun, P. Kaitaniemi, N. Karakatsanis, M. Karamitros, M. Kelsey, A. Kimura, T. Koi, H. Kurashige, A. Lechner, S. B. Lee, F. Longo, M. Maire, D. Mancusi, A. Mantero, E. Mendoza, B. Morgan, K. Murakami, T. Nikitina, L. Pandola, P. Paprocki, J. Perl, I. Petrovi, M. G. Pia, W. Pokorski, J. M. Quesada, M. Raine, M. A. Reis, A. Ribon, A. R. Fira, F. Romano, G. Russo, G. Santin, T. Sasaki, D. Sawkey, J. I. Shin, I. I. Strakovsky, A. Taborda, S. Tanaka, B. Tom, T. Toshito, H. N. Tran, P. R. Truscott, L. Urban, V. Uzhinsky, J. M. Verbeke, M. Verderi, B. L. Wendt, H. Wenzel, D. H. Wright, D. M. Wright, T. Yamashita,

- J. Yarba, and H. Yoshida. Recent developments in geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225, 2016. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2016.06.125>. URL <https://www.sciencedirect.com/science/article/pii/S0168900216306957>.
- K. A. Anderson and D. C. Enemark. Balloon observations of x rays in the auroral zone ii. *Journal of Geophysical Research (1896-1977)*, 65:3521–3538, 11 1960. ISSN 0148-0227. doi: <https://doi.org/10.1029/JZ065i011p03521>. URL <https://doi.org/10.1029/JZ065i011p03521>.
- V. Angelopoulos, P. Cruce, A. Drozdov, E. W. Grimes, N. Hatzigeorgiu, D. A. King, D. Larson, J. W. Lewis, J. M. McTiernan, D. A. Roberts, C. L. Russell, T. Hori, Y. Kasahara, A. Ku-mamoto, A. Matsuoka, Y. Miyashita, Y. Miyoshi, I. Shinohara, M. Teramoto, J. B. Faden, A. J. Halford, M. McCarthy, R. M. Millan, J. G. Sample, D. M. Smith, L. A. Woodger, A. Mas-son, A. A. Narock, K. Asamura, T. F. Chang, C. Y. Chiang, Y. Kazama, K. Keika, S. Matsuda, T. Segawa, K. Seki, M. Shoji, S. W. Tam, N. Umemura, B. J. Wang, S. Y. Wang, R. Red-mon, J. V. Rodriguez, H. J. Singer, J. Vandegriff, S. Abe, M. Nose, A. Shinbori, Y. M. Tanaka, S. UeNo, L. Andersson, P. Dunn, C. Fowler, J. S. Halekas, T. Hara, Y. Harada, C. O. Lee, R. Lil-lis, D. L. Mitchell, M. R. Argall, K. Bromund, J. L. Burch, I. J. Cohen, M. Galloy, B. Giles, A. N. Jaynes, O. L. Contel, M. Oka, T. D. Phan, B. M. Walsh, J. Westlake, F. D. Wilder, S. D. Bale, R. Livi, M. Pulupa, P. Whittlesey, A. DeWolfe, B. Harter, E. Lucas, U. Auster, J. W. Bonnell, C. M. Cully, E. Donovan, R. E. Ergun, H. U. Frey, B. Jackel, A. Keiling, H. Korth, J. P. McFadden, Y. Nishimura, F. Plaschke, P. Robert, D. L. Turner, J. M. Weygand, R. M. Candey, R. C. Johnson, T. Kovalick, M. H. Liu, R. E. McGuire, A. Breneman, K. Kersten, and P. Schroeder. *The Space Physics Environment Data Analysis System (SPEDAS)*, volume 215. The Author(s), 2019. ISBN 1121401805. doi: 10.1007/s11214-018-0576-4. URL <http://dx.doi.org/10.1007/s11214-018-0576-4>.
- J. R. Barcus and T. J. Rosenberg. Energy spectrum for auroral-zone x rays: 1. diurnal and type

- effects. *Journal of Geophysical Research (1896-1977)*, 71:803–823, 2 1966. ISSN 0148-0227. doi: <https://doi.org/10.1029/JZ071i003p00803>. URL <https://doi.org/10.1029/JZ071i003p00803>. <https://doi.org/10.1029/JZ071i003p00803>.
- R. J. Barlow. Statistics : a guide to the use of statistical methods in the physical sciences., 2013. URL <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1101495>.
- M. J. Berger and S. M. Seltzer. Bremsstrahlung in the atmosphere. *Journal of Atmospheric and Terrestrial Physics*, 34:85–108, 1972. ISSN 00219169. doi: 10.1016/0021-9169(72)90006-2.
- A. Bjerhammar. Rectangular reciprocal matrices, with special reference to geodetic calculations. *Bulletin Godsique (1946-1975)*, 20:188–220, 1951. ISSN 0007-4632. doi: 10.1007/BF02526278. URL <https://doi.org/10.1007/BF02526278>.
- J. B. Blake, P. A. Carranza, S. G. Claudepierre, J. H. Clemons, W. R. Crain, Y. Dotan, J. F. Fennell, F. H. Fuentes, R. M. Galvan, J. S. George, M. G. Henderson, M. Lalic, A. Y. Lin, M. D. Looper, D. J. Mabry, J. E. Mazur, B. McCarthy, C. Q. Nguyen, T. P. OBrien, M. A. Perez, M. T. Redding, J. L. Roeder, D. J. Salvaggio, G. A. Sorensen, H. E. Spence, S. Yi, and M. P. Zdziarski. The magnetic electron ion spectrometer (magentis) instruments aboard the radiation belt storm probes (rbsp) spacecraft. *Space Science Reviews*, 179:383–421, 2013. ISSN 1572-9672. doi: 10.1007/s11214-013-9991-8. URL <https://doi.org/10.1007/s11214-013-9991-8>.
- J. Bortnik, R. M. Thorne, and N. P. Meredith. The unexpected origin of plasmaspheric hiss from discrete chorus emissions. *Nature*, 452:62–66, 2008. ISSN 1476-4687. doi: 10.1038/nature06741. URL <https://doi.org/10.1038/nature06741>.
- H. Bouwmeester, A. Dougherty, and A. V. Knyazev. Nonsymmetric preconditioning for conjugate gradient and steepest descent methods1. *Procedia Computer Science*, 51:276–285, 2015. ISSN 1877-0509. doi: <https://doi.org/10.1016/j.procs.2015.05.241>. URL <https://www.sciencedirect.com/science/article/pii/S1877050915010492>.

- J. C. Brown, A. G. Emslie, G. D. Holman, C. M. JohnsKrull, E. P. Kontar, R. P. Lin, A. M. Massone, and M. Piana. Evaluation of algorithms for reconstructing electron spectra from their bremsstrahlung hard xray spectra. *The Astrophysical Journal*, 643:523–531, 2006. ISSN 0004-637X. doi: 10.1086/501497.
- R. R. Brown. Features of the auroral electron energy spectrum inferred from observations of ionospheric absorption. 1965. URL <https://www.osti.gov/biblio/4601621><https://www.osti.gov/servlets/purl/4601621>.
- R. R. Brown. Electron precipitation in the auroral zone. *Space Science Reviews*, 5:311–387, 1966. ISSN 1572-9672. doi: 10.1007/BF02653249. URL <https://doi.org/10.1007/BF02653249>.
- H. Burkhardt, V. Grichine, P. Gumplinger, V. N. Ivanchenko, R. Kokoulin, M. Maire, and L. Urban. *Geant4 standard electromagnetic package for HEP applications*, volume 3. 11 2004. ISBN 0-7803-8700-7. doi: 10.1109/NSSMIC.2004.1462617.
- C. Cattell, J. R. Wygant, K. Goetz, K. Kersten, P. J. Kellogg, T. von Rosenvinge, S. D. Bale, I. Roth, M. Temerin, M. K. Hudson, R. A. Mewaldt, M. Wiedenbeck, M. Maksimovic, R. Ergun, M. Acuna, and C. T. Russell. Discovery of very large amplitude whistler-mode waves in earth’s radiation belts. *Geophysical Research Letters*, 35, 1 2008. ISSN 0094-8276. doi: <https://doi.org/10.1029/2007GL032009>. URL <https://doi.org/10.1029/2007GL032009><https://doi.org/10.1029/2007GL032009>.
- E. W. Cheney and D. D. R. Kincaid. *Numerical mathematics and computing / Ward Cheney, David Kincaid*. Brooks/Cole, sixth edit edition, 2008. ISBN 9780495114758. Includes bibliographical references (pages 745-753) and index.
- A. B. Christensen and R. Karas. Energy spectra of precipitating electrons from observations of optical aurora, bremsstrahlung x rays, and auroral absorption. *Journal of Geophysical Research (1896-1977)*, 75:4266–4278, 8 1970. ISSN 0148-0227. doi: <https://doi.org/10.1029/JA075i008p04266>

//doi.org/10.1029/JA075i022p04266. URL <https://doi.org/10.1029/JA075i022p04266>.
<https://doi.org/10.1029/JA075i022p04266>.

S. R. Church and R. M. Thorne. On the origin of plasmaspheric hiss: Ray path integrated amplification. *Journal of Geophysical Research: Space Physics*, 88:7941–7957, 10 1983. ISSN 0148-0227. doi: <https://doi.org/10.1029/JA088iA10p07941>. URL <https://doi.org/10.1029/JA088iA10p07941>. <https://doi.org/10.1029/JA088iA10p07941>.

A. Dresden. The fourteenth western meeting of the american mathematical society. *Bulletin of the American Mathematical Society*, 26:385–396, 6 1920. doi: bams/1183425340. URL <https://doi.org/>.

J. E. Foat, R. P. Lin, D. M. Smith, F. Fenrich, R. Millan, I. Roth, K. R. Lorentzen, M. P. McCarthy, G. K. Parks, and J. P. Treilhou. First detection of a terrestrial mev x-ray burst. *Geophysical Research Letters*, 25:4109–4112, 1998. ISSN 00948276. doi: 10.1029/1998GL900134.

B. J. Fraser, H. J. Singer, W. J. Hughes, J. R. Wygant, R. R. Anderson, and Y. D. Hu. Cres poynting vector observations of electromagnetic ion cyclotron waves near the plasma-pause. *Journal of Geophysical Research: Space Physics*, 101:15331–15343, 7 1996. ISSN 0148-0227. doi: <https://doi.org/10.1029/95JA03480>. URL <https://doi.org/10.1029/95JA03480>. <https://doi.org/10.1029/95JA03480>.

G. P. Ginnet, T. P. OBrien, S. L. Huston, W. R. Johnston, T. B. Guild, R. Friedel, C. D. Lindstrom, C. J. Roth, P. Whelan, R. A. Quinn, D. Madden, S. Morley, and Y.-J. Su. Ae9, ap9 and spm: New models for specifying the trapped energetic particle and space plasma environment. *Space Science Reviews*, 179:579–615, 2013. ISSN 1572-9672. doi: 10.1007/s11214-013-9964-y. URL <https://doi.org/10.1007/s11214-013-9964-y>.

J. Goeman, R. Meijer, and N. Chaturvedi. L1 and l2 penalized regression models. *Cran. R-Project. or*, pages 1–20, 2012. URL <ftp://ftp.yz.yamagata-u.ac.jp/mirror7/cran/web/packages/penalized/vignettes/penalized.pdf>.

- G. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 5 1979. doi: 10.1080/00401706.1979.10489751.

S. Guatelli, A. Mantero, B. Mascialino, P. Nieminen, M. G. Pia, and S. Saliceti. Geant4 atomic relaxation. volume 4, pages 2178–2181 Vol. 4, 2004. ISBN 1082-3654 VO - 4. doi: 10.1109/NSSMIC.2004.1462694.

J. Hadamard. *Sur les problems aux derivees partielles et leur signification physique*, volume 13. 1902. URL <http://ci.nii.ac.jp/naid/10030001312/en/>.

A. J. Halford, S. L. McGregor, K. R. Murphy, R. M. Millan, M. K. Hudson, L. A. Woodger, C. A. Cattel, A. W. Breneman, I. R. Mann, W. S. Kurth, G. B. Hospodarsky, M. Gkioulidou, and J. F. Fennell. Barrel observations of an icme-shock impact with the magnetosphere and the resultant radiation belt electron loss. *Journal of Geophysical Research: Space Physics*, 120:2557–2570, 2015. ISSN 21699402. doi: 10.1002/2014JA020873.

S. R. Hanha. Air quality model evaluation and uncertainty. *JAPCA*, 38:406–412, 4 1988. ISSN 0894-0630. doi: 10.1080/08940630.1988.10466390. URL <https://doi.org/10.1080/08940630.1988.10466390>.

M. Hanke. Limitations of the l-curve method in ill-posed problems. *BIT Numerical Mathematics*, 36:287–301, 1996. doi: 10.1007/BF01731984. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-1542498779&doi=10.1007%2FBF01731984&partnerID=40&md5=7af81269745eb6e61a8dc1efcbe4b3ec>. Cited By :144;br/Export Date: 22 September 2021.

P. C. Hansen. No title. *SIAM Rev.*, pages 561–580, 1992. Cited By :1;br/Export Date: 22 September 2021.

P. C. Hansen and D. P. O’Leary. The use of the l-curve in the regularization of discrete ill-posed

problems. *SIAM J. Sci. Comput.*, 14:1487–1503, 1993. Cited By :68;br/;br/;Export Date: 22 September 2021.

P. C. Hansen, J. G. Nagy, and D. P. O’Leary. *Deblurring Images*. 2006. ISBN 9780898716184. doi: 10.1137/1.9780898718874.

C. Huang and C. Goertz. Ray-tracing studies and path-integrated gains of elf unducted whistler mode waves in the earths magnetosphere. *Journal of Geophysical Research*, 88:6181–6187, 8 1983. doi: 10.1029/JA088iA08p06181.

V. K. Jordanova, C. J. Farrugia, R. M. Thorne, G. V. Khazanov, G. D. Reeves, and M. F. Thomsen. Modeling ring current proton precipitation by electromagnetic ion cyclotron waves during the may 1416, 1997, storm. *Journal of Geophysical Research: Space Physics*, 106:7–22, 1 2001. ISSN 0148-0227. doi: <https://doi.org/10.1029/2000JA002008>. URL <https://doi.org/10.1029/2000JA002008>. <https://doi.org/10.1029/2000JA002008>.

kamiyaha h. Ionization and excitation by precipitating electrons. *Rep. Ion. Space Res. Jpn.*, 20: 171–187, 1966. URL <http://ci.nii.ac.jp/naid/10030112507/en/>.

S. Kasahara, Y. Miyoshi, S. Yokota, T. Mitani, Y. Kasahara, S. Matsuda, A. Kumamoto, A. Mat-suoka, Y. Kazama, H. U. Frey, V. Angelopoulos, S. Kurita, K. Keika, K. Seki, and I. Shinohara. Pulsating aurora from electron scattering by chorus waves. *Nature*, 554:337–340, 2018. ISSN 14764687. doi: 10.1038/nature25505.

C. F. Kennel and H. E. Petschek. Limit on stably trapped particle fluxes. *Journal of Geophysical Research (1896-1977)*, 71:1–28, 1 1966. ISSN 0148-0227. doi: <https://doi.org/10.1029/JZ071i001p00001>. URL <https://doi.org/10.1029/JZ071i001p00001>. <https://doi.org/10.1029/JZ071i001p00001>.

K. R. Lorentzen, M. P. McCarthy, G. K. Parks, J. E. Foat, R. M. Millan, D. M. Smith, R. P. Lin, and J. P. Treilhou. Precipitation of relativistic electrons by interaction

with electromagnetic ion cyclotron waves. *Journal of Geophysical Research: Space Physics*, 105:5381–5389, 3 2000. ISSN 2156-2202. doi: 10.1029/1999JA000283. URL <https://onlinelibrary.wiley.com/doi/full/10.1029/1999JA000283><https://onlinelibrary.wiley.com/doi/abs/10.1029/1999JA000283><https://agupubs.onlinelibrary.wiley.com/doi/10.1029/1999JA000283>.

K. R. Lorentzen, J. B. Blake, U. S. Inan, and J. Bortnik. Observations of relativistic electron microbursts in association with vlf chorus. *Journal of Geophysical Research: Space Physics*, 106:6017–6027, 2001. doi: 10.1029/2000ja003018. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77954910240&doi=10.1029%2f2000ja003018&partnerID=40&md5=a53d00937b1b40e3033155f4af2c397f>.

Cited By :190 Export Date: 22 November 2021.

T. M. Loto’aniu, B. J. Fraser, and C. L. Waters. Propagation of electromagnetic ion cyclotron wave energy in the magnetosphere. *Journal of Geophysical Research: Space Physics*, 110, 7 2005. ISSN 0148-0227. doi: <https://doi.org/10.1029/2004JA010816>. URL <https://doi.org/10.1029/2004JA010816>. <https://doi.org/10.1029/2004JA010816>.

L. R. Lyons and R. M. Thorne. Equilibrium structure of radiation belt electrons. *Journal of Geophysical Research (1896-1977)*, 78:2142–2149, 5 1973. ISSN 0148-0227. doi: <https://doi.org/10.1029/JA078i013p02142>. URL <https://doi.org/10.1029/JA078i013p02142>. <https://doi.org/10.1029/JA078i013p02142>.

L. R. Lyons, R. M. Thorne, and C. F. Kennel. Pitch-angle diffusion of radiation belt electrons within the plasmasphere. *Journal of Geophysical Research (1896-1977)*, 77:3455–3474, 7 1972. ISSN 0148-0227. doi: <https://doi.org/10.1029/JA077i019p03455>. URL <https://doi.org/10.1029/JA077i019p03455>. <https://doi.org/10.1029/JA077i019p03455>.

N. P. Meredith, R. M. Thorne, R. B. Horne, D. Summers, B. J. Fraser, and R. R. Anderson. Statistical analysis of relativistic electron energies for cyclotron resonance with emic waves observed

on crres. *Journal of Geophysical Research: Space Physics*, 108, 6 2003. ISSN 0148-0227. doi: <https://doi.org/10.1029/2002JA009700>. URL <https://doi.org/10.1029/2002JA009700>.

N. P. Meredith, R. B. Horne, R. M. Thorne, D. Summers, and R. R. Anderson. Substorm dependence of plasmaspheric hiss. *Journal of Geophysical Research: Space Physics*, 109, 2004. doi: 10.1029/2004JA010387. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-41849107567&doi=10.1029%2f2004JA010387&partnerID=40&md5=5d9a6bdd59487c7ace986a8f961e7915>. Cited By :260 Export Date: 21 November 2021.

N. P. Meredith, R. B. Horne, T. Kersten, W. Li, J. Bortnik, A. Sicard, and K. H. Yearby. Global model of plasmaspheric hiss from multiple satellite observations. *Journal of Geophysical Research: Space Physics*, 123:4526–4541, 6 2018. ISSN 2169-9380. doi: <https://doi.org/10.1029/2018JA025226>. URL <https://doi.org/10.1029/2018JA025226>.

R. M. Millan and D. N. Baker. Acceleration of particles to high energies in earths radiation belts. *Space Science Reviews*, 173:103–131, 2012. ISSN 1572-9672. doi: 10.1007/s11214-012-9941-x. URL <https://doi.org/10.1007/s11214-012-9941-x>.

R. M. Millan, R. P. Lin, D. M. Smith, and M. P. McCarthy. Observation of relativistic electron precipitation during a rapid decrease of trapped relativistic electron flux. *Geophysical Research Letters*, 34:1–5, 2007. ISSN 00948276. doi: 10.1029/2006GL028653.

R. M. Millan, M. P. McCarthy, J. G. Sample, D. M. Smith, L. D. Thompson, D. G. McGaw, L. A. Woodger, J. G. Hewitt, M. D. Comess, K. B. Yando, A. X. Liang, B. A. Anderson, N. R. Knezeck, W. Z. Rexroad, J. M. Scheiman, G. S. Bowers, A. J. Halford, A. B. Collier, M. A. Clilverd, R. P. Lin, M. K. Hudson, and L. D. Thompson. The balloon array for rbsp relativistic electron losses (barrel) abstract barrel is a multiple-balloon investigation designed to study electron losses from

- earth's radiation belts. selected as a nasa living with a star mission of opportu. *Space Sci Rev*, 179:503–530, 2013. ISSN 00386308. doi: 10.1007/978-1-4899-7433-4-15.
- K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50:157–175, 7 1900. ISSN 1941-5982. doi: 10.1080/14786440009463897. URL <https://doi.org/10.1080/14786440009463897>. doi: 10.1080/14786440009463897.
- A. Pfeier, G. Cosmo, B. Ferrero-Merlino, R. Giannitrapani, F. Longo, P. Nieminen, M. G. Pia, and G. Santin. Architecture of collaborating frameworks: simulation, visualisation, user interface and analysis, 9 2001. URL <https://cds.cern.ch/record/519005>.
- J. M. Picone, A. E. Hedin, D. P. Drob, and A. C. Aikin. Nrlmsise-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research: Space Physics*, 107:SIA 15–1–SIA 15–16, 12 2002. ISSN 21699402. doi: 10.1029/2002JA009430. URL <https://doi.org/10.1029/2002JA009430>. doi: 10.1029/2002JA009430.
- G. Pini and G. Gambolati. Is a simple diagonal scaling the best preconditioner for conjugate gradients on supercomputers? *Advances in Water Resources*, 13:147–153, 1990. ISSN 0309-1708. doi: [https://doi.org/10.1016/0309-1708\(90\)90006-P](https://doi.org/10.1016/0309-1708(90)90006-P). URL <https://www.sciencedirect.com/science/article/pii/030917089090006P>.
- C. Polk. Electron density distribution in ionosphere and exosphere. *Proceedings of the IEEE*, 53: 334–335, 1965. ISSN 1558-2256 VO - 53. doi: 10.1109/PROC.1965.3737.
- M. H. Rees. Auroral ionization and excitation by incident energetic electrons. *Planetary and Space Science*, 11:1209–1218, 1963. ISSN 00320633. doi: 10.1016/0032-0633(63)90252-6.
- J. R. Rice. A theory of condition. *SIAM Journal on Numerical Analysis*, 3:287–310, 6 1966.

ISSN 0036-1429. doi: 10.1137/0703023. URL <https://doi.org/10.1137/0703023>. doi: 10.1137/0703023.

S. F. Singer. "radiation belt" and trapped cosmic-ray albedo. *Physical Review Letters*, 1:171–173, 9 1958. doi: 10.1103/PhysRevLett.1.171. URL <https://link.aps.org/doi/10.1103/PhysRevLett.1.171>.

H. E. Spence, G. D. Reeves, D. N. Baker, J. B. Blake, M. Bolton, S. Bourdarie, A. A. Chan, S. G. Claudepierre, J. H. Clemmons, J. P. Cravens, S. R. Elkington, J. F. Fennell, R. H. W. Friedel, H. O. Funsten, J. Goldstein, J. C. Green, A. Guthrie, M. G. Henderson, R. B. Horne, M. K. Hudson, J.-M. Jahn, V. K. Jordanova, S. G. Kanekal, B. W. Klatt, B. A. Larsen, X. Li, E. A. MacDonald, I. R. Mann, J. Niehof, T. P. OBrien, T. G. Onsager, D. Salvaggio, R. M. Skoug, S. S. Smith, L. L. Suther, M. F. Thomsen, and R. M. Thorne. Science goals and overview of the radiation belt storm probes (rbsp) energetic particle, composition, and thermal plasma (ect) suite on nasas van allen probes mission. *Space Science Reviews*, 179:311–336, 2013. ISSN 1572-9672. doi: 10.1007/s11214-013-0007-5. URL <https://doi.org/10.1007/s11214-013-0007-5>.

H. E. Spence, G. D. Reeves, D. N. Baker, J. B. Blake, M. Bolton, S. Bourdarie, A. A. Chan, S. G. Claudepierre, J. H. Clemmons, J. P. Cravens, S. R. Elkington, J. F. Fennell, R. H. W. Friedel, H. O. Funsten, J. Goldstein, J. C. Green, A. Guthrie, M. G. Henderson, R. B. Horne, M. K. Hudson, J.-M. Jahn, V. K. Jordanova, S. G. Kanekal, B. W. Klatt, B. A. Larsen, X. Li, E. A. MacDonald, I. R. Mann, J. Niehof, T. P. OBrien, T. G. Onsager, D. Salvaggio, R. M. Skoug, S. S. Smith, L. L. Suther, M. F. Thomsen, and R. M. Thorne. Science goals and overview of the radiation belt storm probes (rbsp) energetic particle, composition, and thermal plasma (ect) suite on nasas van allen probes mission bt - the van allen probes mission, 2014. URL https://doi.org/10.1007/978-1-4899-7433-4_10.

A. Tarantola. *Inverse problem theory*. 2004. ISBN 0898715725. URL http://www.ipgp.fr/~tarantola/Inverse_Problems/.

<fr/~tarantola/Files/Professional/SIAM/InverseProblemTheory.pdf%0Apapers2://publication/uuid/62A7A164-F0E3-4298-AA2B-222FF673BF4E>.

R. M. Thorne, S. R. Church, and D. J. Gorney. On the origin of plasmaspheric hiss: The importance of wave propagation and the plasmapause. *Journal of Geophysical Research*, 84: 5241–5248, 9 1979. ISSN 0148-0227. doi: 10.1029/JA084iA09p05241. URL <https://ui.adsabs.harvard.edu/abs/1979JGR....84.5241T>.

R. M. Thorne, L. J. Andreoli, and L. J. Andreoli. Mechanisms for intense relativistic electron precipitation, 1981.

R. M. Thorne, T. P. O'Brien, Y. Y. Shprits, D. Summers, and R. B. Horne. Timescale for mev electron microburst loss during geomagnetic storms. *Journal of Geophysical Research: Space Physics*, 110, 9 2005. ISSN 0148-0227. doi: <https://doi.org/10.1029/2004JA010882>. URL <https://doi.org/10.1029/2004JA010882>. <https://doi.org/10.1029/2004JA010882>.

R. M. Thorne, B. Ni, X. Tao, R. B. Horne, and N. P. Meredith. Scattering by chorus waves as the dominant cause of diffuse auroral precipitation. *Nature*, 467:943–946, 2010. ISSN 00280836. doi: 10.1038/nature09467.

A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola. Numerical methods for the solution of ill-posed problems. 1995. doi: 10.1007/978-94-015-8480-7.

B. T. Tsurutani and E. J. Smith. Postmidnight chorus: A substorm phenomenon. *Journal of Geophysical Research (1896-1977)*, 79:118–127, 1 1974. ISSN 0148-0227. doi: <https://doi.org/10.1029/JA079i001p00118>. URL <https://doi.org/10.1029/JA079i001p00118>. <https://doi.org/10.1029/JA079i001p00118>.

N. A. Tsyganenko. Global quantitative models of the geomagnetic field in the cislunar magnetosphere for different disturbance levels, 1987.

C. R. Vogel. Non-convergence of the l-curve regularization parameter selection method. *Inverse Problems*, 12:535–547, 1996. doi: 10.1088/0266-5611/12/4/013. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0001041472&doi=10.1088%2F0266-5611%2F12%2F4%2F013&partnerID=40&md5=aa5ea097016cd616faa2f9a2fee64c74>. Cited By :191
Export Date: 22 September 2021.

J. R. Winckler, L. Peterson, R. Arnoldy, and R. Hoffman. X-rays from visible aurorae at minneapolis. *Phys. Rev.*, 110:1221–1231, 6 1958. doi: 10.1103/PhysRev.110.1221. URL <https://link.aps.org/doi/10.1103/PhysRev.110.1221>.

W. Xu and R. A. Marshall. Characteristics of energetic electron precipitation estimated from simulated bremsstrahlung x-ray distributions. *Journal of Geophysical Research: Space Physics*, 124:2831–2843, 2019. ISSN 21699402. doi: 10.1029/2018JA026273.