

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXII

Advances in Mixture Modelling for Model-Based Clustering: Two Case Studies in Astronomy

Coordinatore del Corso: Prof. Massimiliano Caporin

Supervisore: Dott. Mauro Bernardi

Co-supervisore: Prof. Alessandra R. Brazzale, Prof. David A. van Dyk,
Prof. Roberto Trotta

Dottorando: Andrea Sottosanti

2nd December 2019

*to Mum and Dad, for the unceasing support
to Luana, for the patient love
to the memory of my beloved Grandparents*

*and above all
to the glory of God*

Acknowledgements

For all I care, I wanted to get the PhD since I was a Bachelor student. Now this widely desired goal is really, surprisingly very close, and it is not easy to summarise all the feelings I have and the things I want to say. Anyway, I will do my best.

First of all, I thank my supervisors in Padova, Mauro and Alessandra, for guiding me into the PhD journey with patience and precious advices. In the same way, I want to thank my supervisors in London, David and Roberto, for giving me the opportunity to work with them and for transmitting me their precious knowledge and experience.

I thank Prof. Sylvia Frühwirth-Schnatter and Prof. Michele Guidani for reviewing in detail this thesis and for the most helpful comments and suggestions they gave.

Thanks also to David Stenning and especially to Alex Geringer-Sameth for the most helpful discussions and for the patience with me and with my “astro doubts”.

Thanks to Luana, my love, my friend. Every time I had to take a choice you were there, ready to support me, already aware of what would make me happy. This little, big success is as such mine as it is yours.

A huge thank goes to my parents, my sustain and my constant encouragement, and to my entire family. Thanks to Giuseppe, who is like a grandfather to me in Padova. During this three-years journey I realised that, while you’re running the race of the life, you might turn and see that someone is not following you anymore. This thesis is to the memory of my two wonderful and beloved grandparents Marco and Michelina. They unexpectedly left me, but they surely would have rejoiced so much with me today.

Thanks to the mates of the XXXII cycle: Ale, Ana, Fede, Huiting, Ila, Meme, Mohammed and Moin. It has been a pleasure to spend the last three years with you, growing together not only from a professional point of view, but also from a human perspective.

Thanks to the Assembly of God church in Padova, my home and my shield, and to the friends, the true and authentic ones, so precious and always ready to support me. Last, but not least, thanks to the people in London: the old friends (Marco and Marta), the new ones at the Imperial College, and the people from the Kingdom Faith Church, who so kindly welcomed me as member of the family for my entire staying.

“To God only wise, be glory through Jesus Christ for ever. Amen.”

(Epistle to the Romans 16:27)

Abstract

Efficient classification techniques are nowadays in huge demand in observational astronomy, which is probably the oldest area of application of statistical science to the study of nature. The rapid evolution through technology has determined a paradigm shift where rich and/or massive data sets are becoming the dominant source of information. A typical sky survey collects terabytes of data each night on millions of possible objects (stars, galaxies, etc.) which themselves may encompass many observed properties. Some success stories are the Sloan Digital Sky Survey whose whelm of very high quality photometric data allowed astronomers to pinpoint the position of over one million of galaxies, and NASA's *Fermi* space telescope, thanks to which we are still deepening our knowledge of high-energy phenomena. Discoveries in these fields are of utmost relevance as they contain a wealth of information about the history of our Galaxy, and impact on the understanding of our own Solar System. However, the simple collection of these data sets is just the beginning of the process. A key step in astronomical breakthrough research is the meaningful analysis of the collected information. The complex structure of the data, combined with the availability of a tremendous number of observations, represents a non-negligible statistical challenge from both the theoretical and computational viewpoint. Statisticians must work at the forefront of this area.

This thesis aims to enlarge the set of clustering and classification techniques available for astronomy, providing new flexible solutions to cope with the requirements imposed by the vast diversity of celestial objects. In the first part, we propose an innovative approach based on Bayesian nonparametric methods to the signal extraction of high-energy astronomical sources immersed in a strong background contamination. Our model simultaneously clusters the photons and gives an estimate of the number of sources using a Dirichlet Process (DP) mixture, while separating them from the irregularly shaped background, that is reconstructed using a novel Bayesian nonparametric technique based

on B-spline functions. The resultant is then a hierarchical model in the class of mixtures of DP mixtures. We provide a suitable Markov Chain Monte Carlo algorithm to conduct the inference, and a post-processing procedure to quantify the information coming from the discovered clusters. We finally test the capacity of the model in locating and extracting the signal of the sources using several artificial datasets, and a further application on the *Fermi* LAT map is proposed.

In the second part, we propose a novel statistical approach to separate the different levels of brightness in the emission activity of high-energy sources. The method analyses the variation of flux in time and clusters the observations to infer on the latent states of variability that correspond to distinct physical mechanisms of the source. We model the transition among the latent states with a continuous-time Markov chain, and the flux measurements in each state with an Ornstein-Uhlenbeck (OU) process. The resultant technique belongs to the class of continuous-time hidden Markov models (HMMs) and can be fitted via maximum likelihood estimation using an efficient EM algorithm. In addition, we assess the properties of the model with a proper bootstrap algorithm. We finally illustrate the efficiency of the method on a light curve from a blazar discovered by the *Fermi* LAT.

Sommario

Negli ultimi anni, si è assistito ad un crescente interesse per lo sviluppo di avanzati strumenti di classificazione per l'astronomia osservativa, la più antica area di applicazione della statistica per lo studio dei fenomeni naturali. La rapida evoluzione degli strumenti tecnologici ha infatti portato ad un decisivo cambio di paradigma, dove imponenti moli di dati sono diventate le più frequenti fonti di informazione disponibili. Durante una tipica fase di esplorazione del cielo vengono registrati terabyte di dati relativi a milioni di possibili oggetti astronomici, tra cui stelle e galassie; tra le più annoverate esplorazioni, la Sloan Digital Sky Survey, che ha permesso agli astronomi di localizzare la posizione di oltre un milione di galassie, e il telescopio spaziale *Fermi* della NASA, grazie al quale la conoscenza dei fenomeni ad alta energia è in continua crescita. Scoperte di questa portata hanno una grandissima rilevanza nella comprensione non solo dell'origine e dell'evoluzione del nostro sistema solare, ma anche dell'intera galassia in cui ci troviamo. Tuttavia, la semplice raccolta di dati è soltanto l'inizio del processo, in quanto il vero punto di svolta nella ricerca astronomica è dato dall'analisi dell'informazione raccolta. Le ingenti e complesse moli di dati disponibili rappresentano infatti una vera e propria sfida per gli statistici, che sono chiamati a rispondere a queste richieste con sviluppi e innovazioni sia su aspetti teorici che computazionali della statistica.

Questa tesi mira a fornire quegli strumenti flessibili per la classificazione e il *clustering* dei dati che l'astronomia moderna richiede, facendo fronte all'enorme varietà di oggetti astronomici identificabili nell'universo. Nella prima parte, viene proposto un innovativo approccio Bayesiano nonparametrico per l'estrazione del segnale di sorgenti astronomiche ad alta energia immerse in un'irregolare contaminazione di fondo. Il modello proposto è in grado di raggruppare i fotoni osservati e fornire una stima del numero di sorgenti tramite una mistura di *Dirichlet Process*, separando il loro segnale dal rumore di fondo, che viene ricostruito attraverso un innovativo approccio Bayesiano nonparametrico basato sulle funzioni B-spline. Il modello gerarchico che ne risulta è una

mistura di misture di *Dirichlet Process*. Viene proposto inoltre un algoritmo di simulazione nella classe dei metodi Markov Chain Monte Carlo per condurre l'inferenza, e un algoritmo che quantifica l'informazione proveniente dai gruppi identificati sfruttando i valori campionati dalla distribuzione a posteriori. I risultati ottenuti su diversi insiemi di dati simulati confermano la capacità del modello proposto di localizzare e quantificare il segnale di sorgenti astronomiche. Viene infine proposta un'applicazione del metodo ad una mappa di fotoni proveniente dal *Fermi* LAT.

Nella seconda parte, viene proposto un innovativo strumento statistico per separare i diversi livelli di luminosità nell'attività di emissione di sorgenti astronomiche ad alta energia. Il metodo analizza la variazione di flusso nel tempo e raggruppa le osservazioni al fine di svolgere inferenza sul numero di stati latenti, che corrispondono a differenti meccanismi fisici della sorgente. La transizione fra i diversi stati latenti viene modellata da una catena di Markov a tempo continuo, e il flusso in ciascuno stato con un processo di Ornstein-Uhlenbeck (OU). Il modello che ne risulta appartiene alla classe di modelli *hidden Markov* e può essere stimato via massima verosimiglianza con un algoritmo EM; in aggiunta, è possibile risalire alle proprietà del modello attraverso un opportuno algoritmo *bootstrap*. Viene infine mostrata un'applicazione del metodo proposto sul segnale proveniente da una sorgente blazar osservata dal *Fermi* LAT.

Contents

List of Figures	xv
List of Tables	xviii
Introduction	3
Overview	3
Main contributions of the thesis	5
1 Source extraction and background separation: a Bayesian nonparametric approach	7
1.1 Introduction	7
1.1.1 Astrophysical source extraction	8
1.1.2 Main goals and outline	9
1.2 The statistical model	11
1.2.1 The source model	11
1.2.2 The background model	12
1.2.3 Including the spectral information	14
1.2.4 Misclassification effect	15
1.3 Model generalization and simulation algorithm	16
1.3.1 A mixture of Dirichlet Process mixtures	16
1.3.2 The MCMC algorithm	18
1.4 Numerical experiments	20
1.4.1 An illustrative example	21
1.4.2 Model comparison	26
1.5 Application to the <i>Fermi</i> LAT data	28
1.6 Discussion	30
2 Continuous time hidden Markov models for flare detection in gamma-ray light curves	33
2.1 Introduction	33
2.2 Hidden Markov modelling approach	35
2.2.1 Continuous time observations	35
2.2.2 Multiple states modelling	36
2.3 Model estimation	37
2.3.1 Parameters estimation via EM algorithm	37

2.3.2	Model assessment via bootstrap algorithm	40
2.4	Application to the <i>Fermi</i> LAT data	41
2.5	Discussion	47
3	Conclusions	49
	Appendix	53
	Bibliography	61

List of Figures

1.1	Photons emitted from an astronomical source located at $(0, 0)$ in different energy bins. Increasing the energy range (from left-top to right-top, and from left-bottom to right-bottom) the events become rarer and concentrate around the direction of the source.	11
1.2	Directed Acyclic Graph (DAG) which illustrates the dependency structure among the variables of the model. The solid lines represent Model (1.1), and both solid and dashed lines describe Model (1.7).	17
1.3	Map of 10,000 events obtained by the generating method described at the beginning of Section 1.4, using 9 equally bright sources and a moderate background contamination.	21
1.4	Left: posterior distribution of the number of active components of $s(\cdot \cdot)$ across the 10,000 iterations, after a burn-in period of 20,000. The modal value is 10, with 3644 visit times. Right: draws from the posterior distribution of $\boldsymbol{\mu}$ when the number of components is $k_s = 10$. Using Algorithm 2, we distinguish 11 modal values to use as reference points.	22
1.5	Left column: trace plots of μ_x (top) and μ_y (bottom) when the number of active components is 10. Right column: relabelled trace plots from μ_x (top) and μ_y (bottom). We distinguish two new components, labelled as 10 and 11, that previously belonged to the 10 th component.	23
1.6	Left: $10^\circ \times 10^\circ$ region from the Fermi LAT telescope. The available dataset based on 9 years of observation is composed by 98,598 events, binned into a 200×200 grid. The scale colour value varies with respect to the number of counts per pixel observed. Right: contour plot of the estimated background density using the spatial-only model, based on 20,000 records sampled at random from the starting dataset.	28
2.1	DAG of the continuous time hidden Markov model. Grey circles are the data and white circles represent the latent Markov process.	36
2.2	The γ -ray light curve from the blazar PKS 1510-05 recorded by the <i>Fermi</i> LAT telescope over 630 observation times. The most frequent time gap is $\Delta_i = 2$, and the largest is $\Delta_i = 60$. The light curve is evidently punctuated by flares over the entire observation period.	42
2.3	Left: γ -ray light curve of the blazar PKS 1510-05 (solid black line) against the mean (solid red line) and 95% confidence interval (dashed lines) of the predictive density. Right: histogram of the flux compared to the limit distributions of the process in Formula (2.7) (red line) and of a single log-OU process (blue line).	43

-
- 2.4 Left: autocorrelation function of the model residuals. Right: plot of the classification of the flux measurements. For every data point i , the colour represents $\hat{\xi}_i(2)$, the estimated probability to be a flaring emission. A shift toward red states that the observation is more probable to come from the flaring activity. 44
- 2.5 Bootstrap distributions of the estimators of $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\tau}$ in the normal (top line) and log-normal (bottom line) distributions based on $B = 200$ replicates. 45
- 2.6 Distribution of $\max(Y)$ based on $B = 200$ bootstrap replicates. The red line corresponds to the maximum in the observed light curve \mathcal{Y} . The proportion of bootstrap values larger than the observed maximum is 0.16. 46

List of Tables

1.1	Details of the results given by the relabelling procedure. From left to right, the columns display: (1) the cluster labels, (2) the source IDs when the sources are within 0.1° from the estimated direction, (3) the true direction of the source (4) the expected number of photons from the estimated cluster, (5) the 90% HPD interval of the cluster size, (6) the true number of photons from the source and (7) the estimated probability that the clusters correspond to real sources.	25
1.2	Results obtained fitting the proposed models on the three simulated datasets. Each row displays, for a given scenario (first column), the number of true and spurious clusters with the type of model in the second column. For ease of presentation, we distinguish the results according to the estimated probabilities of being a real source, and we group them into three equally large intervals called respectively low [0 - 0.333], medium [0.334 - 0.666] and large [0.667 - 1]. For every interval, we report the number of sources discovered (denoted as true) and the spurious clusters (denoted as noise).	26
1.3	Results for the <i>Fermi</i> LAT data. The estimated directions of the clusters are shown in the second and in the fourth columns, while the third and the fifth display the estimated posterior probabilities of being a source. The first cluster receives a non negligible posterior evidence from both models. Its closest source is 0.93° away, but its presence is not supported by our models. Cluster IDs 2 and 4 are less than 0.1° away from two sources recorded in the third <i>Fermi</i> LAT catalogue. The two models diverge about the direction of third cluster, as the estimated points are 0.57° away. Finally, the fifth cluster is selected only by the spatial model with very low posterior probability.	29
2.1	Estimates of the model parameters in the two latent states. From left to right: mean, square of the volatility, speed of mean reversion and probability to remain in the same state after an interval $\Delta_i = 2$. The standard errors obtained with $B = 200$ bootstrap replicates are given in parenthesis. The last two columns show the mean and the variance of the limit distributions in the two states.	43
A1	Lower and upper bounds of the full-conditional distribution of the spline knots. $\omega_{(1)}$ is the lowest value in $\{\omega_i\}_{i=1}^m$, while $\omega_{(n)}$ is the largest.	54

Introduction

*“It was here already, long ago;
it was here before our time.”*

Solomon - *Ecclesiastes*

Overview

Statistical learning, by which we mean the ability of discovering patterns and regularities in the data, plays a central role in knowledge discovery. This also includes allocating objects to a pre-assigned or unknown number of groups according to a set of observed attributes or features, which is a natural activity of any science. A major distinction is made depending on whether the groups are defined, and known a priori, or need be detected using the data. Clustering, or unsupervised learning, considers the latter situation. A surge of techniques has been proposed over the years, which differ significantly in their definition of what a cluster is and how to identify it. A precise statistical notion of what a group is, is provided by the density-based approach. Here, the clusters are associated with some specific features of the probability distribution which is assumed to underlie the data. The model-based or parametric approach has received a large attention in literature, and aims to represent the probability distribution of the data as a mixture of parametric distributions. A cluster is associated with each component of the mixture and the observations are allocated to the cluster with maximal density among the components. Standard accounts are the seminal works of [Fraley and Raftery \(1998\)](#) and [Fraley and Raftery \(2002\)](#).

Discovering and locating high-energy γ -ray sources in the whole sky map is a declared target of the *Fermi* Gamma-ray Space Telescope collaboration. The knowledge about these celestial objects can improve our understanding of high-energy astrophysical phenomena, and will help to resolve the mystery of the fundamental nature of dark matter. The data collected by the *Fermi* Large Area Telescope (LAT) are counts of

γ -ray photons which cover the energy range from about 20 MeV to more than 300 GeV. Each detected photon represents an event which is characterized by several variables, the primary ones are the direction of the photon, expressed in Galactic coordinates, its energy content and the so-called event type which expresses the quality of the measurement. Using these data, astrophysicists are interested in determining the number of the emitting extra-galactic high energy sources, measuring their intensities, and pooling the individual photon counts into the corresponding clusters. A major challenge is how to efficiently separate the signal of the emitting source from noise. The *Fermi* LAT data, in particular, are characterized by two types of noise: measurement error associated with the components of the LAT (tracker, calorimeter etc.)¹ and the diffuse γ -ray background which spreads over the entire area observed by the telescope. A consistent literature on signal extraction methods for observational astronomy has been developed across the years [Hobson *et al.* 2010, Section 7]. Many contributions require the whole sky map to be split into small regions, where to search for one source at time. The ongoing sky surveys provide however billions of photon counts from sky regions which are too large to be analysed with small area techniques. This motivated our research, whose aims are to expand the current statistical toolbox of the analysis of photon count maps by providing advanced statistical techniques that can simultaneously locate multiple sources, while covering a very large sky area, and separate them from the heavy and irregularly shaped background contamination.

Straight after the signal detection, astronomers need to classify the nature of the sources measuring the variation of brightness in time. Many observed sources in the high-energy astrophysics, such as blazars, supermassive black holes and merging neutron stars, are intrinsically time-varying in their light emission due to several physical mechanisms. Establishing when such variations initiate and identifying their duration is a crucial step towards the comprehension and classification of the nature of the occurring phenomena. The literature of the last two decades has largely focused on low-energy sky surveys and on the analysis of celestial objects with prevalently regular emission activity during time. This has allowed astrophysicists to formulate ad-hoc models for any specific case. However, the recent observations of high-energy phenomena are characterised by the presence of irregular patterns and aperiodic variations in brightness [Dodds-Eden *et al.* 2011, Ramakrishnan *et al.* 2015], which are likely due to multiple underlying physical processes. This increased the demand of advanced statistical methods to describe the emission activity of time-varying sources. The hope is that the more accurate models used to fit the light emission of high-energy sources across time will

¹<https://www.nasa.gov/content/goddard/fermi-spacecraft-and-instruments>

lead to a deeper understanding of the various phenomena of the Universe.

Main contributions of the thesis

This research project aims to investigate the suitability of advanced parametric and non-parametric mixture models from both the frequentist and the Bayesian perspectives, for the effective analysis of spatial and temporal observations in high-energy astrophysics.

Contribution 1: A hierarchical Bayesian nonparametric model for source extraction and background separation

In the first part of the thesis, we propose an innovative approach which uses Bayesian nonparametric techniques to extract the signal of astronomical sources in gamma-ray count maps under the presence of a strong background contamination. Our model simultaneously induces clustering on the photons using their spatial information and gives an estimate of the number of sources, while separating them from the irregularly shaped background which extends over the entire map. From the statistical perspective, the signal of the sources is modelled using a Dirichlet Process mixture, which allows us to discover and locate a possible infinite number of clusters. The background contamination on the other hand is completely reconstructed using a novel flexible Bayesian nonparametric model based on B-spline basis functions. The result can be then thought of as a hierarchical mixture of nonparametric mixtures for flexible clustering of highly contaminated signals. We provide also a Markov chain Monte Carlo algorithm to infer the posterior distribution of the model parameters which does not require any tuning parameter, and a suitable post-processing algorithm to quantify the information coming from the detected clusters. The analyses of several artificial datasets confirm the capacity of the model to discover and locate the sources in the map, to quantify their intensities and to estimate and account for the presence of the background contamination. Last, we will give an illustration based on the *Fermi* LAT data.

Contribution 2: Continuous time hidden Markov modelling for flare detection in astronomical gamma-ray light curves

In the second part of the project, we develop a new model for astronomical gamma-ray light curves which overcomes the limitations of the existing methods and is capable of separating the distinct states in the emission activity of a source. The detection of a flare can be posed as a clustering problem where flux measurements correspond to either a resting or to a flaring phase. We propose a novel multi-state model for gamma-ray

light curves with the aim of disentangling the flaring activity. The switching between resting and flaring is modelled through a latent continuous time Markov process that can accommodate irregularly spaced observation times. We adopt distinct OU processes to characterize the flux variation in the discovered phases. The resulting compound model can be fit via maximum likelihood estimation using an efficient EM algorithm. We develop also a suitable bootstrap algorithm to evaluate the adequacy of the model and to test whether evidence of additional states emerges from the data. We finally propose an illustrative example based on the signal of a blazar monitored by the *Fermi* LAT. Results confirm the adequacy of our model to successfully identify the emission phases of the light curve which correspond to distinct physical mechanisms.

Chapter 1

Source extraction and background separation: a Bayesian nonparametric approach

“Not all those who wander are lost;”
John R. R. Tolkien - *The Lord of the Rings*

1.1 Introduction

Source extraction in astronomy aims to discover new astronomical sources and study their physical characteristics. The ultimate purpose is to extend our knowledge about the incredibly various energetic phenomena that originate in the universe.

In the last twenty years, new technologies have massively increased the precision of detectors and the size of storage systems. The resulting richness of data resources poses significant statistical challenges and has stimulated the development of new statistical techniques and advanced computational methods that culminate into more powerful and precise source extraction approaches.

From a statistical perspective, extracting a source means quantifying the evidence of its presence in an observed region of the sky using a collection of photons, also known as events. This is challenging in practice because the source signals are embedded in and thus masked by background contamination, which is generally a significant component of the data. The intensity of the contamination varies both with sky coordinates and with the energy, i.e., across the joint domain of the image and the spectrum, that is the distribution of the energy. Regions in the so-called *galactic space*, for example, are subject to contamination from the nearby Milky Way galaxy, while more distant

locations in the *extragalactic space* are dominated by the isotropic γ -ray background (IGRB) component and thus are not subject to heavy contamination.

1.1.1 Astrophysical source extraction

Source extraction has received growing interest both in the astronomical and the statistical literatures. [Hobson *et al.* \(2010\)](#) differentiate methods according to whether they are designed to identify a single source or multiple sources simultaneously. While the former has received considerable attention since the early 1990s [[Kraft *et al.* 1991](#), [Mattox *et al.* 1996](#), [van Dyk *et al.* 2001](#), [Protassov *et al.* 2002](#), [Park *et al.* 2006](#), [Weisskopf *et al.* 2007](#), [Knoetig 2014](#)], the latter have received a growing interest only more recently, both for X-ray [[Guglielmetti *et al.* 2009](#), [Primini and Kashyap 2014](#), [Jones *et al.* 2015](#)] and γ -ray data [[Acero *et al.* 2015](#), [Selig *et al.* 2015](#)]. The high computational costs of simultaneous extraction have doubtlessly delayed their development. However, their features make them more attractive than single extraction methods as they allow to analyse a whole region of the sky all at once.

Signal extraction methods must duly concern for the presence of the background component in the data. This is particularly true for simultaneous extraction as the intensity of the contamination can significantly vary across the map. When the background is small or relatively constant over the region of interest, one can model the spatial and the energy components of the background as uniform as proposed by [Jones *et al.* \(2015\)](#), or alternatively consider the Bayesian aperture photometry approach of [Primini and Kashyap \(2014\)](#) for low count images.

For intense and prominent background contaminations which incorporate many heterogeneous processes, [Guglielmetti *et al.* \(2009\)](#) define a Poisson-based mixture model for X-ray count images that account for the background using thin-plate splines. As an alternative, the recent work of [Acero *et al.* \(2016\)](#) develops a detailed reconstruction of the γ -ray background for the whole sky using 8 years of observations from the *Fermi* LAT telescope; this precise background estimate can be used for example with the approach of [Stein *et al.* \(2015\)](#), which is thought to detect unknown structures in an image for which part of the information, and thus also the background, is known. [Acero *et al.* \(2015\)](#) perform multiple source extraction exploiting a preliminary version of the [Acero *et al.* \(2016\)](#) background estimate. However, the model of [Acero *et al.* \(2016\)](#) may include any source that was not accounted for as part of the background, masking its signal in a subsequent study. In parallel, [Selig *et al.* \(2015\)](#) exploit the method of [Selig and Enßlin \(2015\)](#) based on *Information Field Theory* to propose their

own reconstruction of the *Fermi* LAT sky. However, the problems encountered with the model of [Acero *et al.* \(2016\)](#) are still extant.

The data are collected by high-energy detectors in the form of photons with spatial direction in the rectangular map

$$\mathcal{X} = \{(x, y) : x \in (x_{\min}, x_{\max}), y \in (y_{\min}, y_{\max})\},$$

and energy in the interval

$$\mathcal{E} = \{E : E \in (E_{\min}, E_{\max})\}.$$

We propose here a novel approach to simultaneously extract the signal of high-energy sources and reconstruct the background contamination that extends over a map. The method exploits the spatial coordinates and the energy of the photons (x, y, E) to probabilistically allocate them to the corresponding sources, for which secondary analyses can be accomplished. In practice, continuous measurements of sky coordinates and energies are not always accessible, in particular when the resolution of the telescope is low. Data are therefore gathered into bins. Nevertheless, our method can be applied also to binned data by treating the centroid of the bins as continuous measures. This strategy will of course work best with high resolution data, i.e., when pixels of the instrument are small.

1.1.2 Main goals and outline

Through this chapter, we largely exploit Bayesian nonparametric modelling to overcome the limitations of the existent approaches in locating the sources in a map highly contaminated by background. In particular, the new method we propose has the following features: (i) it dynamically determines the number of sources in the map, (ii) it clusters the photons according to their spatial direction and energy, and (iii) it flexibly estimates the background contamination without neither relying on previous knowledge of the background map, nor using empirical reconstructions of it.

Most of the available simultaneous extraction methods require to set the number of sources at the beginning of the analysis [[Guglielmetti *et al.* 2009](#), [Ray *et al.* 2011](#), [Acero *et al.* 2015](#)] or are computationally limited [[Primini and Kashyap 2014](#)]. The more recent work of [Jones *et al.* \(2015\)](#), which inspired this research, considers a Bayesian extraction method based on mixtures, and the number of sources is inferred using the reversible

jump Markov chain Monte Carlo algorithm [Green 1995]. This chapter exploits the advantages of mixture modelling and extends the model of Jones and coauthors using an infinite mixture induced by a Dirichlet Process prior [Ferguson 1973]. In this way, the model is conceptually appropriate as it dynamically extracts the sources in a map and admits also new unobserved signals that would become detectable when new photons will be available. In addition, the Gibbs sampling algorithms for Bayesian nonparametric methods [Müller *et al.* 2015] scale the mixture size faster than Green’s algorithm, guaranteeing practical advantages in the model estimation.

In parallel to the signal extraction, our method provides also an accurate and flexible model for the background contamination. This component is crucial in the detection of new sources, as a wrong choice of the background distribution may lead to misleading results and causes the identification of spurious sources. Jones *et al.* (2015) model the background as uniform either on the map and the spectrum. However, this assumption is unrealistic for heavily contaminated regions as the ones close to the *Galactic space*. We propose a new model that combines innovative Bayesian nonparametric techniques with B-spline functions [de Boor 2001] for flexible reconstruction of bivariate irregular signals. Some comparable approaches in literature [Denison *et al.* 1998, Biller 2000, DiMatteo *et al.* 2001, Sharef *et al.* 2010] adopt the reversible jump algorithm to select the active spline functions in the model. As discussed above, simulation algorithms for Bayesian nonparametric methods are particularly advantageous and lead to a fast model selection. With respect to Guglielmetti *et al.* (2009) and Schellhase and Kauermann (2012), we do not rely on any minimisation problem that in practice can be unfeasible. We provide also a Gibbs sampler for spline knots that is free from any tuning parameter.

The Chapter is structured in 6 sections. Section 1.2 presents our innovative model based on Bayesian nonparametric mixtures for signal extraction from highly contaminated astronomical data. We first propose a model that only accounts for the sky coordinates, and then extend the model to also account for photon energies. Section 1.3 shows that the model developed in Section 1.2 is a mixture of Dirichlet Process mixtures and shows how it can be fit with a Gibbs sampler. The model is validated in Section 1.4 with a suite of simulation studies, first illustrating the method and then demonstrating the advantage of considering spectral data. Some results from a region of the *Fermi* LAT sky are presented in Section 1.5. Final discussion appear in Section 1.6, and technical details of the Gibbs sampling algorithms in the Appendix.

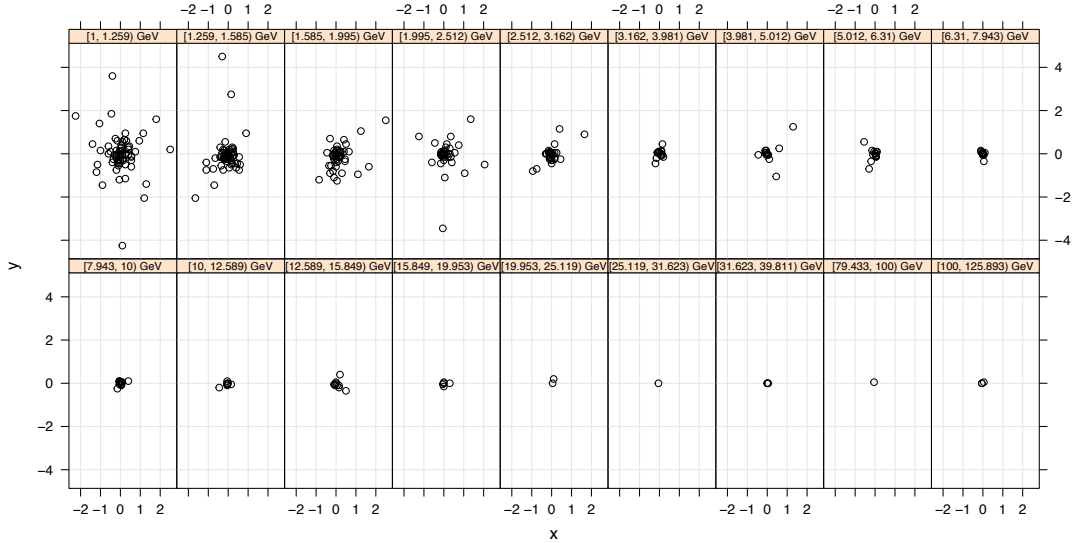


FIGURE 1.1: Photons emitted from an astronomical source located at $(0, 0)$ in different energy bins. Increasing the energy range (from left-top to right-top, and from left-bottom to right-bottom) the events become rarer and concentrate around the direction of the source.

1.2 The statistical model

Let $i = 1, \dots, n$ denotes a collection of observed photons with recorded sky coordinates $\mathbf{x}_i = (x_i, y_i) \in \mathcal{X}$ and energy $E_i \in \mathcal{E}$. As a single photon may originate from either one of the astronomical sources or from the background, we formalize the statistical model for the sky coordinates as

$$f(\mathbf{x}_i|\Theta) = \delta s(\mathbf{x}_i|\vartheta_s) + (1 - \delta)b(\mathbf{x}_i|\vartheta_b), \quad (1.1)$$

where $s(\cdot|\cdot)$ and $b(\cdot|\cdot)$ are the density functions for the combined sources and for the background, respectively parametrised by the vectors ϑ_s and ϑ_b . Finally, $\delta \in (0, 1)$ is a mixing parameter, which we assume to be a priori $Beta(\lambda, \lambda)$ distributed, where $\lambda > 0$ is a fixed hyperparameter. Thus, the full parameter is $\Theta = \{\vartheta_s, \vartheta_b, \delta\}$.

1.2.1 The source model

We investigate a modelling approach for $s(\cdot|\cdot)$ starting from the simplest case, a unique source with spatial direction $\boldsymbol{\mu} = (\mu_x, \mu_y)$. Although the considered source is point-like, and thus the direction of every emanated photon \mathbf{x}_i should be equal to $\boldsymbol{\mu}$, the signal is spread around the true location of the source because of the instrument response function, also known as *Point Spread Function* (PSF). Telescopes are sensible to the energetic value of the incoming events: high-energy photons are rarer and concentrate

around the source direction $\boldsymbol{\mu}$, while low energy photons are more frequent and appear much more spread. Figure 1.1 illustrates the variation of both the number of incoming photons and their spread around the centre of the source $(0^\circ, 0^\circ)$ as the energy increases according to the LAT PSF [Ackermann *et al.* 2012].

The LAT PSF is known in pixelized form; a parametric formulation is not available. For ease of interpretation and computation, we opt for a parametric approximation of the PSF. Specifically, following King (1962), we use

$$\mathbf{x}_i | E_i, \boldsymbol{\mu} \sim t_{\nu_i}(\boldsymbol{\mu}, \sigma_i^2 I), \quad (1.2)$$

where $t_d(\mathbf{a}, \mathbf{C})$ is a bivariate Student t distribution with vector of location parameters \mathbf{a} , scale matrix \mathbf{C} and d degrees of freedom, I is the 2×2 identity matrix, $\sigma_i^2 = \sigma^2(E_i)$ and $\nu_i = \nu(E_i)$ are known coefficients which depend on the energy E_i . Details on the approximation of the PSF are given in the Appendix A1.

Once the single source model is established, some aspects must be considered before extending to multiple sources. First, the number of underlying sources in a map is unknown. Second, each source can have a different intensity, and third, no a priori information on their locations in space is available. We hence translate all these assumptions into the model

$$s(\mathbf{x}_i | E_i, \mathcal{F}) = \int p(\mathbf{x}_i | \boldsymbol{\mu}, \sigma_i^2, \nu_i) \mathcal{F}(d\boldsymbol{\mu}), \quad \mathcal{F} \sim \mathcal{DP}(\alpha_s, \mathcal{F}_0), \quad (1.3)$$

where $\mathcal{DP}(r, \mathcal{P}_0)$ is a Dirichlet Process prior with concentration parameter r and base measure \mathcal{P}_0 and p is the density function of a bivariate Student t distribution as in (1.2). Model (1.3) is known in the literature as Dirichlet Process mixture and estimates $s(\cdot | \cdot)$ mixing the kernel p with respect to the unknown measure \mathcal{F} . \mathcal{F}_0 has the role of prior distribution over the location parameter $\boldsymbol{\mu}$ and it is taken to be uniform over \mathcal{X} . For a review on Bayesian nonparametrics, see Müller *et al.* (2015).

The proposed model allows then a generic photon i to come from a possibly infinite number of sources that a priori can be anywhere in the map. In practice, just a finite number of observations, say n_s , come from the sources, and for this reason (1.3) will appear as a finite summation of at most n_s components.

1.2.2 The background model

The second component of (1.1), $b(\cdot | \cdot)$, represents the total amount of background in the map. As mentioned in Section 1.1, high-energy count images use to be heavily affected

by an irregular contamination at every level of energy; additionally, the background noise is fruit of several astronomical phenomena that cannot be predicted and prevented. These considerations translate into the practical impossibility of deriving an analytical distribution for the background, but encourage to abandon any parametric distribution in favour of flexible non-parametric solutions.

We recall here the definition of B-spline basis function of order m defined on a set of knots $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{m+1})$:

$$\mathcal{B}_m(x|\boldsymbol{\tau}) = \frac{x - \tau_1}{\tau_m - \tau_1} \mathcal{B}_{m-1}(x|\boldsymbol{\tau}_{-(m+1)}) + \frac{\tau_{m+1} - x}{\tau_{m+1} - \tau_2} \mathcal{B}_{m-1}(x|\boldsymbol{\tau}_{-1}), \quad (1.4)$$

where $\mathcal{B}_1(x|(a, b)) = \mathbb{1}_{[a, b)}(x)$ and $\boldsymbol{\tau}_{-j}$ is the vector $\boldsymbol{\tau}$ without its j -th element. By construction, the function (1.4) is always positive between τ_1 and τ_{m+1} , while it is zero outside and unimodal for $m > 1$. Furthermore, (1.4) can assume many different shapes depending on the location of the knots, and can be easily normalised by $\tilde{\mathcal{B}}_m(x|\boldsymbol{\tau}) = m/(\tau_{m+1} - \tau_1) \mathcal{B}_m(x|\boldsymbol{\tau})$; thus, $\tilde{\mathcal{B}}(\cdot|\cdot)$ is a distribution to all effects. See [de Boor \(2001\)](#) for a full review on spline functions.

These characteristics motivate the use of B-splines for modelling complex data, such as the background contamination in this study. Let us consider an event with direction \mathbf{x}_i from the background component. First, we define the bivariate density function

$$\varphi(\mathbf{x}_i|\boldsymbol{\ell}, \mathbf{b}) = \tilde{\mathcal{B}}_4(x_i|\boldsymbol{\ell}) \tilde{\mathcal{B}}_4(y_i|\mathbf{b}), \quad (1.5)$$

where $\boldsymbol{\ell} = (\ell_1, \dots, \ell_5)$ and $\mathbf{b} = (b_1, \dots, b_5)$ denote respectively the knots of the longitude and of the latitude B-splines, both with order fixed to 4 to guarantee enough flexibility of the density. We model the background as

$$b(\mathbf{x}_i|\mathcal{G}) = \int \varphi(\mathbf{x}_i|\boldsymbol{\ell}, \mathbf{b}) \mathcal{G}(d\boldsymbol{\ell}, d\mathbf{b}), \quad \mathcal{G} \sim \mathcal{DP}(\alpha_b, \mathcal{G}_0). \quad (1.6)$$

Formula (1.6) is an infinite mixture of normalised B-spline functions induced by a Dirichlet Process prior, although the model will appear in practice as a finite mixture of at most n_b components, where n_b is the unknown number of photons from the background. The base measure of the process \mathcal{G}_0 works as a prior distribution for the set of longitude and latitude knots. The order of the knots in the vector is guaranteed by assuming the first of the five elements to be uniform distributed over the limits of the map, i.e., $\ell_1 \sim \mathcal{U}(x_{min}, x_{max})$ and $b_1 \sim \mathcal{U}(y_{min}, y_{max})$, and the remaining to be uniform distributed as well, but bounded to the left of their precedents. So, the knots for the longitude coordinate distribute as $\ell_k \sim \mathcal{U}(\ell_{k-1}, x_{max})$, and for the latitude as $b_k \sim \mathcal{U}(b_{k-1}, y_{max})$,

for $k = 2, \dots, 5$.

1.2.3 Including the spectral information

The convolution of the two Dirichlet Process mixtures in Formula (1.1) exploits only the spatial information of the recorded photons. A direct extension includes photons energy as a third response variable of the model. Jones *et al.* (2015) explored this approach and stated that, in the X-ray framework, the extended full model improves considerably the detection performance compared to the only spatial version and reveals itself a good solution for disentangling overlapping sources. In the γ -ray field, Acero *et al.* (2015) showed that a variation in the emission activity of many sources monitored by the LAT, caused by an increment of the energy, has a power-law decay. Because of the complex structure of the background contamination, it is challenging to define an appropriate model for the energy of the photons. In this section, we propose a simple extension of (1.1) as follows:

$$f^{ext}(\mathbf{x}_i, E_i | \Theta^{ext}) = \delta s^{ext}(\mathbf{x}_i, E_i | \vartheta_s^{ext}) + (1 - \delta) b^{ext}(\mathbf{x}_i, E_i | \vartheta_b^{ext}), \quad (1.7)$$

where the unknown functions s^{ext} and b^{ext} are equal to

$$\begin{aligned} s^{ext}(\mathbf{x}_i, E_i | \vartheta_s^{ext}) &= s(\mathbf{x}_i | E_i, \mathcal{F}) g(E_i | E_{\min}, \eta_s), \\ b^{ext}(\mathbf{x}_i, E_i | \vartheta_b^{ext}) &= b(\mathbf{x}_i | \mathcal{G}) g(E_i | E_{\min}, \eta_b), \end{aligned} \quad (1.8)$$

and $g(\cdot | e, \eta)$ is the density function of a Pareto distribution, that has a power-law kernel, with shape parameter η and scale $e = E_{\min}$. Some alternatives listed by Acero *et al.* (2015) are the exponential and the log-normal distributions; however, we are not going to exploit them in this work. The Gamma distribution for the shape parameter is conjugate with the Pareto model, so we take $\eta_s \sim \text{Gamma}(a_{\eta_s}, b_{\eta_s})$ and $\eta_b \sim \text{Gamma}(a_{\eta_b}, b_{\eta_b})$.

In (1.8), we model the photon energy of the sources and of the background with a single Pareto distribution. This assumption is rather simplistic, but it is useful in a first stage of the analysis to explore whether the energy variable will be helpful or not to increase the detection performance of the model. Future developments will focus on how to appropriate model the energy of the photons from the background, and on how to allow the point sources to have distinct spectral models.

1.2.4 Misclassification effect

From a physical perspective, the astronomical sources and the background contamination show different characteristics: the background tends to be much smoother than the signal of point sources, and huge spikes are considerably unlikely. However, real data are heavily contaminated and show often irregular behaviours which may cause misclassification effects when models are fitted. We thus distinguish two types of error.

We refer to the first as *Type I* error: it occurs when groups of photons from the background are spatially gathered in a way similar to the signal from a point source. This phenomenon is fairly common in the analysis of astronomical maps and cannot be prevented, as it strongly depends from how the photons from the background are spatially distributed. In Section 1.4.1 we will discuss a proper method to evaluate whether the components of the mixture $s(\cdot|\cdot)$ represent astronomical sources using the draws from the posterior distribution of the model parameters.

The second type of error, called *Type II* error, occurs when the signal from a point source is captured by the background model. This effect is mainly caused by the extreme flexibility of B-spline functions combined with Bayesian nonparametric methods; then, the probability that the mixture $b(\cdot|\cdot)$ embeds some weak sources is not negligible. In order to prevent this effect, we control the shape of the B-splines kernel imposing some restrictions on the parameter space. Specifically, these restrictions work on the variance of the normalized B-spline functions. From Carlson (1991), if a random variable X is distributed according to the density function $\tilde{\mathcal{B}}_m(\cdot|\boldsymbol{\tau})$, then its variance is

$$\text{Var}(X) = \mathcal{V}(\boldsymbol{\tau}) := \frac{\sum_{p=1}^m \sum_{q=p+1}^{m+1} (\tau_p - \tau_q)^2}{(m+1)^2(m+2)}. \quad (1.9)$$

According to (1.2), the covariance matrix of the spatial location of a photon i from a point source is $\text{Var}(\mathbf{x}_i|E_i) = v_i I$, where

$$v_i = \frac{\nu_i \sigma_i^2}{\nu_i - 2}$$

is a function of the known set of parameters (σ_i^2, ν_i) and $\nu_i > 2$. Let $\tilde{v} = \max_i v_i$. Then, for every component l of the background mixture model, we want the set of knots $\boldsymbol{\ell}_l$ and \mathbf{b}_l to respect the following constraint,

$$\mathcal{V}(\boldsymbol{\ell}_l) > c\tilde{v}, \quad \mathcal{V}(\mathbf{b}_l) > c\tilde{v}, \quad (1.10)$$

where c is tuned to control the smoothness of the background component. If c is small,

the above constraint becomes inefficient to prevent the misclassification; on the contrary, if c is too large, the posterior distribution of $\boldsymbol{\ell}$ and \mathbf{b} under (1.10) becomes computationally infeasible to treat. In this work, we suggest to guide the selection of c using a pre-existing background model, when available, as the one provided by Acero *et al.* (2016). In details, we can fit just the B-spline Dirichlet Process mixture (1.6) on a dataset drawn from the available background model and tune the value of c by assessing the variances of the mixture components, taking into account also the grade of smoothness we want our background model to have. In this work, we set $c = 1.5$ as it approximately coincides with the posterior modes of $\mathcal{V}(\boldsymbol{\ell})/\tilde{v}$ and $\mathcal{V}(\mathbf{b})/\tilde{v}$. We will further show in Section 1.4 that this value guarantees a good separation of the background from the sources and at the same time prevents the *Type II* misclassification. In alternative, one can set a softer constraint by considering for example $c = \min(\mathcal{V}(\boldsymbol{\ell}))/\tilde{v}$, which guarantees a more flexible background reconstruction but augments the risk of embedding some sources.

If instead $\nu_i \leq 2 \forall i$, the variance of the Student t distribution is undefined and the constraint (1.10) becomes just

$$\mathcal{V}(\boldsymbol{\ell}_i) > c, \quad \mathcal{V}(\mathbf{b}_i) > c.$$

1.3 Model generalization and simulation algorithm

1.3.1 A mixture of Dirichlet Process mixtures

A generic mixture of Dirichlet Process mixtures can be expressed as follows. Let $\{\mathbf{y}_i\}_{i=1}^n$ where $\mathbf{y}_i \in \mathcal{Y} \subseteq \mathbb{R}^p$. The observations originate from

$$f(\mathbf{y}_i | \boldsymbol{\delta}, \mathcal{K}_1, \dots, \mathcal{K}_J) = \sum_{j=1}^J \delta_j g_j(\mathbf{y}_i | \mathcal{K}_j), \quad (1.11)$$

$$g_j(\mathbf{y}_i | \mathcal{K}_j) = \int q_j(\mathbf{y}_i | \theta_j) \mathcal{K}_j(d\theta_j), \quad \mathcal{K}_j \sim \mathcal{DP}(\alpha_j, \mathcal{K}_{0,j}).$$

where $\boldsymbol{\delta} \sim \text{Dir}(\boldsymbol{\lambda})$. We furthermore assume that J is fixed, $\theta_j \in \Theta_j$ and the size $\dim(\Theta_j)$ may differ for any j .

The spatial model (1.1) directly derives from the above formulation by setting $J = 2$, $g_1(\mathbf{x}_i | \mathcal{K}_1) = s(\mathbf{x}_i | E_i, \mathcal{F})$, $g_2(\mathbf{x}_i | \mathcal{K}_2) = b(\mathbf{x}_i | \mathcal{G})$, $q_1(\mathbf{x}_i | \theta_1) = p(\mathbf{x}_i | \boldsymbol{\mu}, \sigma_i^2, \nu_i)$ and $q_2(\mathbf{x}_i | \theta_2) = \varphi(\mathbf{x}_i | \boldsymbol{\ell}, \mathbf{b})$. The base measure $\mathcal{K}_{0,1} = \mathcal{F}_0$ is flat over \mathcal{X} and $\mathcal{K}_{0,2} = \mathcal{G}_0$ is a sequence of uniform distributions as described in Section 1.2.2. To extend the model as described

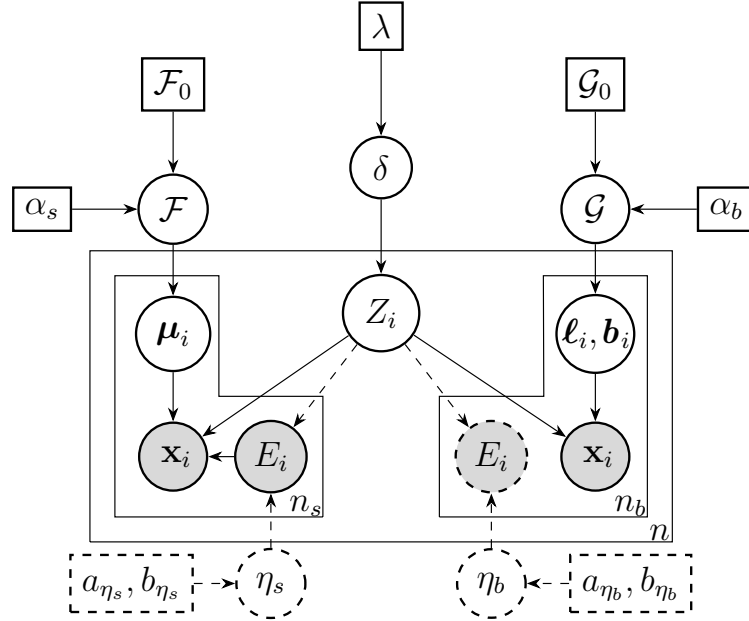


FIGURE 1.2: Directed Acyclic Graph (DAG) which illustrates the dependency structure among the variables of the model. The solid lines represent Model (1.1), and both solid and dashed lines describe Model (1.7).

in Section 1.2.3, we just multiply the densities $g_1(\mathbf{x}_i|\mathcal{K}_1)$ by $g(E_i|E_{\min}, \eta_s)$ and $g_2(\mathbf{x}_i|\mathcal{K}_2)$ by $g(E_i|E_{\min}, \eta_b)$.

We describe now the dependency structure of quantities in the proposed models using a Directed Acyclic Graph (DAG), as shown in Figure 1.2. The edges represent the objects of the models: the data are shown as grey circles, the model parameters are white circles and the hyperparameters are white rectangles. The arrows outline the relationship among the elements: if an arrow connects an edge A to an edge B , then B depends from A . The dependency structure of the spatial model $f(\cdot|\cdot)$ is described only by the solid lines, while both solid and dashed lines display the extended model $f^{ext}(\cdot|\cdot)$.

We display in the DAG also the latent group variable Z_i that highlights whether the photon i is deemed to come from the astronomical sources or from the background. Thus, the n events are split into a first group of n_s observations assigned to the sources, and into a second group of n_b observations assigned to the background. On the left side of the edge Z_i , we represent the mixture model $s(\cdot|\cdot)$ used to perform the signal extraction of the sources. Following the directions of the arrows we can follow how the Bayesian nonparametric model reconstructs the unknown generating process of the data. First, the set of location parameters $\{\mu_1, \dots, \mu_{n_s}\}$ is sampled from a random probability measure \mathcal{F} that comes from $\mathcal{DP}(\alpha_s, \mathcal{F}_0)$. Then, conditioning on μ_i , the model generates the direction of the photon \mathbf{x}_i using a bivariate Student t distribution whose scale and

degrees of freedom are (σ_i^2, ν_i) and given by the photon energy E_i . On the right side of Z_i we display the mixture model $b(\cdot|\cdot)$ which represents the background. The model parameters $\{(\ell_1, \mathbf{b}_1), \dots, (\ell_{n_b}, \mathbf{b}_{n_b})\}$ come from a probability measure \mathcal{G} sampled from a Dirichlet Process prior $\mathcal{DP}(\alpha_b, \mathcal{G}_0)$. The spatial location of photon i is then generated using the bivariate spline function $\varphi(\cdot|\ell_i, \mathbf{b}_i)$.

According to the extension of the model presented in Section 1.2.3, the energy E_i is Pareto distributed with scale η_s when i is assigned to the sources, and with scale η_b when i is assigned to the background.

Because of the discreteness of random probability measures sampled by Dirichlet Process priors [Ferguson 1973, Blackwell 1973], we will observe just a subset of k_s distinct values in $\{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{n_s}\}$, and a subset of k_b in $\{(\ell_1, \mathbf{b}_1), \dots, (\ell_{n_b}, \mathbf{b}_{n_b})\}$. Therefore, the model produces two distinct levels of clustering: the first is given by the values of Z , and the second by the photons with the same value of the model parameters. This aspect is particularly relevant for the model $s(\cdot|\cdot)$, as each component of the mixture has a physical interpretation and represents the signal of a distinct astronomical source. The components of $b(\cdot|\cdot)$ instead do not have a particular physical meaning and represent just separate regions of the background contamination.

1.3.2 The MCMC algorithm

We develop a simulation algorithm in the class of Markov chain Monte Carlo methods for posterior inference of mixtures of Dirichlet Process mixture models. The pseudocode of Code-Box 1 extends the Gibbs sampling scheme for Bayesian nonparametric mixtures based on the Chinese Restaurant Process [Escobar 1994, Escobar and West 1995, Section 3.3.1 of Müller and Rodriguez 2013] to include the two orders of clustering described at the end of the previous section. The algorithm refers to the generic mixture model (1.11) and can be used to fit both the proposed models $f(\cdot|\cdot)$ and $f^{ext}(\cdot|\cdot)$. We keep track of the first level of clustering through a vector \mathbf{Z} of size n : $Z_i = j$ means that the algorithm assigns the i -th observation to the j -th mixture model, for $i = 1, \dots, n$ and $j = 1, \dots, J$. In accordance with Section 3.3.1 of Müller and Rodriguez (2013), we introduce for all Dirichlet Process mixtures an n -size vector, \mathbf{H}_j , to store the second level of clustering. Thus, $H_{ji} = l$ states that observation i is assigned to the l -th component of the j -th Dirichlet Process mixture, for $l = 1, \dots, k_j$, where k_j is the set of distinct values from the random probability measure \mathcal{K}_j . Additionally, $H_{ji} = 0$ if $Z_i \neq j$. The proposed algorithm is iterative, so the quantity involved are marked with (t) to denote the t -th iteration. When $J = 1$, the algorithm reduces to the standard collapsed Gibbs sampler for Bayesian nonparametric inference.

Algorithm 1: Gibbs sampling for mixture of Dirichlet Process mixtures

The following pseudo-code outlines an iterative algorithm for posterior inference in mixture models as in (1.11). For ease of presentation, we show just a single iteration t given the output from iteration $t - 1$. The procedure must be run for a large number of times T .

Input: $\boldsymbol{\theta}_j^{(t-1)} = (\theta_{j1}, \dots, \theta_{jk_j})^{(t-1)}$ is the set of $k_j^{(t-1)}$ distinct values from \mathcal{K}_j ,
 $\mathbf{g}_j^{(t-1)} = (g_{j1}, \dots, g_{jn})^{(t-1)}$ is the estimate of the density $g_j(\cdot|\cdot)$ over the n observations, $\mathbf{H}_j^{(t-1)}$, an integer h_j , $\delta_j^{(t-1)}$, and (optional) $\alpha_j^{(t-1)}$, for $j = 1, \dots, J$.

begin

Sample a set of additional parameters $\tilde{\boldsymbol{\theta}}_j = (\tilde{\theta}_{j1}, \dots, \tilde{\theta}_{jh_j})$ from $\mathcal{K}_{0,j}$, for $j = 1, \dots, J$. Let $\mathbf{H}_j^* = \mathbf{H}_j^{(t-1)}$, $\boldsymbol{\theta}_j^* = \boldsymbol{\theta}_j^{(t-1)}$ and $k_j^* = k_j^{(t-1)}$.

Step [1] Draw $Z_i^{(t)}$ from

$$p(Z_i^{(t)} = j | g_{ji}^{(t-1)}, \delta_j^{(t-1)}) \propto \delta_j^{(t-1)} g_{ji}^{(t-1)}, \quad j = 1, \dots, J,$$

and draw $\boldsymbol{\delta}^{(t)}$ from $Dir(\mathbf{n}^{(t)} + \boldsymbol{\lambda})$, where $\mathbf{n}^{(t)} = (n_1, \dots, n_J)^{(t)}$ and $n_j^{(t)} = \sum_{i=1}^n \mathbf{1}(Z_i^{(t)} = j)$.

Step [2] for $j = 1, \dots, J$ do

Step [2.a] For $i = 1, \dots, n$, if $Z_i^{(t)} \neq j$, set $H_{ji}^* = 0$. For $l = 1, \dots, k_j^*$, if there isn't any $H_{ji}^* = l$, remove θ_{jl} from $\boldsymbol{\theta}_j^*$ and decrease k_j^* by one.

Step [2.b] For all i such that $Z_i^{(t-1)} \neq j$ and $Z_i^{(t)} = j$, draw from

$$p(H_{ji}^* = l | \mathbf{H}_{j,-i}^*, \boldsymbol{\theta}_j^*, \tilde{\boldsymbol{\theta}}_j, \alpha_j^{(t-1)}) \propto \begin{cases} n_{jl}^* q_j(\mathbf{y}_i | \theta_{jl}^*) & l = 1, \dots, k_j^*, \\ \frac{\alpha_j^{(t-1)}}{h_j} q_j(\mathbf{y}_i | \tilde{\theta}_{jh}) & l = k_j^* + h, \\ & h = 1, \dots, h_j, \end{cases}$$

according to Algorithm 8 of Neal (2000), where $n_{jl}^* = \sum_{i=1}^n \mathbf{1}(H_{ji}^* = l)$ and $\mathbf{H}_{j,-i}^*$ is the vector \mathbf{H}_j^* without the i -th element. If $l > k_j^*$, then $H_{ji}^* = k_j^* + 1$ and increase k_j^* by one. Draw a value for $\theta_{j(k_j^*+1)}$ from

$$\pi(\theta | \mathbf{y}_i) \propto q_j(\mathbf{y}_i | \theta) \mathcal{K}_{0,j}(\theta) \quad (1.12)$$

and add it to $\boldsymbol{\theta}_j^*$.

Step [2.c] For all i such that $Z_i^{(t)} = j$, update H_{ji}^* using the probabilities given by [2.b]. If a cluster becomes empty, remove it and decrease k_j^* by one. If a new component is accepted, increase k_j^* by one, sample a new value from (1.12) and add it to $\boldsymbol{\theta}_j^*$. Finally, conditioning on \mathbf{H}_j^* , update the model parameters $\boldsymbol{\theta}_j^*$. At the end, $\mathbf{H}_j^{(t)} = \mathbf{H}_j^*$, $k_j^{(t)} = k_j^*$ and $\boldsymbol{\theta}_j^{(t)} = \boldsymbol{\theta}_j^*$.

Step [2.d] Let $n_{jl}^{(t)} = \sum_{i=1}^n \mathbf{1}(H_{ij}^{(t)} = l)$. Compute

$$g_{ij}^{(t)} = \sum_{l=1}^{k_j^{(t)}} \frac{n_{jl}^{(t)}}{n_j^{(t)} + \alpha_j^{(t-1)}} q_j(\mathbf{y}_i | \theta_{jl}^{(t)}) + \sum_{h=1}^{h_j} \frac{\alpha_j^{(t-1)}}{h_j} q_j(\mathbf{y}_i | \tilde{\theta}_{jh}), \quad i = 1, \dots, n.$$

Step [2.e] (optional) Update the concentration parameter α_j with the data augmentation strategy of Escobar and West (1995), using $k_j^{(t)}$, $n_j^{(t)}$, a_j and b_j .

Output: $\mathbf{Z}^{(t)}$ and $k_j^{(t)}$, $\mathbf{H}_j^{(t)}$, $\boldsymbol{\theta}_j^{(t)}$, $\mathbf{g}_j^{(t)}$, $\alpha_j^{(t)}$, for $j = 1, \dots, J$.

For every iteration, Step [1] pulls the observations into the J distinct mixture models. Then, for $j = 1, \dots, J$, Step [2] works as follows: [2.a] it removes the observations which previously belonged to mixture j and changed at the current iteration the first level of clustering; [2.b] it assigns the second clustering indicator to the new entries; [2.c] it updates the labels and sample from the posterior distribution of the model parameters; and [2.d] it estimates the density value of the n observations. Step [2.b] and [2.c] work on the indicators of the second clustering level and either assign the data to an existent component or add a new one. With respect to Page 30 of Müller and Rodriguez (2013), we implement Algorithm 8 of Neal (2000) which does not require us to compute the predictive density function, that is often computationally demanding.

Afterwards, the algorithm samples new values from the posterior distribution of the model parameters. The Student t distribution used does not have a conjugate prior for the non centrality parameter $\boldsymbol{\mu}$; however, a data augmentation strategy [van Dyk and Meng 2001] can be implemented to achieve a simple and efficient Gibbs sampler. For what concerns the set of knots $(\boldsymbol{\ell}, \mathbf{b})$ of the background model, a closed conditional distribution is not easily accessible. Thus, we implement a rejection sampler which updates each knot once at a time while conditioning on the values assumed by the other knots. As every knot is bounded, a uniform distribution can be used as proposal. Before accepting the knots, Condition (1.10) must be respected. Details are given in Appendix A2. Lastly, we include the optional Step [2.e] to learn the Dirichlet Process concentration parameters from the data by exploiting the data augmentation strategy of Escobar and West (1995) with $\alpha_j \sim \text{Gamma}(a_j, b_j)$.

When we fit the extended model $f^{ext}(\cdot|\cdot)$, we need just an additional step to sample from the posterior distribution of η_s and η_b . Because of the conjugacy between the Gamma and the Pareto distributions, the simulation is straightforward.

1.4 Numerical experiments

In this section, we illustrate how the proposed signal extraction method works through some numerical experiments. We simulate artificial datasets in a spatial region of 200×200 pixels of size 0.05° across 25 energy levels. The energy bins are \log_{10} -equispaced with size 0.1 in the interval 0 and 2.5, spanning an energy range between 1 GeV and 316.2278 GeV . Let S be the number of sources in the artificial dataset, with directions (ψ_1, \dots, ψ_S) : we simulate the counts in pixel (u, v, z) drawing from the Poisson

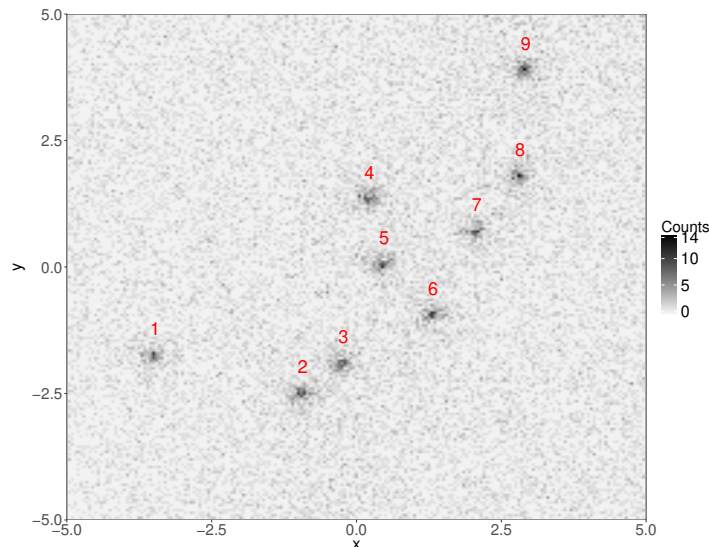


FIGURE 1.3: Map of 10,000 events obtained by the generating method described at the beginning of Section 1.4, using 9 equally bright sources and a moderate background contamination.

distribution

$$Y_{uvz} \sim Poi \left(\sum_{s=1}^S \Lambda_{uvz}^{\psi_s} + \Lambda_{uvz}^b \right), \quad \begin{aligned} u, v &= 1, \dots, 200, \\ z &= 1, \dots, 25, \end{aligned}$$

where $\Lambda_{uvz}^{\psi_s}$ is the expected number of photons from the source located in ψ_s and Λ_{ijk}^b are the expected counts from the background. In order to conduct realistic experiments, we draw the signal of a source using the *Fermi* LAT PSF and a power-law spectral model

$$\Lambda_{uvz}^{\psi_s} = PSF(\psi_s)_{uvz} \cdot F_{0,s} \left(\frac{E_k}{E_0} \right)^{-\varrho_s}, \quad (1.13)$$

where the parameters $F_{0,s}$ and ϱ_s control the brightness of the source, E_0 is the minimum level of energy observed (here $E_0 = 1\text{GeV}$) and E_k denotes the centroid of the k -th energy bin. $PSF(\psi_s)_{uvz}$ is the probability that a photon with energy l is recorded in the pixel jk . The background expectation Λ_{uvz}^b is given by [Acero et al. \(2016\)](#).

1.4.1 An illustrative example

We first propose an illustrative example based on 9 equally bright sources with the same spectral parameters $F_{0,s} = 1 \cdot 10^{-9}$ and $\varrho_s = 2$, for $s = 1, \dots, 9$, and a moderate background contamination. We select a dataset of 10,000 photons at random from the initial 25,140 generated by the simulation to reduce computational time. The reduced map of events, that we call \mathcal{D}_X , is shown in Figure 1.3; the sources are labelled from 1 to 9. The aim of this first part is to illustrate the functioning of the model and how

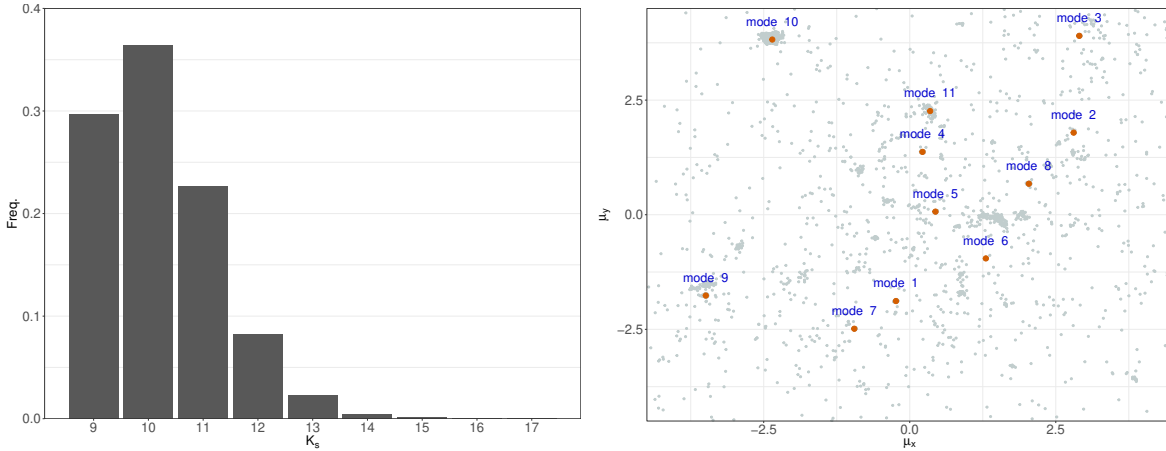


FIGURE 1.4: Left: posterior distribution of the number of active components of $s(\cdot|\cdot)$ across the 10,000 iterations, after a burn-in period of 20,000. The modal value is 10, with 3644 visit times. Right: draws from the posterior distribution of $\boldsymbol{\mu}$ when the number of components is $k_s = 10$. Using Algorithm 2, we distinguish 11 modal values to use as reference points.

to perform a post-processing analysis on the draws from the posterior distribution of the model parameters. For this reason, we fit only the spatial model $f(\cdot|\cdot)$ of Formula (1.1), postponing the comparison of performance with the extended model $f^{ext}(\cdot|\cdot)$ to the next section.

We set the total number of iterations to 30,000 and use the first 20,000 as burn-in. Neal’s algorithm used in Steps [2.b] and [2.c] of our MCMC sampler requires the additional parameters h_s and h_b : we set $h_s = h_b = 5$ which guarantees a fast exploration of the support of the posterior distribution without increasing computational time. Last, the concentration parameters α_s and α_b are learned from the data using Step [2.e] of Algorithm 1. We set $\alpha_s \sim \text{Gamma}(9, 3)$ and $\alpha_b \sim \text{Gamma}(4, 2)$ to have the expectation of α_s to be larger than the one of α_b ; so, $\mathbb{E}(\alpha_s) = 3$ and $\mathbb{E}(\alpha_b) = 2$, while the variance is one for both the parameters. By doing so, the a priori number of clusters induced by the Dirichlet Process priors, taking $\lambda = 1$, is $\mathbb{E}(k_s) \approx 21.18$ and $\mathbb{E}(k_b) \approx 14.8$.

The left plot of Figure 1.4 shows the posterior distribution of the number of components in the mixture model of the sources k_s , across the 10,000 iterations: with more than 36.44% visits, 10 is the most visited size, and almost 89% of the times the mixture size is between 9 and 11. The trace plots of the draws from the posterior distributions of $(\mu_{x,1}, \dots, \mu_{x,10})$ and $(\mu_{y,1}, \dots, \mu_{y,10})$ are displayed in the left column of Figure 1.5.

The algorithm achieves convergence for the parameters of the first nine components and the draws reveal unimodal distributions, while the posterior of $\boldsymbol{\mu}_{10}$ is prevalently multimodal. This effect arises when the MCMC sampler repeatedly adds to and removes a mixture component from the model, due to the small number of corresponding

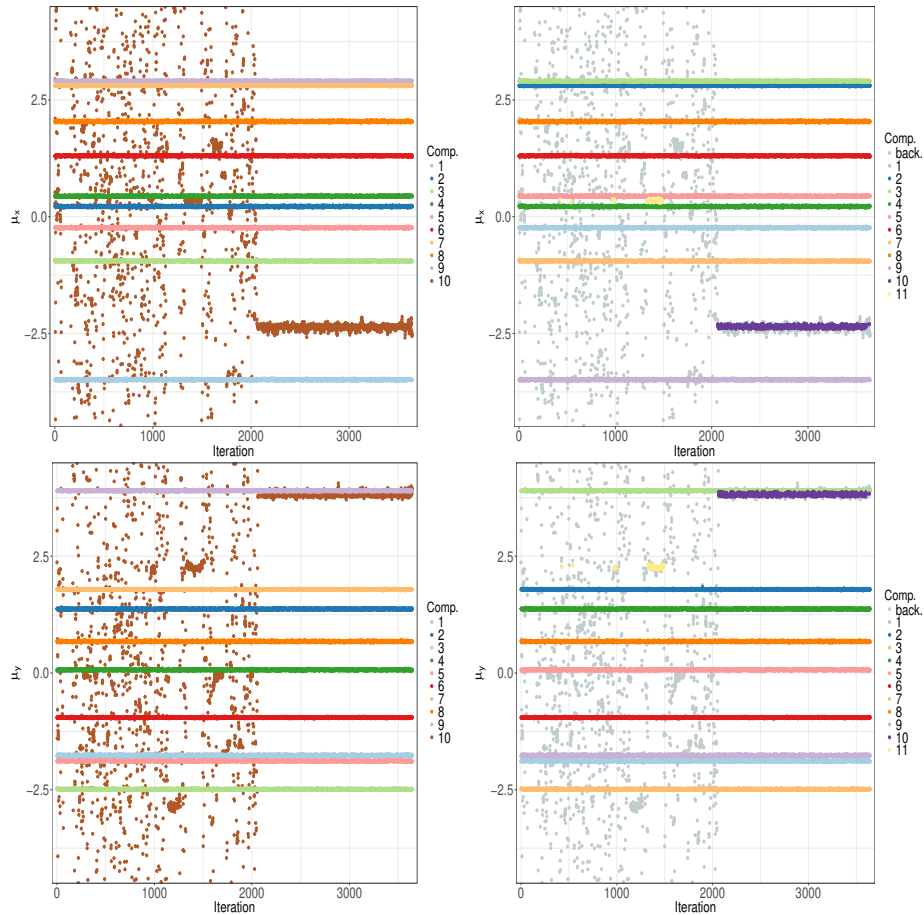


FIGURE 1.5: Left column: trace plots of μ_x (top) and μ_y (bottom) when the number of active components is 10. Right column: relabelled trace plots from μ_x (top) and μ_y (bottom). We distinguish two new components, labelled as 10 and 11, that previously belonged to the 10th component.

observations. The sparse groups of photons which make the algorithm adding the tenth component are located at very different points of the map, causing the posterior distribution of μ_{10} to be multimodal. The multiple modes can then either be sparse signals from the background which are not captured by $b(\cdot)$, or alternatively they represent the signal of faint sources made by very few photons.

Following Malsiner-Walli *et al.* (2016), we develop a post-processing algorithm to relabel the values sampled from $\mu|\mathcal{D}_{\mathcal{X}}$ so as to avoid multimodal posterior distributions, as it is the case for the 10th component in the left column of Figure 1.5. The pseudo-code is illustrated in Code-Box 2.

During Step [2] of Algorithm 2, we fit the kernel density estimator of Wand (1994) over the draws from the posterior distribution of $\mu|\mathcal{D}_{\mathcal{X}}$ and we select the modal values to perform the relabelling. The kernel estimator works on a grid that must be carefully chosen to highlight the modes of the posterior distribution of μ while avoiding overfitting; we chose here an equally spaced grid of 150 bins on both dimensions. The number

Algorithm 2: Relabelling algorithm

Input: $k_s^{(t)}$ and $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k_s})^{(t)}$, for all the iterations t after the burn-in period.

Let $\tilde{d} = 0.06$ be a given threshold, that corresponds to the radius of the 99% credibility interval of $\boldsymbol{\mu}$, when the data are known to come from a source.

begin

Step [1] Let k_s^+ be the modal value inside $k_s^{(1)}, \dots, k_s^{(T)}$; for $r = 1, 2, \dots$ such that $k_s^{(r)} = k_s^+$, we consider $\boldsymbol{\mu}_+^{(r)} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k_s})^{(r)}$.

Step [2] As the density of $\boldsymbol{\mu}|\mathcal{D}_X$ is multimodal, run a bivariate nonparametric density estimator over $(\boldsymbol{\mu}_+^{(1)}, \boldsymbol{\mu}_+^{(2)}, \dots)$ as the one proposed by Wand (1994) based on bivariate Gaussian kernels. Sort the modes in decreasing order, starting from the highest, and stop when the density value is smaller than a given threshold ξ . Let $(\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_P^*)$ be the P directions detected.

Step [3] For $r = 1, 2, \dots$ and $j = 1, \dots, k_s^+$, compute

$$d_j^{(r)} = \min_{p=1, \dots, P} \|\boldsymbol{\mu}_j^{(r)} - \boldsymbol{\mu}_p^*\|$$

and if $d_j^{(r)} < \tilde{d}$ then label $\boldsymbol{\mu}_j^{(r)}$ as p , otherwise call it *back*. Relabel as p also the observations assigned to the j -th component of the source mixture $s(\cdot|\cdot)$ during the r -th iteration of the MCMC algorithm.

of modes P to use as reference points is chosen according to an heuristic threshold ξ : any mode whose density value is larger than ξ is taken as a reference points for the relabelling procedure, otherwise is discarded. Although one can take an arbitrary small value of ξ , a too small threshold augments the risk of using spurious modes as reference point for the relabelling. Here we set $\xi = 0.1$.

In the right plot of Figure 1.4, we show the draws from the posterior distribution of $\boldsymbol{\mu}$ when the size of the mixture $s(\cdot|\cdot)$ is equal to 10; we furthermore choose to discard the values of the corners and close to the edges, restricting our attention on the region $(-4.5^\circ, 4.5^\circ) \times (-4.5^\circ, 4.5^\circ)$. Our post-processing algorithm selects $P = 11$ modal values, which are marked as orange points in the right plot of Figure 1.4. Finally, Step [3] relabels the draws according to their distance to the P reference points. The quantity $d_j^{(r)}$ represents the distance of $\boldsymbol{\mu}_j^{(r)}$, the location parameter of the j -th component drawn at the r -th iteration, to the closest reference point. If $d_j^{(r)}$ is larger than a threshold \tilde{d} , then it is marked with the label of that reference point, otherwise it is believed to be generated by a spurious cluster of photons that was wrongly accounted for by the source mixture $s(\cdot|\cdot)$, and for this reason it is relabelled as *back*.

The right column of Figure 1.5 shows the draws from $\boldsymbol{\mu}|\mathcal{D}_X$ after Algorithm 2 was applied. The relabelling procedure has highlighted two new distinct components that were previously part of a unique one, the tenth. The rest of the draws from the tenth

TABLE 1.1: Details of the results given by the relabelling procedure. From left to right, the columns display: (1) the cluster labels, (2) the source IDs when the sources are within 0.1° from the estimated direction, (3) the true direction of the source (4) the expected number of photons from the estimated cluster, (5) the 90% HPD interval of the cluster size, (6) the true number of photons from the source and (7) the estimated probability that the clusters correspond to real sources.

clusterID	sourceID	$\mathbb{E}(\boldsymbol{\mu} \dots)$	$\boldsymbol{\psi}$	$\mathbb{E}(N \dots)$	HPD _{90%}	#Counts	$\mathbb{P}(\text{source} \dots)$
1	3	(-0.237, -1.882)	(-0.225, -1.875)	144.68	(126, 163)	122	1.000
2	8	(2.817, 1.787)	(2.825, 1.825)	133.24	(117, 150)	130	1.000
3	9	(2.89, 3.907)	(2.925, 3.925)	133.03	(117, 149)	125	1.000
4	4	(0.2, 1.379)	(0.225, 1.375)	180.16	(162, 199)	142	1.000
5	5	(0.418, 0.075)	(0.475, 0.075)	175.49	(156, 195)	123	1.000
6	6	(1.29, -0.985)	(1.325, -0.925)	143.43	(126, 161)	142	0.999
7	2	(-0.964, -2.453)	(-0.925, -2.475)	165.51	(146, 186)	139	0.999
8	7	(2.017, 0.646)	(2.075, 0.725)	121.43	(105, 139)	131	0.979
9	1	(-3.508, -1.801)	(-3.475, -1.725)	116.73	(100, 133)	128	0.944
10	//	(-2.345, 3.826)	//	31.02	(20, 44)	//	0.291
11	//	(0.345, 2.276)	//	22.64	(9, 37)	//	0.041

component are remarked as *back*, as they cannot be associated with any modal value.

Using Algorithm 2, we can also perform the second level of clustering in the mixture $s(\cdot|\cdot)$. Since for any $\boldsymbol{\mu}_j^{(r)}$ there exists at least one photon assigned to the j -th component of $s(\cdot|\cdot)$ during the r -th iteration, the procedure changes the labels of both $\boldsymbol{\mu}_j^{(r)}$ and the relative photons, forming new clusters which should represent the astronomical sources in the map. We can quantify their intensities by counting the photons assigned to them. In addition, at some iterations of the MCMC sampler, a cluster might become empty; the number of times the cluster has at least one observation over the length of the Markov chain gives an estimate of the probability that the cluster truly coincide with a real astronomical source.

Table 1.1 summarises the results for the 11 clusters which were identified by the algorithm. The first two columns show respectively the cluster ID and the corresponding source ID in case that the estimated direction of the source (Column 3) is 0.1° away from the true direction (Column 4). Columns 5 and 6 summarise the posterior distribution of the counts in each cluster, and can be compared with the seventh column, which shows the true number of photons from the sources. Finally, the last column gives an estimate of the probability that the discovered clusters coincide with the existent sources.

The model precisely locates all 9 true sources with very high posterior probability. Moreover, we select two additional spurious clusters with small expected number of counts and small estimated probabilities, which do not correspond to any source.

TABLE 1.2: Results obtained fitting the proposed models on the three simulated datasets. Each row displays, for a given scenario (first column), the number of true and spurious clusters with the type of model in the second column. For ease of presentation, we distinguish the results according to the estimated probabilities of being a real source, and we group them into three equally large intervals called respectively low [0 - 0.333], medium [0.334 - 0.666] and large [0.667 - 1]. For every interval, we report the number of sources discovered (denoted as true) and the spurious clusters (denoted as noise).

Scen.	#Sources	Model	low [0 - 0.333]		medium [0.334 - 0.666]		large [0.667 - 1]	
			true	noise	true	noise	true	noise
1	15	Spatial	0	0	0	0	15	0
1		Extended	0	1	1	0	14	0
2	15	Spatial	0	1	0	0	15	0
2		Extended	0	2	0	0	15	0
3	25	Spatial	4	1	1	0	18	0
3		Extended	0	0	2	1	20	0

1.4.2 Model comparison

We carried out a number of numerical experiments to evaluate the performance of the models $f(\cdot|\cdot)$ and $f^{ext}(\cdot|\cdot)$. The goal is to determine whether the presence of the additional energy variable impacts on cluster recognition and on background separation. We consider three different artificial datasets which were generated with the same background region used for the above illustrative example, but which largely differ in the number of sources and in their spectral parameters. The first scenario considers a map with 15 equally bright sources ($F_{0,s} = 1 \cdot 10^{-9}$ and $\varrho_s = 2$ for $s = 1, \dots, 9$). The second scenario is made by 15 sources located of the same directions of the previous scenario, but with different spectral parameters. The last scenario is made by 25 sources with different spectral parameters; in addition, it includes two pairs of overlapping sources. In the first pair, the source with ID 1 is 0.07° from the one with ID 2, and in the second pair the source ID 22 is 0.11° from ID 23. The characteristics of the simulated sources used in the second and in the third scenarios were randomly chosen from the catalogue of [Acero et al. \(2015\)](#). Every dataset consists of Poisson counts as described at the beginning of Section 1.4. The size of each simulated dataset is reduced to 10,000 records to reduce computational cost. We run the MCMC sampler 20,000 times to fit both $f(\cdot|\cdot)$ and $f^{ext}(\cdot|\cdot)$, and we used half of the iterations as burn-in. The relabelling procedure in Code-Box 2 is then applied to the draws from the posterior distributions with the same set-up illustrated in Section 1.4.1. The results are summarised in Table 1.2. Further details on the simulations are given in Appendix B.

The three simulations evaluate the two models in terms of the total number of sources

discovered and the number of spurious clusters located. In addition, for every mixture component which effectively locates a source, we say that the signal is correctly extracted if the 90% HPD interval of the cluster size includes the true number of photons from that source.

In the first numerical example, the spatial model clearly locates the 15 sources without any spurious cluster and correctly extracts 13 of them. The extended model performs slightly worse in terms of signal recognition. It locates 14 sources in the high probability range and 1 in the middle range. It also pinpoints a false positive with very low probability (0.03), but correctly extract the signal of 14 sources, one more than the model $f(\cdot|\cdot)$ in (1.1). For the second scenario, both models perform remarkably well, discovering all 15 sources with large posterior probabilities and few spurious clusters. In addition, the extended model $f^{ext}(\cdot|\cdot)$ correctly extracts 10 of the 15 sources, against the 8 of the baseline model $f(\cdot|\cdot)$. Regarding the final scenario, which is the most complex one, both models locate most of the sources in the high probability range, although some aspects must be considered. Model $f(\cdot|\cdot)$ discovers 4 sources in the low probability range and 1 in the middle range, showing in addition a spurious cluster. Model $f^{ext}(\cdot|\cdot)$ shows instead higher posterior probabilities, but at the same time it completely misses the 21-th source and locates a spurious cluster with probability in the middle range. Both models do not clearly disentangle the two pairs of overlapping sources, but reveal a unique component in the direction of interest. The sources of the first pair (IDs 1 and 2) emit respectively 238 and 157 photons, while the second (IDs 22 and 23) 229 and 30. The component of $f(\cdot|\cdot)$ referred to the first pair exhibits around 389 expected counts, against the 369 expected from $f^{ext}(\cdot|\cdot)$. In the same way, the component from the spatial model which detect the second pair has 229 expected counts, against the 240 expected from $f^{ext}(\cdot|\cdot)$. Both models fail in distinguishing the two pairs of overlapping sources, but correctly extract their total intensities. In total, both $f(\cdot|\cdot)$ and $f^{ext}(\cdot|\cdot)$ correctly extract the signal of 18 celestial objects.

In conclusion, the two models revealed high quality performances in locating and extracting the signal of the sources over the three simulated maps of photons. In the first scenario, the spatial model detects 100% of the sources in high posterior probability range, against a 93% of the extended model. In the second scenario, both models correctly locate all the sources and show just few spurious clusters. Finally, in the last scenario the two models discover 72% and 80% of the sources, respectively; in addition, these rates become 76% and 88% if we consider also the middle probability range.

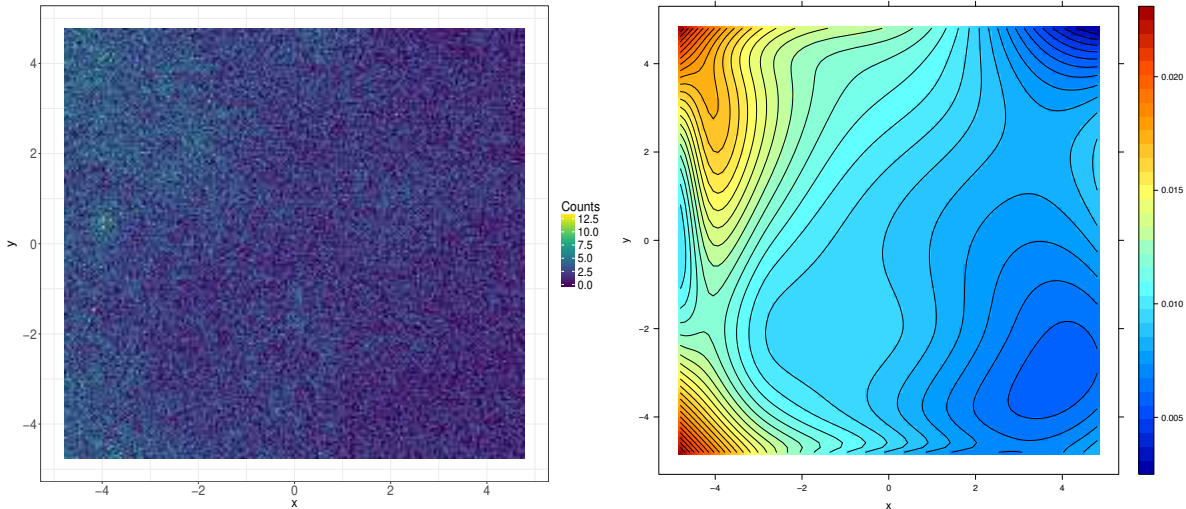


FIGURE 1.6: Left: $10^\circ \times 10^\circ$ region from the Fermi LAT telescope. The available dataset based on 9 years of observation is composed by 98,598 events, binned into a 200×200 grid. The scale colour value varies with respect to the number of counts per pixel observed. Right: contour plot of the estimated background density using the spatial-only model, based on 20,000 records sampled at random from the starting dataset.

1.5 Application to the *Fermi* LAT data

We finally use the proposed models to analyse a γ -ray dataset collected by the Large Area Telescope (LAT) onboard the *Fermi* telescope. The data refer to the set of events detected within a region surrounding the newly discovered dwarf galaxy *Antlia 2*. This region is relatively close to the galactic plane so there is a large contribution from diffuse background processes which have strong gradients across the field of view. A first background model developed by the *Fermi* Collaboration [Acero *et al.* 2016] is made up of an isotropic component plus a diffuse component which represents galactic processes. The physical processes which give rise to these diffuse emissions are very difficult to model. We do not expect the *Fermi* background model to capture all the detailed morphological and spectral characteristics of the true background. Therefore, a nonparametric method as the one we propose would be more suitable for estimating the noise component of the data.

We consider events from a squared region of 200×200 spatial bins and 30 \log_{10} -equispaced energy bins spanning the range $0.5 \text{ GeV} - 500 \text{ GeV}$.

The 98,598 detected counts are displayed in the left plot of Figure 1.6: each point in the map represents a spatial pixel, coloured according to the number of photons which fall into it. A noise component extends rather evidently over the whole map which is more prevalent at the left side of the plot and slowly fades out moving towards the right

TABLE 1.3: Results for the *Fermi* LAT data. The estimated directions of the clusters are shown in the second and in the fourth columns, while the third and the fifth display the estimated posterior probabilities of being a source. The first cluster receives a non negligible posterior evidence from both models. Its closest source is 0.93° away, but its presence is not supported by our models. Cluster IDs 2 and 4 are less than 0.1° away from two sources recorded in the third *Fermi* LAT catalogue. The two models diverge about the direction of third cluster, as the estimated points are 0.57° away. Finally, the fifth cluster is selected only by the spatial model with very low posterior probability.

ClusterID	<i>Spatial</i>		<i>Spatial and Spectral</i>	
	$\mathbb{E}(\boldsymbol{\mu} \dots)$	$\mathbb{P}(\text{source} \dots)$	$\mathbb{E}(\boldsymbol{\mu} \dots)$	$\mathbb{P}(\text{source} \dots)$
1	(1.86, 0.90)	0.79	(1.86, 0.90)	0.47
2	(-3.94, 0.47)	0.62	(-3.94, 0.47)	0.91
3	(-2.10, 1.67)	0.27	(-1.59, 1.85)	0.20
4	(0.04, -1.79)	0.27	(0.05, -1.79)	0.68
5	(-2.02, 2.39)	0.07	//	//

side.

Signal extraction is performed using both the spatial-only model and its spectral extension to evaluate the accordance of the results. The settings of the algorithm are the same as in the previous numerical examples. We furthermore reduce the dataset size to 20,000 records to streamline computational cost.

The estimate of background morphology given by the spatial model is illustrated using the contour plot on the right side of Figure 1.6. It clearly shows: i) a prominent surface which extends over the left part of the image, and which is more pronounced at the left-top and left-bottom edges, ii) a moderate contamination in the middle part of the image, and iii) a progressive decrease of the noise intensity moving forward towards to the x-axis.

Once the background contamination is filtered appropriately, we proceed with signal extraction. Table 1.3 displays the results obtained from the two models after the post-processing algorithm was applied. The two models select respectively 5 and 4 clusters. The first is located in the same direction from both models and is supported to be a source by a large estimated probability. However, the closest source in the third *Fermi* LAT catalogue is 0.93° away and both models do not show any evidence of that. We thus believe that further investigation about the true location of this source are to be considered. The second and the fourth clusters are less than 0.1° from two sources discovered by the *Fermi* collaboration; the spatial and spectral model is more accurate in terms of posterior evidence than the only spatial one. The two models are discordant with respect to the direction of the third cluster, as the first points 0.57° away from the other. In addition, the two reported probabilities fall into the low

range as they are less than 0.33. We hence suppose that the third cluster is currently locating a sky region contaminated by a prominent background component, whose signal is occasionally captured by the source mixture model. However, the discovered signal could also originate from a pair or a group of faint neighboring sources. In this latter case, a statistical analysis using data from lower energy ranges would be more suitable to reveal their presence. Finally, the fifth cluster is located only by the spatial model and exhibits a very poor posterior evidence. We are confident to conclude that no source is located at this direction.

1.6 Discussion

We introduced in this chapter an innovative approach to signal extraction of astronomical sources in highly contaminated γ -ray count maps using the spatial and spectral information of the events. The method exploits advanced Bayesian nonparametric techniques to detect the directions of the sources and, at the same time, to fit the underlying background component. A constraint on the parameter space is imposed to reduce the risk of absorption of the signals of the sources by the background. In addition, we developed a post-processing algorithm to solve the well-known label switching effect of mixture models and to extrapolate the relevant information when the posterior distribution of some location parameters exhibits multiple modes. We furthermore illustrated how the model works through an illustrative example.

We first introduced a model that exploits only the spatial information of the events; we then extended it by considering also the photon energy. We thus carried out some numerical experiments to test the detection performance of our method. We simulated the datasets using the *Fermi* LAT *Point Spread Function* and a power law spectral model in the range of energies $1 \text{ GeV} - 316.2278 \text{ GeV}$; the spectral parameters of the sources were chosen at random from the third *Fermi* LAT catalogue to include both bright and faint sources. The results of the numerical experiments are very promising. Both models successfully locate most of the sources with high posterior evidence and correctly extract their signals. The two models exhibit similar detection performances and reveal only few spurious clusters which can be easily recognised using the post-processing algorithm proposed. However, it emerges that the extended model $f^{ext}(\cdot|\cdot)$ helps in quantifying the source intensity, but at the same time it slightly increases the number of *Type I* errors. Finally, the models were applied over a region of the sky surrounding the dwarf galaxy *Antlia 2*, observed by the LAT component of the *Fermi* telescope, and produced concordant results.

With respect to the reversible jump approach of Jones *et al.* (2015), the Dirichlet Process mixture has both conceptual and computational advantages. Considering that the number of components in Bayesian nonparametric mixtures is potentially infinite, a larger number of records usually implies a larger size of the mixture. Thus, a model based on a Dirichlet Process is more appropriate than a finite mixture given data, with astronomical data, new observations become constantly available and may guide to the identification of new sources. Moreover, the available algorithms for Bayesian nonparametric models are more efficient for posterior inference and scale the mixture size faster than the reversible jump algorithm.

Further developments will concentrate on carefully including the spectral information into the Dirichlet Process models, with particular attention to the background component, and on developing a formal disentangling procedure for reconstructing the signal of overlapping sources.

Chapter 2

Continuous time hidden Markov models for flare detection in gamma-ray light curves

“I watched C-beams glitter in the dark near the Tannhäuser Gate.”

Rutger Hauer - *Blade Runner*

2.1 Introduction

The statistical analysis of time varying astronomical sources is an interdisciplinary field which combines both astronomical and statistical methods to investigate the physical mechanisms that characterise celestial objects. This type of analysis works with photons which fall on the detector surface of a telescope during the monitoring activity of a source in time. Then, a common practice is to convert counts into flux to standardise the observations with respect to some features, as the exposure time of the instrument to the portion of the sky where they come from and the size of the instrument’s surface, even known as effective area. The flux time series that describes the emission activity of a source as a function of time is called *light curve*.

Up to now, several authors modelled the light curves of astronomical sources either by adopting a single stochastic process [Kelly *et al.* 2009, MacLeod *et al.* 2010, Zu *et al.* 2013], with particular emphasis on optical data, or by their extensions which linearly combine multiple stochastic processes [Kelly *et al.* 2011, Sobolewska *et al.* 2014]. In addition, Dodds-Eden *et al.* (2011) and Witzel *et al.* (2012) accounted also for low variations in the signal due to the instrumental error with a Gaussian distribution.

Many works in the literature focus on the analysis of optical and X-ray light curves, that generally present moderate variations in the flux [Uttley *et al.* 2005, Sobolewska and Papadakis 2009, Kelly *et al.* 2011]. However, other types of light curves are highly variable and show evidence in favour of multiple states. This is the case, of example, of the Galactic black hole associated to the source *Sagittarius A** (Sgr A*), whose emission activity was previously modelled with a unique stochastic process [Witzel *et al.* 2012], and more recently Meyer *et al.* (2014) proposed a two latent states model, in accord with the conclusions of Dodds-Eden *et al.* (2011). Lots of phenomena in the universe cover a wide range of energies in the electromagnetic spectrum up to the γ -rays, the most energetic type of radiation in the universe. However, γ -ray light curves are known also to be highly complex. For example, Ramakrishnan *et al.* (2015) show the presence of abrupt spikes in the γ -ray light curves of several blazars, monitored by the *Fermi* LAT. In addition, Kushwaha *et al.* (2017) pass beyond the use of Gaussian distributions to model the flux of blazars and Active Galactic Nuclei (AGNs) [Sobolewska *et al.* 2014], and showed that skewed distributions are preferable as they successfully account for the jets in the emission activity. Thus, establish when such flares occur is a key task to comprehend the physical mechanisms at the base of high-energy sources, but a formal procedure to learn these mechanisms using γ -ray data has not been provided yet.

We introduce in this chapter a new formal statistical technique to distinguish the multiple states of variability in γ -ray light curves. The statistical method we propose finds its roots in the class of latent states models. In particular, we exploit the well-known hidden Markov models (HMMs), that have been largely studied in the statistical literature [Zucchini *et al.* 2016] and extensively applied in many areas, such as engineering [Wei *et al.* 2002, de Gunst and Schouten 2003, Lee *et al.* 2004, Roberts and Ephraim 2008], genomics [Krogh *et al.* 2001, Schliep *et al.* 2003, Hobolth and Stone 2009, Städler and Mukherjee 2013, Cui *et al.* 2015] and finance [Genon-Catalot *et al.* 2000, Mamon and Elliott 2014]. The Markovian assumption claims that the current value of the process depends solely on the value in the previous step. Although in some cases this assumption is not properly respected and requires to consider further extensions [Barbu and Limnios 2009], it leads to versatile and interpretable statistical models, and computational advantages in the estimation step.

This chapter aims to substantially improve the statistical toolbox available to classify the nature of high-energy astronomical sources. Among the challenges to overcome, the presence of multiple latent states of variability and of abrupt increments of brightness, and the high dependence between subsequent observations have a central role. The contribution given by the chapter is organised in four sections as follows: Section 2.2

extends the methods already exploited in literature and defines a new model that accounts for different levels of variability in γ -ray light curves, Section 2.3 outlines the estimation algorithm and a bootstrap procedure to assess the properties of the model, Section 2.4 shows a case-study using the γ -ray light curve from blazar PKS 1510-05, and Section 2.5 concludes by discussing the main contributions of the chapter and further extensions that can be explored.

2.2 Hidden Markov modelling approach

2.2.1 Continuous time observations

We consider a collection of n observations $\mathcal{Y} = \{y_{t_i}\}_{i=1}^n$ representing the flux of an astronomical source over a discrete sequence of observation times $(t_1, \dots, t_n) \in (T_{\min}, T_{\max})$, where $T_{\max} - T_{\min}$ is the entire observation period. The emission activity of a source is constantly monitored and thus, after transforming the photon counts into flux, the observations should be available at constant time bins. In practice, data usually show big gaps in the observation times due to the limited telescope time, the night/day cycles and instrument failures which might not make the records available at some points. So, $\Delta_i = t_i - t_{i-1}$ is not constant.

This issue led Kelly *et al.* (2009) to consider a more appropriate model for continuous time observations. A generic process $\{y_t\}_t$ is said to follow an *Ornstein-Uhlenbeck* (OU) process if its dynamic can be described by the stochastic differential equation $\partial y_t = \tau(\mu - y_t)\partial t + \sigma\partial Z_t$, where ∂Z_t is the increment of a standard Brownian motion with $Z_t \sim \mathcal{N}(0, 1)$, μ is a real parameter which represents the mean, σ is the volatility and τ is the speed of mean reversion [Ornstein and Uhlenbeck 1930]. The solution to the differential equation above leads $y_{t_{i-1}+\Delta_i}|y_{t_{i-1}}$ to be Gaussian with mean and variance equal to

$$\mu_i^* = y_{t_{i-1}}e^{-\tau\Delta_i} + \mu(1 - e^{-\tau\Delta_i}), \quad \sigma_i^{2*} = \frac{\sigma^2(1 - e^{-2\tau\Delta_i})}{2\tau}, \quad (2.1)$$

for every observation time t_i and for every arbitrary time interval Δ_i . This modelling approach has the advantage of directly solving the time gaps issue described above without performing any data imputation. Taking $\Delta_i \rightarrow \infty$ we obtain $\mu_i^* = \mu$ and $\sigma_i^{2*} = \sigma^2/2\tau$, the parameters of the limit distribution of the process and also of the first observation y_{t_1} .

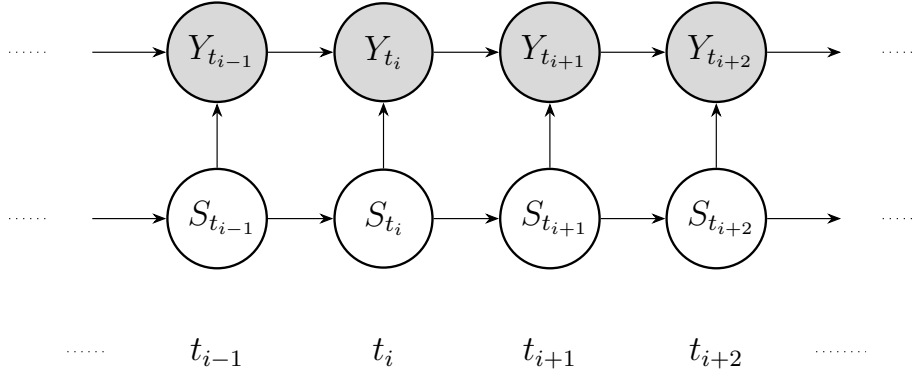


FIGURE 2.1: DAG of the continuous time hidden Markov model. Grey circles are the data and white circles represent the latent Markov process.

2.2.2 Multiple states modelling

The recent analyses of [Sobolewska *et al.* \(2014\)](#), [Ramakrishnan *et al.* \(2015\)](#) and [Kushwaha *et al.* \(2017\)](#) has brought out that γ -ray light curves from blazars and AGNs are characterised by distinct levels of variability and large peaks in the brightness, suggesting the presence of multiple heterogeneous mechanisms that occur at the base of the source. So, an ideal statistical model to fit these kinds of light curves must account also for the presence of different emission phases. The only attempt in the astronomy literature to use a state-space model was made by [Meyer *et al.* \(2014\)](#), who proposed a two states hidden Markov model (HMM) to separate the measurement error from the true signal of the source Sgr A*, fitting an X-ray light curve. The transition between the source dominated and the noise dominated states was governed by a discrete time Markov chain.

We propose in this section to model the signal of time varying γ -ray sources with a multistate continuous time hidden Markov model [[Zucchini *et al.* 2016](#), Chapter 11]. We adopt an OU process for modelling the flux measurements \mathcal{Y} [[Kelly *et al.* 2009](#), [Meyer *et al.* 2014](#)], but we consider also a latent continuous time Markov process $S_t \in \mathcal{S} = \{1, \dots, \mathcal{S}\}$ with initial probability vector $\boldsymbol{\delta}$ and generator matrix $\mathbf{Q} = \{q_{ij}, i, j \in \mathcal{S}\}$, with $q_{ij} \geq 0$ when $i \neq j$, and $q_{ii} = -\sum_{j \neq i} q_{ij}$. Thus, conditioning to the current value of the latent state, for any time gap Δ_i the statistical model is $y_{t_{i-1}+\Delta_i} | y_{t_{i-1}}, S_{t_{i-1}+\Delta_i} = s \sim \mathcal{N}(\mu_{i,s}^*, \sigma_{i,s}^{2*})$, where

$$\mu_{i,s}^* = y_{t_{i-1}} e^{-\tau_s \Delta_i} + \mu_s (1 - e^{-\tau_s \Delta_i}), \quad \sigma_{i,s}^{2*} = \frac{\sigma_s^2 (1 - e^{-2\tau_s \Delta_i})}{2\tau_s}.$$

A graphical representation of the model is given in Figure 2.1 using a Directed Acyclic Graph (DAG), which describes the dependency relation among the involved random

Algorithm 3: EM algorithm for continuous time HMMs and OU processes

Input: $(\hat{\Theta}, \hat{\mathbf{Q}}, \hat{\boldsymbol{\delta}})^{(0)}$, choose an arbitrary threshold ϵ and set $\Delta^{(0)} > \epsilon$, $r = 0$, $\ell^{(0)} = -\infty$ and Δ_t arbitrary large.

while $\Delta^{(r)} > \epsilon$ **do**

1. $r = r + 1$;

— *E-step* —

2. For $i = 1, \dots, n$, compute $\mathbf{c}^{(r)}$ and $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)^{(r)}$ using (2.4).

3. Decompose $\mathbf{Q}^{(r-1)} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$. For all $s, s' \in \mathcal{S}$, compute $\Lambda(s, s')$ and store $\hat{m}_{s,s'}^{(r)}$.

4. $\hat{T}_s^{(r)} = \hat{m}_{s,s}^{(r)} / \hat{q}_{s,s}^{(r-1)}$, for all $s \in \mathcal{S}$.

5. For $i = 1, \dots, n$, store the n -size vector $\hat{\boldsymbol{\xi}}_i^{(r)} = \tilde{\boldsymbol{\alpha}}_i^{(r)} \odot \tilde{\boldsymbol{\beta}}_i^{(r)}$, where \odot is the term-by-term product.

— *M-step* —

6. For all $s, s' \in \mathcal{S}$ and $s \neq s'$, $\hat{q}_{s,s'}^{(r)} = \hat{m}_{s,s'}^{(r)} / \hat{T}_s^{(r)}$; then, $\hat{q}_{s,s}^{(r)} = -\sum_{s'=1}^{\mathcal{S}} \hat{q}_{s,s'}^{(r)}$.

Compute $\hat{\mathbf{P}} = \exp\{\hat{\mathbf{Q}}\Delta_t\}$ and set $\hat{\boldsymbol{\delta}}^{(r)}$ equal to the first row of $\hat{\mathbf{P}}$.

7. For all $s \in \mathcal{S}$, update $(\hat{\mu}_s, \hat{\sigma}_s^2)^{(r)}$ using (2.6) and $\hat{\tau}_s^{(r)}$ with a Newton-Rhapson step.

8. Let $\ell^{(r)} = \sum_{i=1}^n \log c_i^{(r)}$. Compute the distance $\Delta^{(r)} = \|\ell^{(r-1)}, \ell^{(r)}\|$.

Output: The parameter estimates $(\hat{\Theta}, \hat{\mathbf{Q}}, \hat{\boldsymbol{\delta}})^{(r)}$.

variables. We display the flux measurements of the source with grey circles, and the latent process S with white circles.

Let us denote with $\mathbf{f}(y_{t_i}; \boldsymbol{\Theta})$ the $\mathcal{S} \times \mathcal{S}$ diagonal matrix whose s -th diagonal element is $\phi(y_{t_i}; \mu_{i,s}^*, \sigma_{i,s}^{2*})$, ϕ is the density function of a Gaussian distribution, $\mathbf{1}$ is the unitary vector of length \mathcal{S} and $\boldsymbol{\Theta} = \cup_{s=1}^{\mathcal{S}} (\mu_s, \sigma_s^2, \tau_s)$. According to Zucchini *et al.* (2016), the likelihood function for the model parameters $(\boldsymbol{\Theta}, \mathbf{Q})$, assuming that $\boldsymbol{\delta}$ is known, is

$$\mathcal{L}(\boldsymbol{\Theta}, \mathbf{Q}) = \boldsymbol{\delta}^T \mathbf{f}(y_{t_1}; \boldsymbol{\Theta}) \prod_{i=2}^n \exp\{\mathbf{Q}\Delta_i\} \mathbf{f}(y_{t_i}; \boldsymbol{\Theta}) \mathbf{1}. \quad (2.2)$$

Since \mathbf{Q} is a matrix, $\exp\{\mathbf{Q}\}$ refers to the matrix exponential, that is defined as $\exp\{\mathbf{Q}\} := \sum_{k=0}^{\infty} \mathbf{Q}^k / k!$ [Norris 1998, page 62].

2.3 Model estimation

2.3.1 Parameters estimation via EM algorithm

We outline an efficient Expectation-Maximization (EM) algorithm for estimating $(\boldsymbol{\Theta}, \mathbf{Q})$. We do not focus on the estimation of the initial probability distribution $\boldsymbol{\delta}$ as it has negligible long term effects on the likelihood. A first version of the algorithm for estimating continuous time HMMs was proposed by Rydén (1996) for arrival counts data, and later

improved by Roberts *et al.* (2006) and Metzner *et al.* (2007). We show that the same algorithm states with slight modifications also for our model. Although the likelihood function of the model is given in Formula (2.2), the estimation procedure we exploit works on the complete representation of the likelihood that elicits both the observed time series \mathcal{Y} and the latent path of the process S . According to Albert (1962), Rydén (1996) and Bladt and Sørensen (2005), the complete log-likelihood function for the model parameters (Θ, \mathbf{Q}) is

$$\begin{aligned} \log \mathcal{L}_c(\Theta, \mathbf{Q}) \propto & \sum_{s=1}^S \mathbb{1}(S_0 = s) \log \delta_s - \sum_{s=1}^S T_s q_s + \sum_{s \neq s'} m_{s,s'} \log q_{s,s'} \\ & - \frac{1}{2} \sum_{s=1}^S \sum_{i=1}^n \xi_i(s) \left\{ \log \sigma_{i,s}^{2*} - \frac{(y_{t_i} - \mu_{i,s}^*)^2}{\sigma_{i,s}^{2*}} \right\}. \end{aligned} \quad (2.3)$$

With respect to (2.2), Formula (2.3) includes some quantities related to the latent Markov chain S : $m_{s,s'}$ is the number of times the process S jumps from a state s to s' , T_s is the total time spent in the s -th state, and $\xi_i(s)$ says whether S is in the state s or not at the observation time t_i . However, the mentioned quantities are unknown and cannot be directly quantified from the observed light curve. For this reason, the EM first takes the expectation of the unknown quantities, and then it maximizes the likelihood to get the parameter estimates. The estimation algorithm is an iterative two-steps procedure that must be run until the convergence is reached. In the next, we illustrate the two steps; in addition, a pseudo-code of the algorithm is given in Code-Box 3.

E-step:

The *E-step* of the algorithm provides an estimate of $m_{s,s'}$, T_s and $\xi_i(s)$ conditioning to the current values of $(\hat{\Theta}, \hat{\mathbf{Q}})$. Given $c_1 = \delta^T \mathbf{f}(y_{t_1}; \hat{\Theta}) \mathbf{1}$, $\alpha_1 = \delta^T \mathbf{f}(y_{t_1}; \hat{\Theta}) / c_1$ and $\beta_n = \mathbf{1}$, we first compute the forward densities $(\alpha_2, \dots, \alpha_n)$ and the backward densities $(\beta_1, \dots, \beta_{n-1})$ as

$$\begin{aligned} \alpha_i &= \frac{\alpha_{i-1} \exp\{\hat{\mathbf{Q}} \Delta_i\} \mathbf{f}(y_{t_i}; \hat{\Theta})}{c_i}, & i = 2, \dots, n, \\ \beta_i &= \frac{\exp\{\hat{\mathbf{Q}} \Delta_{i+1}\} \mathbf{f}(y_{t_{i+1}}; \hat{\Theta}) \beta_{i+1}}{c_{i+1}}, & i = 1, \dots, n-1, \end{aligned} \quad (2.4)$$

where $c_i = \alpha_{i-1} \exp\{\hat{\mathbf{Q}} \Delta_i\} \mathbf{f}(y_{t_i}; \hat{\Theta}) \mathbf{1}$. The quantities in Formula (2.4) differ by a rescaling factors (c_1, \dots, c_n) from the forward and backward densities of Rydén (1996), that

are widely known to be numerically unstable, as they tend to zero or to infinity exponentially fast with the sample size n [Leroux and Puterman 1992, Zucchini *et al.* 2016, Chapter 3]. For this reason, Roberts *et al.* (2006) proposed the above adjustment. It follows the estimate of $m_{s,s'}$ as

$$\hat{m}_{s,s'} = \hat{q}_{s,s'} \sum_{i=2}^n \frac{\alpha_{i-1} \Lambda_i(s, s') \beta_i}{c_i}, \quad (2.5)$$

where

$$\Lambda_i(s, s') = \int_{t_{i-1}}^{t_i} \exp\{\mathbf{Q}(t - t_{i-1})\} \mathbf{e}_s \mathbf{e}_{s'}^T \exp\{\mathbf{Q}(t_i - t)\} \mathbf{f}(y_{t_i}; \Theta) dt$$

and \mathbf{e}_s is a vector of length \mathcal{S} whose elements are 0 except for the s -th, which is 1. To solve the above integral of a matrix exponential, we first need to elicit each element of the integrand matrix. If \mathbf{Q} is diagonalisable, then we can write $\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, where \mathbf{V} is the matrix of eigenvectors and $\mathbf{D} = \text{diag}(d_1, \dots, d_{\mathcal{S}})$ is the diagonal matrix of eigenvalues. In this way, the integrand can be simplified by exploiting the relation $\exp\{\mathbf{Q}t\} = \mathbf{V} \exp\{\mathbf{D}t\} \mathbf{V}^{-1}$ and $\exp\{\mathbf{D}t\} = \text{diag}\{\exp(d_1 t), \dots, \exp(d_{\mathcal{S}} t)\}$. By substituting in the above formula, we derive that the pq -th element of Λ_i is equal to

$$\Lambda_{i,p,q}(s, s') = \sum_{u=1}^{\mathcal{S}} \sum_{v=1}^{\mathcal{S}} V_{pu} V_{us}^{-1} V_{s'v} V_{vq}^{-1} \phi(y_{t_i}; \hat{\mu}_{i,s'}^*, \hat{\sigma}_{i,s'}^2) \mathcal{J}_i(d_u, d_v),$$

where

$$\mathcal{J}_i(d_u, d_v) = e^{d_v t_i - d_u t_{i-1}} \int_{t_{i-1}}^{t_i} e^{t(d_u - d_v)} dt = \begin{cases} \Delta_i e^{d_u \Delta_i} & \text{if } d_u = d_v, \\ \frac{e^{d_u \Delta_i} - e^{d_v \Delta_i}}{d_u - d_v} & \text{if } d_u \neq d_v. \end{cases}$$

As an alternative, the evaluation of the above integral of matrix exponential can be replaced by an evaluation of a matrix exponential of higher order [Roberts and Ephraim 2008, Formulas (28)-(31)].

The estimation of the remaining quantities directly follows from the estimation of $m_{s,s'}$: in particular, $\hat{T}_s = \hat{m}_{s,s} / \hat{q}_{s,s}$ and $\hat{\xi}_i(s) = \mathbb{E}(\xi_i(s) | \Theta, \mathbf{Q}, \mathcal{Y}) = \alpha_{i,s} \beta_{i,s}$.

M-step:

The second step updates the estimates of the model parameters (Θ, \mathbf{Q}) using the quantities given by the *E-step*. The maximum likelihood estimator of the elements of \mathbf{Q} outside of the diagonal is

$$\hat{q}_{s,s'} = \frac{\hat{m}_{s,s'}}{\hat{T}_s},$$

Algorithm 4: Bootstrap for continuous time HMMs and OU processes

Input: The maximum likelihood estimates $(\hat{\Theta}, \hat{\mathbf{Q}})$ and $\hat{\delta}$. Set B arbitrary large.

for $b = 1, \dots, B$ **do**

1. Set $k = 1$, $U_1 = T_{\min}$ and draw $Z_1 \sim \hat{\delta}$.

while $U_k < T_{\max}$ **do**

2. $k = k + 1$;

3. sample $u_k \sim \text{Exp}(-\hat{q}_{Z_{k-1}, Z_{k-1}})$ and set $U_k = U_{k-1} + u_k$;

4. sample Z_k from $\mathbb{P}(Z_k = s | Z_{k-1} = r) = -\hat{q}_{r,s} / \hat{q}_{r,r}$, with $r, s \in \mathcal{S}$.

5. For $i = 1, \dots, n$, let $k(i)$ an integer s.t. $t_i \in \mathbf{U}_{k(i)}$ and $\mathbf{U}_{k(i)} = (U_{k(i)}, U_{k(i)+1})$.

Set $S_{t_i}^{(b)} = Z_{k(i)}$.

6. For $i = 1, \dots, n$, sample $y_{t_i}^{(b)} | y_{t_{i-1}}^{(b)}, S_{t_i}^{(b)} = s \sim \mathcal{N}(\mu_{i,s}^*, \sigma_{i,s}^{2*})$. Store the n -size bootstrap replicate $\mathcal{Y}^{(b)}$.

7. Fit the model on $\mathcal{Y}^{(b)}$ to get $(\hat{\Theta}, \hat{\mathbf{Q}})^{(b)}$.

Output: The bootstrap replicates $\mathcal{Y}^{(b)}$ and the estimates $(\hat{\Theta}, \hat{\mathbf{Q}})^{(b)}$, for $b = 1, \dots, B$.

and $\hat{q}_{s,s} = -\sum_{s' \neq s} \hat{q}_{s,s'}$. The solutions to the score equations for the mean μ_s and the square of the volatility σ_s^2 are available in closed form:

$$\begin{aligned} \hat{\mu}_s &= \frac{\sum_{i=2}^n \frac{\hat{\xi}_i(s)}{1 - e^{-2\hat{\tau}_s \Delta_i}} (y_{t_i} - y_{t_{i-1}} e^{-\hat{\tau}_s \Delta_i}) (1 - e^{-\hat{\tau}_s \Delta_i})}{\sum_{i=1}^n \frac{\hat{\xi}_i(s)}{1 - e^{-2\hat{\tau}_s \Delta_i}} (1 - e^{-\hat{\tau}_s \Delta_i})^2}, \\ \hat{\sigma}_s^2 &= \frac{\sum_{i=2}^n 2\hat{\tau}_s \frac{\hat{\xi}_i(s)}{1 - e^{-2\hat{\tau}_s \Delta_i}} (y_{t_i} - y_{t_{i-1}} e^{-\hat{\tau}_s \Delta_i} - \hat{\mu}_s (1 - e^{-\hat{\tau}_s \Delta_i}))^2}{\sum_{i=1}^n \hat{\xi}_i(s)}, \end{aligned} \tag{2.6}$$

while the speed of mean reversion τ_s requires a Newton-Raphson step to be updated. We further update the initial probability vector $\hat{\delta}$ using the stationary distribution of the Markov chain when $\Delta_i \rightarrow \infty$, that can be easily computed by taking the transition probability matrix $\exp\{\hat{\mathbf{Q}}\Delta_t\}$ and Δ_t large. It can be proved that the rows of the resultant matrix become equal to the stationary distribution of the process [Norris 1998, Chapter 3]. Last, by exploiting the definition of the rescaling factor c_i , the log-likelihood value is given by $\ell(\hat{\Theta}, \hat{\mathbf{Q}}) = \sum_{i=1}^n \log c_i$.

2.3.2 Model assessment via bootstrap algorithm

In the steps that follow the parameter estimation, we focus on assessing the model accuracy and the distribution of the parameter estimators. We furthermore need to test the presence of outliers with respect to the model, and thus if an additional state is required to adequately fit the light curve. Depending on the type of the source, the proposed hidden Markov model requires in fact a different number of regimes to

accurately describe the underlying physical mechanisms that cause the flares.

The EM algorithm is widely used in statistical literature for estimating the model parameters, and can be further exploited to compute the Fisher information and get the standard errors of the estimates. The approach has been investigated also for continuous time models [Rydén 1996, Lu 2017]; however, the information matrix is often not available in closed-form, and thus its derivation can be computationally demanding. Furthermore, under the hidden Markov modelling assumption, the distribution of the maximum value admitted by the process $\max(Y)$ has not a closed form, and thus it cannot be used to test the presence of outliers.

We provide an alternative approach to the model assessment based on bootstrap methods [Davison and Hinkley 1997, Zucchini *et al.* 2016 in Section 3.6.2]. The parametric bootstrap exploits the generating process induced by the model, described also by the DAG plot in Figure 2.1, and draws new replicates of the original light curve. The replicates are then compared to the observed time series to evaluate the adequacy of the model in describing the data generating process. In addition, the parameter estimates obtained on the bootstrap samples allow us to infer the distribution of the maximum likelihood estimators, and thus quantify the bias and the standard errors of the estimates.

The bootstrap algorithm is given in Code-Box 4. Steps 1-4 sample the underlying path of the latent process. The duration time, i.e., the time spent by the process in a state before the next jump, is exponentially distributed according to Albert (1962) and Bladt and Sørensen (2005); the transition probabilities to the other states are given in Step 4. Steps 5-7 draw the replicates of the flux in the observation times (t_1, \dots, t_n) conditioned to the sampled latent path. In the last step, the algorithm fits the model on the bootstrap replicates and stores the parameter estimates. By taking the maximum of each bootstrap sample $\mathcal{Y}^{(b)}$, we are able also to efficiently infer the distribution of $\max(Y)$.

2.4 Application to the *Fermi* LAT data

In this section, we present an application of our model to a real-case dataset, and in particular we consider a γ -ray light curve from the blazar PKS 1510-05 recorded by the *Fermi* LAT. Blazars are very luminous and energetic sources characterised by an highly variable signal, with heavy fluctuations in brightness on short time intervals.

The available light curve made by 630 flux measurements is displayed in Figure 2.2. Most of the observations are equally spaced in time with $\Delta_i = 2$; however, the light curve

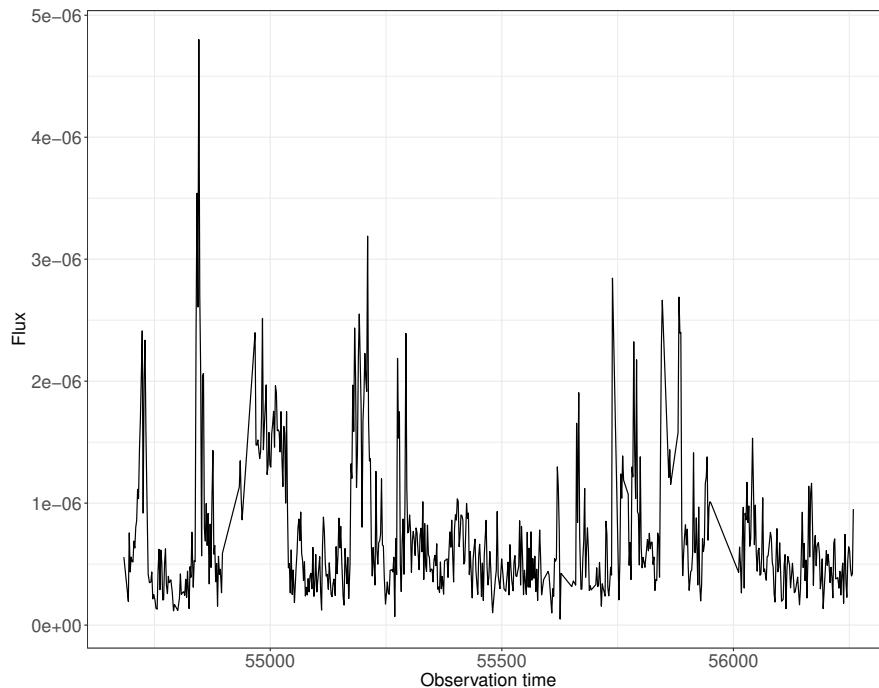


FIGURE 2.2: The γ -ray light curve from the blazar PKS 1510-05 recorded by the *Fermi* LAT telescope over 630 observation times. The most frequent time gap is $\Delta_i = 2$, and the largest is $\Delta_i = 60$. The light curve is evidently punctuated by flares over the entire observation period.

shows some not negligible time gaps that make the continuous time modelling approach adequate to fit this kind of data. Along the entire observation period, we distinguish a prevalent resting phase in the lower part of the graph; in addition, lots of flares in the form of spikes are visible. We conclude that, from a mere graphical inspection, the presence of at least two distinct classes of emission activity emerges from the light curve.

We set a two-states continuous time model: the first is taken as OU and is dedicated to fit the moderate emission activity in the lower part of the graph (between 0 and $1 \cdot 10^{-6}$), while the second is log-normal distributed to account for the intense flares in the light curve [Kushwaha *et al.* 2017]. In this way, we are implicitly assuming that the logarithm of the flux measurements in the second state is OU distributed, which implies that $y_{t_{i-1}+\Delta_i} | y_{t_{i-1}} \sim \log \mathcal{N}(\mu_{i,2}^*, \sigma_{i,2}^{2*})$, and thus the emission activity in the second state is a log-OU process.

Table 2.1 displays the estimates of the model parameters. The fourth column is the estimated diagonal of the transition probability matrix $\exp\{2\hat{\mathbf{Q}}\}$ and gives the probability of S to remain in the same state after the most frequent time gap $\Delta_i = 2$, also known as persistence probability; it further confirms that both the discovered states find large evidence. The second component, labelled as $s = 2$, refers to the log-OU process: it is faster mean reverting as $\hat{\tau}_2$ is smaller $\hat{\tau}_1$, and it has a smaller persistence probability

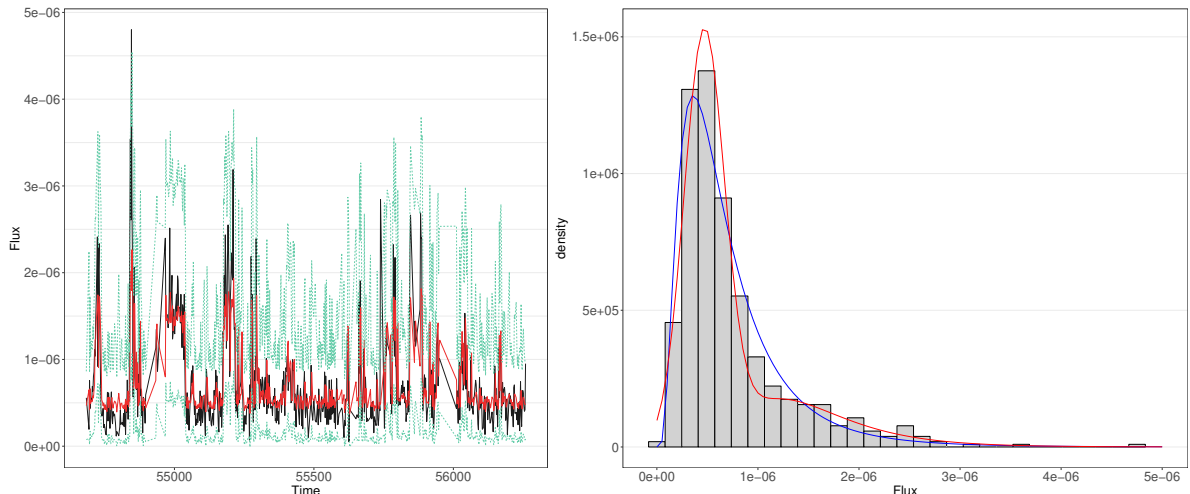


FIGURE 2.3: Left: γ -ray light curve of the blazar PKS 1510-05 (solid black line) against the mean (solid red line) and 95% confidence interval (dashed lines) of the predictive density. Right: histogram of the flux compared to the limit distributions of the process in Formula (2.7) (red line) and of a single log-OU process (blue line).

than the first. The mean and the square of the volatility parameters in the table have different scales and so they are not directly comparable. For this reason, we display in the last two columns the mean and the variance of the limit distributions: when $s = 1$, the limit is $\mathcal{N}(\hat{\mu}_1, \hat{\sigma}_1^2 / (2\hat{\tau}_1))$, while for $s = 2$ it is $\log \mathcal{N}(\hat{\mu}_2, \hat{\sigma}_2^2 / (2\hat{\tau}_2))$. It emerges that the estimated flux in the second state is remarkably larger and more variable than in the first. We can conclude that the log-OU state mainly describes the flaring activity of the source, where a general increase of the average flux is anticipated by prominent spikes in the light curve.

The graphs in Figure 2.3 display the observed light curve against the fitted model, and give an indication of the goodness of fit. In the left plot, we compare the observed flux with the predictive distribution $Y_{t_i} | \mathcal{Y}^{1:(i-1)}$ [Zucchini *et al.* 2016, Chapter 5], where $\mathcal{Y}^{1:(i-1)}$ is the vector of the first $i - 1$ elements in the time series: the solid red line represents the prediction $\hat{y}_{t_i} = \mathbb{E}(Y_{t_i} | \mathcal{Y}^{i-1}, \hat{\Theta}, \hat{Q})$ and describes the trend of brightness

TABLE 2.1: Estimates of the model parameters in the two latent states. From left to right: mean, square of the volatility, speed of mean reversion and probability to remain in the same state after an interval $\Delta_i = 2$. The standard errors obtained with $B = 200$ bootstrap replicates are given in parenthesis. The last two columns show the mean and the variance of the limit distributions in the two states.

	$\hat{\mu}_s$	$\hat{\sigma}_s^2$	$\hat{\tau}_s$	$\hat{p}_{s,s}(\Delta_t = 2)$	$\lim_{\Delta \rightarrow \infty} \mathbb{E}_Y$	$\lim_{\Delta \rightarrow \infty} \mathbb{V}_Y$
$s = 1$	$4.69 \cdot 10^{-7}$ ($1.602 \cdot 10^{-8}$)	$5.563 \cdot 10^{-14}$ ($8.822 \cdot 10^{-15}$)	0.699 (0.08)	0.952 (0.012)	$4.69 \cdot 10^{-7}$	$3.977 \cdot 10^{-14}$
$s = 2$	-13.443 (0.066)	0.172 (0.032)	0.522 (0.119)	0.868 (0.039)	$1.576 \cdot 10^{-6}$	$4.446 \cdot 10^{-13}$

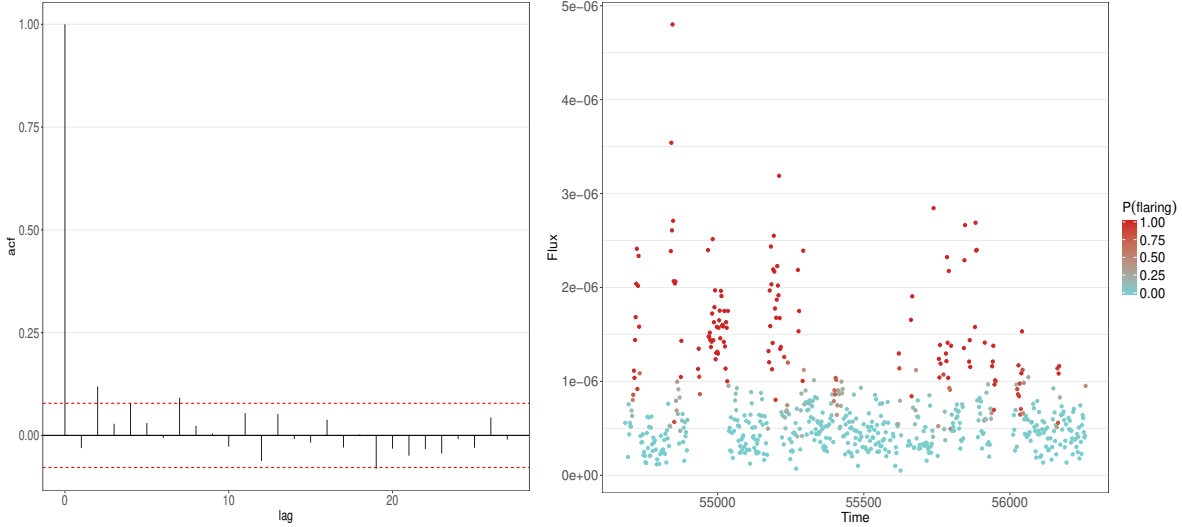


FIGURE 2.4: Left: autocorrelation function of the model residuals. Right: plot of the classification of the flux measurements. For every data point i , the colour represents $\hat{\xi}_i(2)$, the estimated probability to be a flaring emission. A shift toward red states that the observation is more probable to come from the flaring activity.

in time, while the dashed lines are respectively the 2.5% and 97.5% quantiles of the predictive distribution. The empirical coverage of the interval is respected. In the right plot of Figure 2.3, we compare the marginal histogram of the flux with the limit distribution of the proposed HMM, that is the mixture

$$\hat{\delta}_1 \mathcal{N} \left(\hat{\mu}_1, \frac{\hat{\sigma}_1^2}{2\hat{\tau}_1} \right) + \hat{\delta}_2 \log \mathcal{N} \left(\hat{\mu}_2, \frac{\hat{\sigma}_2^2}{2\hat{\tau}_2} \right). \quad (2.7)$$

The mixture kernels are the limit distributions of the OU and log-OU processes, and $\hat{\delta} = (0.732, 0.268)$ is the proportion of resting and flaring emissions in the long-term period. In addition, following Kushwaha *et al.* (2017), we display also the limit distribution of a single log-OU process, fitted via maximum likelihood estimation. It emerges from the plot that the mixture model (2.7) (red line) is globally more adequate than the single log-OU process (blue line), and accounts also for the extreme values in the right tail of the histogram caused by the multiple jets in the light curve. These conclusions are also numerically supported: the maximum of the log-likelihood functions of the continuous time HMM and the single log-OU are 8569.888 and 8524.614, respectively. The chi-square ratio test gives a p-value of $1.26 \cdot 10^{-17}$, confirming that the proposed model is far more appropriate to fit this γ -ray light curve than a single stochastic process.

Figure 2.4, on the left, displays the autocorrelation function of the model residuals $\hat{r}_{t_i} = y_{t_i} - \hat{y}_{t_i}$: the autocorrelation values fall within the red confidence bars after the first lag, confirming that our model correctly fits also the temporal dependency among

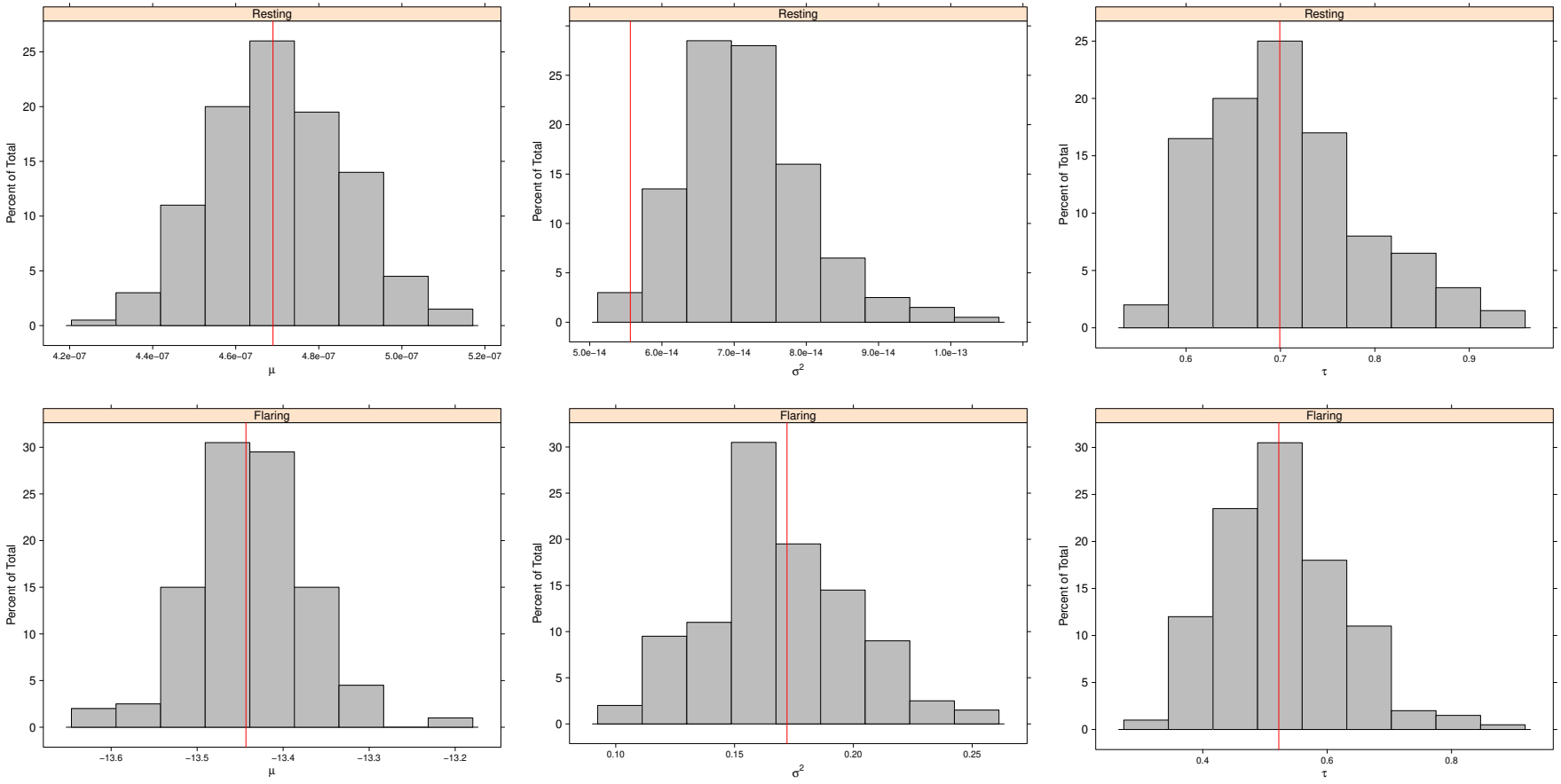


FIGURE 2.5: Bootstrap distributions of the estimators of $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\tau}$ in the normal (top line) and log-normal (bottom line) distributions based on $B = 200$ replicates.

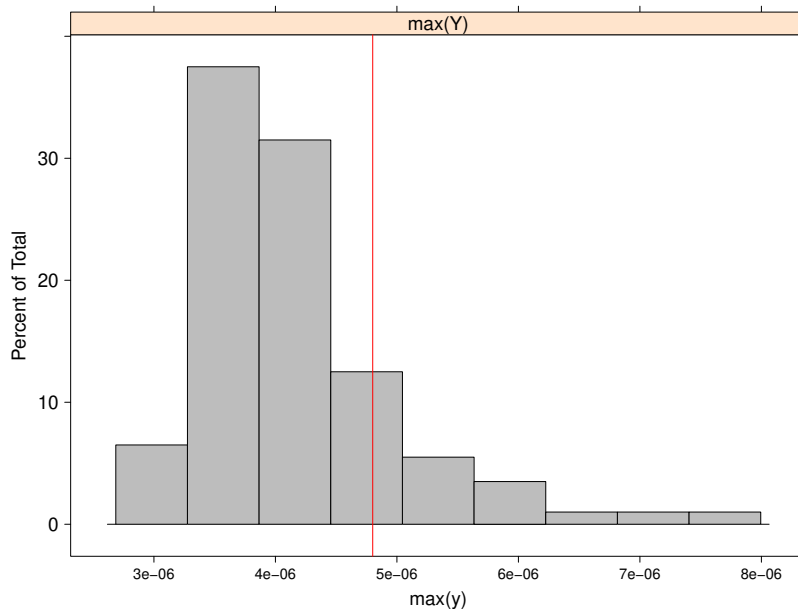


FIGURE 2.6: Distribution of $\max(Y)$ based on $B = 200$ bootstrap replicates. The red line corresponds to the maximum in the observed light curve \mathcal{Y} . The proportion of bootstrap values larger than the observed maximum is 0.16.

the data.

We further classify the flux data points in the light curve using the estimated probabilities $(\hat{\xi}_1, \dots, \hat{\xi}_n)$. In Figure 2.4, data are coloured with respect to $\hat{\xi}_i(2)$, the estimated probability of being a flaring observation. The model clearly performs a distinct separation of the two classes, dividing the flaring activity from the resting phase. The only exception is made by the few observations in the middle of the two states.

We assess the properties of the model using the bootstrap algorithm in Code-Box 4. We sample $B = 200$ replicates from the model to reconstruct the distribution of the parameter estimators in the two phases. We display in Figure 2.5 the histograms of the fitted values on the bootstrap replicates $(\hat{\Theta}_1, \dots, \hat{\Theta}_B)$, and we compare them to the maximum likelihood estimates on the light curve \mathcal{Y} , shown with red lines. The graphs do not evidence the presence of bias in the estimates, except for the square of the volatility parameter in the normal-distributed state (top-middle graph), whose bootstrap density is shifted to the right with respect to the estimated value on \mathcal{Y} . This bias is prevalently caused by the flux observations between the two phases, showed in Figure 2.4 and previously discussed, that cannot properly be assigned to one of the two components.

Last, we assess also the distribution of $\max(Y)$, the maximum value admitted by the model. Figure 2.6 displays the histogram of the maximum in the bootstrap replicates $(\max(\mathcal{Y}^{(1)}), \dots, \max(\mathcal{Y}^{(200)}))$ against $\max(\mathcal{Y})$: among the 200 replicates, the 16% has

a maximum flux measurement larger than $\max(\mathcal{Y})$, allowing to conclude that even the largest spike is not an outlier with respect to the fitted model, and consequently that an additional state in the HMM is not required.

2.5 Discussion

One of the declared goals of the *Fermi* LAT is to monitor the emission activity, i.e., the number of photons emitted during the time, of many high-energy astronomical sources. The emitted photons cover a vast range of energies in the spectrum, including the γ -rays, which are the most energetic type of radiation in the universe. In addition, high-energy sources are known to be highly variable in the brightness due to distinct physical mechanisms that occur at the base. We proposed in this chapter a new approach to formally distinguish and separate the emission phases in γ -ray light curves. Our proposal consists in a continuous time multistate hidden Markov model. A continuous time Markov chain controls the switch between the latent states of the light curve, and an OU process models the brightness in each phase. We adapted the EM algorithm of Rydén (1996) to fit the model, and we provided a suitable parametric bootstrap to evaluate the model accuracy and assess the distributions of the parameter estimators.

We tested the proposed method on a light curve of 630 flux measurements from the blazar PKS 1510-05 recorded by the *Fermi* LAT. The model brings out the presence of two distinct classes of variability in the light curve: the first denotes a resting phase characterised by a roughly constant emission along the entire observation period, and the second outlines instead an highly variable emission phase characterised by jets and spikes. In addition, our method surpasses the simpler competitor proposed by Kushwaha *et al.* (2017) for blazar light curves both in terms of model accuracy and capacity to detect the extreme values.

The model can be further extended to account for many aspects that we did not consider in this chapter. First, we can include a step in the estimation algorithm to automatically select the number of latent states for the Markov process, thereby leading to a model that admits new states of flaring activity as the time goes on. For this purpose, Bayesian methods offer several solutions and powerful estimation algorithms to infer on the number of latent states. A second task would involve the characterisation of the flaring period. The duration of flares is a key element in the analysis of time-varying sources as it plays a central role in the comprehension of the nature of the occurring events. The hidden Markov process previously described intrinsically models the duration of the flaring phase with an exponential distribution. However, it is well

established that in practice this assumption may result as inaccurate or inconsistent for the data [Barbu and Limnios 2009]. Hidden semi-Markov models allow to relax the Markovian assumption by replacing the exponential distribution with other suitable alternatives. Although some parametric distributions can be considered, we are convinced that nonparametric methods will offer a more accurate comprehension of the duration of the flaring activity of a source. A new approach would then not impose any parametric restriction to the persistence time distribution, in order to offer a more detailed and fully data driven comprehension of the flaring activity and of the underlying physical mechanisms.

Chapter 3

Conclusions

“e quindi uscimmo a riveder le stelle.”

(“and thence we came forth to see again the stars.”)

Dante Alighieri - *La Divina Commedia*

The rapid evolution of technology during the last decades has allowed astronomers to collect huge amounts of data from sky surveys in a single day of observation. In parallel to the increasing availability of information, the demand of advanced statistical techniques to extrapolate the relevant knowledge has consistently increased to the extent that Statistics nowadays covers a core role in astrophysical research. This thesis aims to enlarge the available statistical toolbox for the analysis of complex datasets in astronomy, with special emphasis on the study and the explanation of high-energy astrophysical phenomena. In particular, we provided innovative solutions and efficient computational methods to address the following tasks: (i) discovering and locating the sources in the sky, (ii) extracting and quantifying their signal and (iii) characterising their emission activity. We proposed both frequentist and Bayesian solutions which find their roots in the field of unsupervised learning. There, data are assumed to be generated according to several distinct clusters, whose labels need to be unveiled.

The main contributions of the thesis are split in the two chapters 1-2. Chapter 1 focuses on Tasks 1 and 2. It presents a new model which exploits both the spatial information and the energy contents of the photons to locate the emitting sources in the sky. It furthermore clusters the observations and quantifies the intensities of the clusters. We modelled the spatial distribution of the photons with a Dirichlet process mixture, which identifies the extra-galactic sources we are searching for. To separate them from the diffuse gamma-ray background which spreads over the entire map, we represented the latter by a Bayesian nonparametric mixture of B-spline functions. The resultant model is a hierarchical mixture of nonparametric mixtures which allows us to

flexibly cluster highly contaminated signals. We simultaneously determined the number of sources in the map, their directions and their intensities. Inference was carried out by MCMC simulation where no tuning parameter needed to be specified. A suitable post-processing algorithm quantifies the information coming from the detected clusters. Numerical assessment on artificial datasets confirmed the capacity of our model to efficiently discovering and locating the sources in the map. In addition, a comparison study on several simulated datasets highlighted how considering the energy contents of the photons efficiently improved the ability of the algorithm of extracting the signal of the sources. Finally, by applying the models on the *Fermi* LAT data, we confirmed the presence of already known sources, and we located also new candidate points, whose nature must be further investigated. Future developments will focus on how to include the energy contents into the Dirichlet Process models, with particular attention to the background component, and how to develop a formal procedure for disentangling the signal of overlapping sources.

Chapter 2 addresses Task 3. It focuses on the investigation of the emission activity of time-varying astronomical sources, whose signal is known to be highly variable across time. Most contributions in astrophysics use a single continuous time stochastic process to describe the emission activity of a source. However, the recent analyses revealed that some sources are characterised by multiple flaring phases which are due to the distinct underlying physical mechanisms. We proposed a continuous time multistate model to handle the presence of several states of variability. A continuous time Markov chain controlled the switching between the latent phases, while the flux within the states was modelled by a suitable stochastic autoregressive process. The proposed model belongs to the class of hidden Markov models (HMMs) and can be efficiently fitted by maximum likelihood estimation using the Expectation-Maximization (EM) algorithm. In addition, we developed a bootstrap algorithm to assess whether additional latent states in the model are present. A case study on a blazar light curve confirmed the desired capability of the proposed method to correctly separate the multiple emission phases of the analysed source. Further developments will focus on investigating whether an alternative distribution for the duration of flares improves the model accuracy. We will exploit the hidden semi-Markov modelling approach, which allows to relax the Markovian assumption on the latent process, and we will replace the exponential distribution with further parametric and nonparametric alternatives, with the ultimate aim of offering a clearer comprehension of the duration of flares.

Appendix

A1. Adapting the Student t distribution to the *Fermi* LAT PSF

We outline here how to derive the approximation of the *Fermi* LAT PSF based on the bivariate Student t distribution (see Section 1.2.1). The *Fermi* PSF is available from Ackermann *et al.* (2012) as a cube of pixels, where the first two dimensions refer to the spatial coordinates and the third to the energy. Thus, the PSF in the pixel uwz gives the probability that a photon with energy in z -th bin falls into the uw -th spatial bin of the detector. For every energy bin z , we draw a large sample of artificial photons in the form of Poisson counts using the LAT PSF, and we fit a bivariate Student t distribution using maximum likelihood estimation. The resultant parameters estimates are σ_z^2 and ν_z . In order to guarantee the existence of the Student t variance, we first set the degrees of freedom to 2.1 for every energy bin z , and after we derive the maximum likelihood estimate of σ_z^2 conditioned to $\nu_z = 2.1$.

A2. Gibbs sampling algorithms

We outline the Gibbs sampling steps used in the MCMC algorithm described in Section 1.3.2.

Draw the coordinates of the source μ_j

Let $\{\mathbf{x}_i\}_{i=1}^m$ be the set of photons assigned to the j -th component of the source mixture model. Then:

1. $\forall i$, sample from

$$u_i | \mathbf{x}_i, \boldsymbol{\mu}_j^{(t-1)} \sim \text{Gamma} \left\{ \frac{\nu_i}{2} + 1, \frac{1}{2\sigma_i^2} \left(\sigma_i^2 \nu_i + \|\mathbf{x}_i - \boldsymbol{\mu}_j^{(t-1)}\|^2 \right) \right\};$$

2. sample from

$$\boldsymbol{\mu}_j^{(t)} | \{\mathbf{x}_i, u_i\}_i \sim \mathcal{N}_{\mathcal{X}} \left\{ \left(\sum_{i=1}^m \frac{u_i}{\sigma_i^2} \right)^{-1} \left(\sum_{i=1}^m \frac{u_i \mathbf{x}_i}{\sigma_i^2} \right) I, \left(\sum_{i=1}^m \frac{u_i}{\sigma_i^2} \right)^{-1} I \right\},$$

where $\mathcal{N}_{\mathcal{X}}$ is a bivariate Gaussian distribution bounded into \mathcal{X} .

Draw the B-spline knots $(\boldsymbol{\ell}_j, \mathbf{b}_j)$

Let $\{\omega_i\}_{i=1}^m$ be a set of n observations bounded into the interval $(\omega_{min}, \omega_{max})$ with density $\tilde{\mathcal{B}}_4(\cdot | \boldsymbol{\tau})$, and $\boldsymbol{\tau}$ a five-dimensional vector of knots that will be replaced in our algorithm by $\boldsymbol{\ell}$ and \mathbf{b} . At the t -th iteration, the full-conditional distribution of τ_k is

$$\pi(\tau | \tilde{\boldsymbol{\tau}}_{-k}, \{\omega_i\}_i) \propto \prod_{i=1}^m \tilde{\mathcal{B}}_4(\omega_i, \tilde{\boldsymbol{\tau}}) \mathcal{G}_0(\tilde{\boldsymbol{\tau}})$$

where $\tilde{\boldsymbol{\tau}} = (\boldsymbol{\tau}_{1:(k-1)}^{(t)}, \tau, \boldsymbol{\tau}_{(k+1):5}^{(t-1)})$ if $k = 2, 3, 4$, $\tilde{\boldsymbol{\tau}} = (\tau, \boldsymbol{\tau}_{2:5}^{(t-1)})$ if $k = 1$ and $\tilde{\boldsymbol{\tau}} = (\boldsymbol{\tau}_{1:4}^{(t)}, \tau)$ if $k = 5$, $\tilde{\boldsymbol{\tau}}_{-k}$ is the vector without its k -th element and \mathcal{G}_0 is the base measure of a set of knots as specified in Section 1.2.2. Then, for $k = 1, \dots, 5$:

1. let $(\tau_{k,left}, \tau_{k,right})$ the lower and the upper bounds of τ_k given in Table A1, and

$$\tilde{b} = \max_{\tau} \pi(\tau | \tilde{\boldsymbol{\tau}}_{-k}, \{\omega_i\}_i) (\tau_{k,right} - \tau_{k,left});$$

2. sample $\tau^* \sim \mathcal{U}(\tau_{k,left}, \tau_{k,right})$ and $u \sim \mathcal{U}(0, 1)$: if

$$\frac{u \tilde{b}}{(\tau_{k,right} - \tau_{k,left})} < \pi(\tau^* | \tilde{\boldsymbol{\tau}}_{-k}, \{\omega_i\}_i)$$

then $\tau_j^{(t)} = \tau^*$.

TABLE A1: Lower and upper bounds of the full-conditional distribution of the spline knots. $\omega_{(1)}$ is the lowest value in $\{\omega_i\}_{i=1}^m$, while $\omega_{(n)}$ is the largest.

<i>Knot</i>	<i>Left bound</i>	<i>Right bound</i>
τ_1	ω_{min}	$\min(\omega_{(1)}, \tau_2^{(t-1)})$
τ_2	$\tau_1^{(t)}$	$\tau_3^{(t-1)}$
τ_3	$\tau_2^{(t)}$	$\tau_4^{(t-1)}$
τ_4	$\tau_3^{(t)}$	$\tau_5^{(t-1)}$
τ_5	$\max(\omega_{(n)}, \tau_4^{(t)})$	ω_{max}

Finally, check whether the imposed constraints on the B-spline variance are respected using $\boldsymbol{\tau}^{(t)}$ (Section 1.2.4). If not, go back to $\boldsymbol{\tau}^{(t)} = \boldsymbol{\tau}^{(t-1)}$ and restart the procedure. This sampling scheme can be adopted to draw independently from the posteriors of $\boldsymbol{\ell}_j$ and \boldsymbol{b}_j .

B. Details of the results in Section 1.4.2

This Section contains the details of the results obtained from the spatial model and its extension over three scenarios in Section 1.4.2. Every table describes a scenario; results above the horizontal line in the middle refer to the model $f(\cdot|\cdot)$, while the ones under the line are from $f^{ext}(\cdot|\cdot)$.

clusterID	sourceID	ψ	$\mathbb{E}(\mu \dots)$	$\mathbb{E}(N \dots)$	HPD _{90%}	#Counts	$\mathbb{P}(\text{source} \dots)$
1	3	(-3.025, 3.875)	(-3.022, 3.863)	141.228	(126, 156)	130	1.000
2	7	(-1.375, 3.075)	(-1.411, 3.096)	145.457	(129, 163)	132	1.000
3	6	(-1.575, 2.125)	(-1.606, 2.075)	132.911	(114, 152)	118	1.000
4	5	(-1.925, 1.525)	(-1.948, 1.512)	152.561	(134, 172)	144	1.000
5	1	(-3.575, -0.125)	(-3.608, -0.123)	142.652	(128, 158)	126	1.000
6	2	(-3.225, -2.075)	(-3.266, -2.115)	120.686	(106, 136)	132	1.000
7	4	(-2.375, -2.875)	(-2.387, -2.882)	115.630	(101, 131)	131	1.000
8	8	(0.375, -2.625)	(0.347, -2.677)	147.688	(130, 166)	116	1.000
9	9	(0.775, -3.625)	(0.737, -3.648)	135.008	(119, 152)	136	1.000
10	11	(1.425, -0.375)	(1.372, -0.429)	125.579	(109, 144)	124	0.999
11	12	(1.475, -1.525)	(1.421, -1.553)	148.315	(130, 166)	130	1.000
12	14	(3.325, -0.425)	(3.276, -0.429)	130.985	(115, 148)	133	1.000
13	13	(2.675, 2.075)	(2.641, 2.075)	149.686	(133, 167)	139	1.000
14	15	(3.575, 3.125)	(3.569, 3.096)	133.204	(117, 150)	120	1.000
15	10	(0.925, 3.825)	(0.932, 3.761)	122.828	(107, 138)	121	1.000
1	11	(1.425, -0.375)	(1.381, -0.414)	124.271	(108, 141)	124	1.000
2	2	(-3.225, -2.075)	(-3.235, -2.094)	124.935	(110, 141)	132	1.000
3	9	(0.775, -3.625)	(0.744, -3.607)	124.024	(108, 140)	136	1.000
4	5	(-1.925, 1.525)	(-1.961, 1.519)	145.293	(127, 165)	144	1.000
5	15	(3.575, 3.125)	(3.53, 3.116)	132.841	(117, 149)	120	1.000
6	8	(0.375, -2.625)	(0.346, -2.682)	146.642	(130, 163)	116	1.000
7	10	(0.925, 3.825)	(0.903, 3.788)	125.217	(109, 141)	121	1.000
8	12	(1.475, -1.525)	(1.461, -1.59)	143.726	(127, 160)	130	1.000
9	14	(3.325, -0.425)	(3.291, -0.414)	122.639	(107, 138)	133	0.994
10	13	(2.675, 2.075)	(2.654, 2.023)	149.304	(133, 166)	139	1.000
11	3	(-3.025, 3.875)	(-2.996, 3.872)	138.994	(124, 154)	130	0.999
12	6	(-1.575, 2.125)	(-1.643, 2.107)	128.411	(110, 147)	118	0.988
13	7	(-1.375, 3.075)	(-1.404, 3.116)	144.145	(126, 162)	132	0.960
14	1	(-3.575, -0.125)	(-3.633, -0.162)	143.337	(126, 160)	126	0.949
15	4	(-2.375, -2.875)	(-2.359, -2.85)	123.451	(109, 139)	131	0.618
16	//	//	(-0.927, -1.842)	24.674	(14, 37)	//	0.029

Scenario 1.

clusterID	sourceID	ψ	$\mathbb{E}(\mu \dots)$	$\mathbb{E}(N \dots)$	HPD _{90%}	#Counts	$\mathbb{P}(\text{source} \dots)$
1	7	(-1.375, 3.075)	(-1.41, 3.039)	135.653	(120, 151)	129	1.000
2	5	(-1.925, 1.525)	(-1.957, 1.492)	147.515	(131, 165)	130	1.000
3	11	(1.425, -0.375)	(1.407, -0.407)	144.977	(128, 162)	130	1.000
4	8	(0.375, -2.625)	(0.312, -2.657)	156.715	(140, 174)	135	1.000
5	1	(-3.575, -0.125)	(-3.6, -0.196)	73.277	(61, 86)	98	1.000
6	6	(-1.575, 2.125)	(-1.566, 2.125)	109.641	(94, 127)	114	1.000
7	3	(-3.025, 3.875)	(-3.052, 3.813)	129.066	(114, 144)	114	1.000
8	9	(0.775, -3.625)	(0.781, -3.642)	174.268	(158, 190)	160	1.000
9	14	(3.325, -0.425)	(3.284, -0.477)	121.261	(106, 137)	126	0.990
10	4	(-2.375, -2.875)	(-2.426, -2.868)	125.715	(110, 142)	128	1.000
11	12	(1.475, -1.525)	(1.485, -1.532)	131.321	(115, 148)	124	1.000
12	2	(-3.225, -2.075)	(-3.209, -2.095)	219.127	(203, 235)	201	1.000
13	13	(2.675, 2.075)	(2.658, 2.055)	168.953	(152, 186)	137	0.999
14	15	(3.575, 3.125)	(3.519, 3.11)	164.190	(147, 182)	127	0.998
15	10	(0.925, 3.825)	(0.859, 3.813)	64.927	(51, 79)	74	0.821
16	//	//	(-1.097, 0.508)	30.047	(17, 42)	//	0.089
1	13	(2.675, 2.075)	(2.626, 2.048)	166.051	(150, 183)	137	1.000
2	12	(1.475, -1.525)	(1.443, -1.528)	127.151	(112, 143)	124	1.000
3	10	(0.925, 3.825)	(0.898, 3.836)	62.327	(50, 76)	74	0.999
4	8	(0.375, -2.625)	(0.352, -2.645)	154.646	(137, 173)	135	1.000
5	4	(-2.375, -2.875)	(-2.377, -2.869)	133.612	(117, 150)	128	1.000
6	15	(3.575, 3.125)	(3.536, 3.091)	165.479	(149, 182)	127	1.000
7	6	(-1.575, 2.125)	(-1.559, 2.122)	108.325	(92, 126)	114	0.997
8	11	(1.425, -0.375)	(1.352, -0.41)	143.478	(127, 161)	130	0.974
9	3	(-3.025, 3.875)	(-3.014, 3.836)	126.975	(112, 141)	114	0.967
10	1	(-3.575, -0.125)	(-3.56, -0.187)	80.813	(68, 95)	98	0.920
11	9	(0.775, -3.625)	(0.716, -3.614)	171.257	(155, 188)	160	0.999
12	2	(-3.225, -2.075)	(-3.287, -2.124)	227.767	(211, 244)	201	0.999
13	14	(3.325, -0.425)	(3.263, -0.485)	118.948	(103, 135)	126	0.856
14	7	(-1.375, 3.075)	(-1.377, 3.016)	133.400	(118, 149)	129	0.976
15	5	(-1.925, 1.525)	(-1.922, 1.452)	145.496	(129, 162)	130	0.923
16	//	//	(-1.65, -2.794)	28.301	(14, 42)	//	0.050
17	//	//	(-1.104, 0.483)	24.654	(13, 37)	//	0.022

Scenario 2.

clusterID	sourceID	ψ	$\mathbb{E}(\boldsymbol{\mu} \dots)$	$\mathbb{E}(N \dots)$	HPD _{90%}	#Counts	$\mathbb{P}(\text{source} \dots)$
1	1	(-3.925, -3.075)	(-3.946, -3.083)	389.222	(370, 408)	238	1.000
1	2	(-3.875, -3.025)	(-3.946, -3.083)	389.222	(370, 408)	157	1.000
2	13	(0.025, 3.325)	(0.006, 3.277)	239.043	(220, 257)	270	1.000
3	16	(0.625, 0.025)	(0.57, 0.009)	36.981	(25, 50)	27	0.994
4	22	(3.175, -3.475)	(3.231, -3.613)	248.908	(229, 268)	30	1.000
4	23	(3.225, -3.575)	(3.231, -3.613)	248.908	(229, 268)	229	1.000
5	4	(-1.975, -1.475)	(-2.011, -1.493)	529.974	(508, 552)	535	1.000
6	11	(-0.275, -1.725)	(-0.317, -1.758)	39.945	(26, 55)	21	0.926
7	20	(2.575, -1.275)	(2.586, -1.316)	464.563	(442, 487)	467	1.000
8	12	(-0.125, 0.625)	(-0.156, 0.627)	77.552	(64, 91)	69	0.999
9	24	(3.575, 2.725)	(3.554, 2.747)	108.473	(94, 124)	91	0.999
10	19	(2.525, -0.375)	(2.506, -0.344)	206.648	(186, 227)	217	0.977
11	17	(0.675, -3.375)	(0.651, -3.436)	21.941	(12, 33)	24	0.855
12	8	(-1.025, -0.575)	(-1.043, -0.609)	32.534	(19, 47)	18	0.744
13	15	(0.475, -2.125)	(0.409, -2.111)	202.121	(182, 224)	204	0.943
14	7	(-1.225, -2.825)	(-1.285, -2.818)	172.095	(155, 190)	164	0.991
15	9	(-0.775, -2.325)	(-0.801, -2.376)	71.547	(55, 89)	76	0.790
16	25	(3.625, -2.075)	(3.635, -2.111)	48.543	(35, 63)	43	0.622
17	3	(-3.575, 3.275)	(-3.543, 3.277)	98.895	(85, 113)	111	0.678
18	5	(-1.875, 3.875)	(-1.93, 3.807)	46.350	(35, 59)	37	0.662
19	18	(2.475, 2.825)	(2.425, 2.747)	224.684	(205, 244)	255	0.737
20	21	(2.675, -3.925)	(2.667, -3.878)	24.907	(12, 38)	19	0.296
21	14	(0.225, -3.675)	(0.167, -3.79)	53.375	(14, 97)	65	0.003
22	6	(-1.475, 3.625)	(-1.446, 3.542)	19.067	(10, 29)	20	0.261
23	10	(-0.425, 2.975)	(-0.398, 2.924)	19.007	(9, 31)	9	0.206
24	//	//	(0.167, 2.836)	19.632	(7, 34)	//	0.050
1	14	(0.225, -3.675)	(0.225, -3.72)	85.862	(72, 100)	65	1.000
2	9	(-0.775, -2.325)	(-0.772, -2.34)	73.269	(57, 90)	76	0.996
3	22	(3.175, -3.475)	(3.216, -3.575)	244.879	(207, 266)	30	0.979
3	23	(3.225, -3.575)	(3.216, -3.575)	244.879	(207, 266)	229	0.979
4	7	(-1.225, -2.825)	(-1.232, -2.848)	178.056	(160, 196)	164	1.000
5	19	(2.525, -0.375)	(2.526, -0.378)	204.200	(185, 224)	217	1.000
6	15	(0.475, -2.125)	(0.455, -2.122)	199.518	(177, 220)	204	0.985
7	24	(3.575, 2.725)	(3.523, 2.746)	105.659	(91, 121)	91	1.000
8	13	(0.025, 3.325)	(-0.005, 3.328)	243.865	(225, 263)	270	1.000
9	4	(-1.975, -1.475)	(-1.999, -1.54)	530.908	(510, 553)	535	1.000
10	1	(-3.925, -3.075)	(-3.916, -3.066)	368.748	(347, 389)	238	1.000
10	2	(-3.875, -3.025)	(-3.916, -3.066)	368.748	(347, 389)	157	1.000
11	16	(0.625, 0.025)	(0.608, -0.015)	35.528	(24, 48)	27	0.981
12	11	(-0.275, -1.725)	(-0.312, -1.758)	41.006	(27, 56)	21	0.932
13	18	(2.475, 2.825)	(2.449, 2.819)	225.448	(207, 245)	255	0.997
14	5	(-1.875, 3.875)	(-1.922, 3.836)	47.519	(35, 60)	37	0.951
15	20	(2.575, -1.275)	(2.526, -1.322)	459.778	(438, 483)	467	1.000
16	25	(3.625, -2.075)	(3.599, -2.049)	51.958	(39, 65)	43	0.717
17	12	(-0.125, 0.625)	(-0.158, 0.639)	74.206	(61, 88)	69	0.968
18	8	(-1.025, -0.575)	(-1.002, -0.596)	30.488	(18, 44)	18	0.812
19	3	(-3.575, 3.275)	(-3.533, 3.255)	98.712	(84, 113)	111	0.762
20	6	(-1.475, 3.625)	(-1.462, 3.546)	21.315	(10, 33)	20	0.637
21	17	(0.675, -3.375)	(0.685, -3.429)	19.202	(10, 30)	24	0.726
22	//	//	(4.289, -4.011)	27.165	(16, 38)	//	0.428
23	10	(-0.425, 2.975)	(-0.465, 2.964)	18.788	(9, 31)	9	0.438

Scenario 3.

Bibliography

- Acero, F., Ackermann, M., Ajello, M., Albert, A., Atwood, W., Axelsson, M., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D. *et al.* (2015) Fermi large area telescope third source catalog. *The Astrophysical Journal Supplement Series* **218**(2), 23–63.
- Acero, F., Ackermann, M., Ajello, M., Albert, A., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bellazzini, R., Bissaldi, E. *et al.* (2016) Development of the model of galactic interstellar emission for standard point-source analysis of fermi large area telescope data. *The Astrophysical Journal Supplement Series* **223**(2), 26–48.
- Ackermann, M., Ajello, M., Albert, A., Allafort, A., Atwood, W., Axelsson, M., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D. *et al.* (2012) The Fermi large area telescope on orbit: event classification, instrument response functions, and calibration. *The Astrophysical Journal Supplement Series* **203**(1), 4–73.
- Albert, A. (1962) Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Ann. Math. Statist.* **33**, 727–753.
- Barbu, V. S. and Limnios, N. (2009) *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*. Volume 191. Springer Science & Business Media.
- Billar, C. (2000) Adaptive Bayesian regression splines in semiparametric generalized linear models. *J. Comput. Graph. Statist.* **9**(1), 122–140.
- Blackwell, D. (1973) Discreteness of Ferguson selections. *Ann. Statist.* **1**, 356–358.
- Bladt, M. and Sørensen, M. (2005) Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(3), 395–410.
- Carlson, B. C. (1991) *B*-splines, hypergeometric functions, and Dirichlet averages. *J. Approx. Theory* **67**(3), 311–325.

- Cui, S., Guha, S., Ferreira, M. A. R. and Tegge, A. N. (2015) hmmSeq: a hidden Markov model for detecting differentially expressed genes from RNA-seq data. *Ann. Appl. Stat.* **9**(2), 901–925.
- Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap methods and their application*. Volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2.
- de Boor, C. (2001) *A practical guide to splines*. Revised edition, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York. ISBN 0-387-95366-3.
- de Gunst, M. C. M. and Schouten, B. (2003) Model selection for hidden Markov models of ion channel data by reversible jump Markov chain Monte Carlo. *Bernoulli* **9**(3), 373–393.
- Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998) Automatic Bayesian curve fitting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**(2), 333–350.
- DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001) Bayesian curve-fitting with free-knot splines. *Biometrika* **88**(4), 1055–1071.
- Dodds-Eden, K., Gillessen, S., Fritz, T. K., Eisenhauer, F., Trippe, S., Genzel, R., Ott, T., Bartko, H., Pfuhl, O., Bower, G., Goldwurm, A., Porquet, D., Trap, G. and Yusef-Zadeh, F. (2011) The two states of Sgr A* in the near-infrared: bright episodic flares on top of low-level continuous variability. *The Astrophysical Journal* **728**(1), 37–49.
- Escobar, M. D. (1994) Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**(425), 268–277.
- Escobar, M. D. and West, M. (1995) Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**(430), 577–588.
- Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.
- Fraley, C. and Raftery, A. E. (1998) How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal* **41**(8), 578–588.
- Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97**(458), 611–631.
- Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000) Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**(6), 1051–1079.

- Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.
- Guglielmetti, F., Fischer, R. and Dose, V. (2009) Backgroundsource separation in astronomical images with Bayesian probability theory I. The method. *Monthly Notices of the Royal Astronomical Society* **396**(1), 165–190.
- Hobolth, A. and Stone, E. A. (2009) Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *Ann. Appl. Stat.* **3**(3), 1204–1231.
- Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukherjee, P. and Parkinson, D. (eds) (2010) *Bayesian methods in cosmology*. Cambridge University Press, Cambridge. ISBN 978-0-521-88794-6.
- Jones, D. E., Kashyap, V. L. and van Dyk, D. A. (2015) Disentangling overlapping astronomical sources using spatial and spectral information. *The Astrophysical Journal* **808**(2), 137–160.
- Kelly, B. C., Bechtold, J. and Siemiginowska, A. (2009) Are the variations in quasar optical flux driven by thermal fluctuations? *The Astrophysical Journal* **698**(1), 895–910.
- Kelly, B. C., Sobolewska, M. and Siemiginowska, A. (2011) A stochastic model for the luminosity fluctuations of accreting black holes. *The Astrophysical Journal* **730**(1), 52.
- King, I. (1962) The structure of star clusters. I. An empirical density law. *The Astronomical Journal* **67**, 471–485.
- Knoetig, M. L. (2014) Signal discovery, limits, and uncertainties with sparse on/off measurements: An objective Bayesian analysis. *The Astrophysical Journal* **790**(2), 106–113.
- Kraft, R. P., Burrows, D. N. and Nousek, J. A. (1991) Determination of confidence limits for experiments with low numbers of counts. *The Astrophysical Journal* **374**(1), 344–355.
- Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**(3), 567–580.

- Kushwaha, P., Sinha, A., Misra, R., Singh, K. and Dal Pino, E. d. G. (2017) Gamma-ray flux distribution and nonlinear behavior of four LAT bright AGNs. *The Astrophysical Journal* **849**(2), 138–146.
- Lee, J. M., Kim, S.-J., Hwang, Y. and Song, C.-S. (2004) Diagnosis of mechanical fault signals using continuous hidden Markov model. *Journal of Sound and Vibration* **276**(3-5), 1065–1080.
- Leroux, B. G. and Puterman, M. L. (1992) Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* **48**(2), 545–558.
- Lu, S. (2017) A continuous-time HMM approach to modeling the magnitude-frequency distribution of earthquakes. *Journal of Applied Statistics* **44**(1), 71–88.
- MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., Kozłowski, S., Kelly, B., Bullock, E., Kimball, A., Sesar, B., Westman, D., Brooks, K., Gibson, R., Becker, A. C. and de Vries, W. H. (2010) Modeling the time variability of SDSS stripe 82 quasars as a damped random walk. *The Astrophysical Journal* **721**(2), 1014–1033.
- Malsiner-Walli, G., Frühwirth-Schnatter, S. and Grün, B. (2016) Model-based clustering based on sparse finite gaussian mixtures. *Statistics and Computing* **26**(1), 303–324.
- Mamon, R. S. and Elliott, R. J. (eds) (2014) *Hidden Markov models in finance*. Volume 209 of *International Series in Operations Research & Management Science*. Springer, New York. ISBN 978-1-4899-7441-9; 978-1-4899-7442-6.
- Mattox, J. R., Bertsch, D., Chiang, J., Dingus, B., Digel, S., Esposito, J., Fierro, J., Hartman, R., Hunter, S., Kanbach, G. *et al.* (1996) The likelihood analysis of EGRET data. *The Astrophysical Journal* **461**, 396–407.
- Metzner, P., Horenko, I. and Schütte, C. (2007) Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time. *Physical Review E* **76**(6), 066702.
- Meyer, L., Witzel, G., Longstaff, F. and Ghez, A. (2014) A formal method for identifying distinct states of variability in time-varying sources: Sgr A* as an example. *The Astrophysical Journal* **791**(1), 24–32.
- Müller, P., Quintana, F. A., Jara, A. and Hanson, T. (2015) *Bayesian nonparametric data analysis*. Springer Series in Statistics. Springer, Cham. ISBN 978-3-319-18967-3; 978-3-319-18968-0.

- Müller, P. and Rodriguez, A. (2013) *Nonparametric Bayesian inference*. Volume 9 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Beachwood, OH; American Statistical Association, Alexandria, VA. ISBN 978-0-940600-82-9.
- Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9**(2), 249–265.
- Norris, J. R. (1998) *Markov chains*. Volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. ISBN 0-521-48181-3.
- Ornstein, L. S. and Uhlenbeck, G. E. (1930) On the theory of the brownian motion. *Physical Review (Series I)* **36**, 823–841.
- Park, T., Kashyap, V. L., Siemiginowska, A., van Dyk, D. A., Zezas, A., Heinke, C. and Wargelin, B. J. (2006) Bayesian estimation of hardness ratios: Modeling and computations. *The Astrophysical Journal* **652**(1), 610–628.
- Primini, F. A. and Kashyap, V. L. (2014) Determining x-ray source intensity and confidence bounds in crowded fields. *The Astrophysical Journal* **796**(1), 24–37.
- Protassov, R., van, D. A., Connors, A., Kashyap, V. L. and Siemiginowska, A. (2002) Statistics, handle with care: detecting multiple model components with the likelihood ratio test. *The Astrophysical Journal* **571**(1), 545–559.
- Ramakrishnan, V., Hovatta, T., Nieppola, E., Tornikoski, M., Lähteenmäki, A. and Valtaoja, E. (2015) Locating the γ -ray emission site in Fermi/LAT blazars from correlation analysis between 37GHz radio and γ -ray light curves. *Monthly Notices of the Royal Astronomical Society* **452**(2), 1280–1294.
- Ray, P. S., Kerr, M., Parent, D., Abdo, A., Guillemot, L., Ransom, S., Rea, N., Wolff, M., Makeev, A., Roberts, M. *et al.* (2011) Precise γ -ray timing and radio observations of 17 fermi γ -ray pulsars. *The Astrophysical Journal Supplement Series* **194**(2), 17–44.
- Roberts, W. J., Ephraim, Y. and Dieguez, E. (2006) On Rydén’s EM algorithm for estimating MMPPs. *IEEE Signal Processing Letters* **13**(6), 373–376.
- Roberts, W. J. J. and Ephraim, Y. (2008) An EM algorithm for ion-channel current estimation. *IEEE Trans. Signal Process.* **56**(1), 26–33.

- Rydén, T. (1996) An EM algorithm for estimation in Markov-modulated Poisson processes. *Comput. Statist. Data Anal.* **21**(4), 431–447.
- Schellhase, C. and Kauermann, G. (2012) Density estimation and comparison with a penalized mixture approach. *Comput. Statist.* **27**(4), 757–777.
- Schliep, A., Schönhuth, A. and Steinhoff, C. (2003) Using hidden Markov models to analyze gene expression time course data. *Bioinformatics* **19**(suppl_1), i255–i263.
- Selig, M. and Enßlin, T. A. (2015) Denoising, deconvolving, and decomposing photon observations. Derivation of the D³PO algorithm. *Astronomy & Astrophysics* **574**, A74.
- Selig, M., Vacca, V., Oppermann, N. and Enßlin, T. A. (2015) The denoised, deconvolved, and decomposed fermi γ -ray sky-An application of the D³PO algorithm. *Astronomy & Astrophysics* **581**, A126.
- Sharef, E., Strawderman, R. L., Ruppert, D., Cowen, M. and Halasyamani, L. (2010) Bayesian adaptive B-spline estimation in proportional hazards frailty models. *Electron. J. Stat.* **4**, 606–642.
- Sobolewska, M. A. and Papadakis, I. E. (2009) The long-term X-ray spectral variability of AGN. *Monthly Notices of the Royal Astronomical Society* **399**(3), 1597–1610.
- Sobolewska, M. A., Siemiginowska, A., Kelly, B. C. and Nalewajko, K. (2014) Stochastic modeling of the Fermi/Lat γ -ray blazar variability. *The Astrophysical Journal* **786**(2), 143–156.
- Städler, N. and Mukherjee, S. (2013) Penalized estimation in high-dimensional hidden Markov models with state-specific graphical models. *Ann. Appl. Stat.* **7**(4), 2157–2179.
- Stein, N. M., van Dyk, D. A., Kashyap, V. L. and Siemiginowska, A. (2015) Detecting unspecified structure in low-count images. *The Astrophysical Journal* **813**(1), 66–80.
- Uttley, P., McHardy, I. M. and Vaughan, S. (2005) Non-linear X-ray variability in X-ray binaries and active galaxies. *Monthly Notices of the Royal Astronomical Society* **359**(1), 345–362.
- van Dyk, D. A. and Meng, X.-L. (2001) The art of data augmentation. *J. Comput. Graph. Statist.* **10**(1), 1–111. With discussions, and a rejoinder by the authors.

- van Dyk, D. A., Connors, A., Kashyap, V. L. and Siemiginowska, A. (2001) Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal* **548**(1), 224–243.
- Wand, M. P. (1994) Fast computation of multivariate kernel estimators. *J. Comput. Graph. Statist.* **3**(4), 433–445.
- Wei, W., Wang, B. and Towsley, D. (2002) Continuous-time hidden markov models for network performance evaluation. *Performance Evaluation* **49**(1-4), 129–146.
- Weisskopf, M. C., Wu, K., Trimble, V., O’Dell, S. L., Elsner, R. F., Zavlin, V. E. and Kouveliotou, C. (2007) A chandra search for coronal x-rays from the cool white dwarf gd 356. *The Astrophysical Journal* **657**(2), 1026–1036.
- Witzel, G., Eckart, A., Bremer, M., Zamaninasab, M., Shahzamanian, B., Valencia-S, M., Schödel, R., Karas, V., Lenzen, R., Marchili, N. *et al.* (2012) Source-intrinsic near-infrared properties of Sgr A*: total intensity measurements. *The Astrophysical Journal Supplement Series* **203**(2), 18–51.
- Zu, Y., Kochanek, C., Kozłowski, S. and Udalski, A. (2013) Is quasar optical variability a damped random walk? *The Astrophysical Journal* **765**(2), 106–112.
- Zucchini, W., MacDonald, I. L. and Langrock, R. (2016) *Hidden Markov models for time series*. Second edition, volume 150 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL. ISBN 978-1-4822-5383-2.

Andrea Sottosanti

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistical Sciences
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 349 081 4924
e-mail: sottosanti@stat.unipd.it

Current Position

Since October 2016; (expected completion: March 2020)

PhD Student in Statistical Sciences, University of Padova.

Thesis title: *Advances in Mixture Modelling for Model-Based Clustering: Two Case Studies in Astronomy*

Supervisor: Dr. Mauro Bernardi

Co-supervisors: Prof. Alessandra R. Brazzale, Prof. David A. van Dyk (Imperial College London), Prof. Roberto Trotta (Imperial College London)

Research interests

- Astrostatistics
- Bayesian Statistics
- Model-Based Clustering
- Mixture Modelling

Education

October 2014 – September 2016

Master (laurea specialistica/magistrale) degree in Statistical Sciences.

University of Padova, Faculty of Statistical Sciences

Title of dissertation: “Searching for Gamma Ray Sources in the Extra-Galactic Space: A Statistical Analysis of the Fermi LAT Data” (in Italian)

Supervisor: Prof. Alessandra R. Brazzale

Final mark: 110/110 with honours

September 2011 – July 2014

Bachelor degree (laurea triennale) in Statistics and Information Management.

University of Milano - Bicocca, Faculty of Statistics

Final mark: 104/110

Visiting positions

2018, Apr. – Oct. Visiting research scholar, Department of Mathematics, Imperial College London, London (UK).
Supervisors: Prof. David A. van Dyk, Prof. Roberto Trotta

Further educational activities

2017, June 25 – 27 *StartUp Research*, Certosa di Pontignano (SI)

Awards and Scholarship

2019, July Winner of the *Best Student Paper Toyota Motor Corporation Award*, 34th International Workshop on Statistical Modelling (IWSM 2019), Guimarães, Portugal
2018, July Runner-up for the *Best Student Paper Toyota Motor Corporation Award*, 33rd International Workshop on Statistical Modelling (IWSM 2018), Bristol, UK
2017, May Scholarship as *Best Master's degree student from Oggiona con Santo Stefano in 2016*
2016, June Scholarship for a position as PhD student in Statistical Sciences, cycle XXXII, University of Padova, Padova, Italy

Grants

2019, June Travel support for the *12th International Conference on Bayesian Nonparametrics*, Oxford, UK
2018, June Travel support for the *Conference of the International Society of Bayesian Analysis*, Edinburgh, UK

Computer skills

SAS Certified Predictive Modeler using SAS Enterprise Miner 7 (December 2013)

Advanced programming skills in R and good skills in the integration of R and C++ through the package Rcpp. Good knowledge of other statistical computing and programming languages as SAS, SAS Enterprise Miner and SPSS.

Advanced knowledge of LaTeX for creating papers, reports and presentations. Personal interests for the BASIC programming language and Arduino.

Language skills

Italian (native), English (writing C1/speaking B2).

Publications

In volumes

1. SOTTOSANTI, A., Costantin, D., Bastieri, D., and Brazzale, A. R. (2019). Discovering and Locating High-Energy Extra-Galactic Sources by Bayesian Mixture Modelling. *New Statistical Developments in Data Science* (A. Petrucci, F. Racioppi and R. Verde Eds.), Springer Proceedings in Mathematics & Statistics, **288**, 135-148.
2. Caponera, A., Denti, F., Rigon, T., SOTTOSANTI, A., and Gelfand, A. (2018). Hierarchical Spatio-Temporal Modeling of Resting State fMRI Data. *Studies in Neural Data Science* (A. Canale, D. Durante, L. Paci and B. Scarpa Eds.), Springer Proceedings in Mathematics & Statistics, **257**, 111-130.

In Proceedings

3. SOTTOSANTI A., Bernardi M., Campos L., Siemiginowska A. and van Dyk D.A. (2019) Continuous time hidden Markov models for astronomical gamma-ray light curves. *Proceedings of the 34th International Workshop on Statistical Modelling (IWSM 2019)*, 7-12 July 2019, Guimarães (Portugal). **Best student paper.**
4. Busatto, C., SOTTOSANTI, A. and Bernardi, M. (2019). Bayesian Variable Selection for High Dimensional Logistic Regression. *Smart Statistics for Smart Applications: Book of Short Papers SIS2019*, 18-21 June 2019, Milan (Italy).
5. SOTTOSANTI, A., Bernardi, M., Brazzale, A.R. (2018) Mining signals of astronomical sources via Bayesian nonparametric mixture modelling. *Proceedings of the 33rd International Workshop on Statistical Modelling (IWSM 2018)*, 16-20 July, Bristol (UK). **Runner-up as best student paper.**
6. SOTTOSANTI, A., Bastieri, D., Brazzale, A.R. (2017). Bayesian Mixture Models for the Detection of High-Energy Astronomical Sources. *Statistics and Data Science: new challenges, new generations*, 28-30 June 2017, Florence (Italy).

Submitted

7. Costantin, D., SOTTOSANTI, A., Brazzale, A. R., Bastieri, D. and Fan, J. (2019) *Bayesian Mixture Modeling of the High-Energy Photon Counts collected by the Fermi Large Area Telescope.*

Working papers

8. SOTTOSANTI, A., Bernardi, M., Brazzale, A. R., Geringer-Sameth, A., Stenning, D. C., Trotta, R. and van Dyk, D. A. () *Identifying High-Energy Astrophysical Sources via Hierarchical Bayesian Nonparametric Clustering.*
9. SOTTOSANTI, A., Bernardi, M., Brazzale, Campos, L., Siemiginowska, A. and van Dyk, D. A. () *Flare Detection in Gamma-ray Light Curves using Continuous Time Hidden Markov Models.*

Conference presentations

Invited

- 2019, July 32nd European Meeting of Statisticians, Palermo, Italy
2019, July Annual meeting of the International CHASC Astro-Statistics Collaboration, Cambridge, USA
2018, December 11th International Conference of the ERCIM WG on Computational and Methodological Statistics, Pisa, Italy

Contributed

- 2019, July 34th International Workshop on Statistical Modelling, Guimarães, Portugal
2019, June 12th International Conference on Bayesian Nonparametrics, Oxford, UK
2018, July 33rd International Workshop on Statistical Modelling, Bristol, UK
2017, June *Statistics and Data Sciences: New Challenges, New Generations*
Conference of the Italian Statistical Society, Florence, Italy

Poster

- 2018, September Conference of the Royal Statistical Society, Cardiff, UK
2018, June Conference of the International Society of Bayesian Analysis, Edinburgh, UK
2017, September *Astro@Stat 2017: Sino-Italian Workshop on Astrostatistics*, Padova, Italy

Thesis supervision

Master's students

- Claudio Busatto 2019 Master's degree in Statistical Sciences, University of Padova, Padova, Italy. Thesis title: *Bayesian Variable Selection in High-Dimensional Logistic Regression* (in co-supervision with Mauro Bernardi)
Emanuele Degani 2018 Master's degree in Statistical Sciences, University of Padova, Padova, Italy. Thesis title: *Bayesian Inference for Mixtures of Support Vector Machines* (in co-supervision with Mauro Bernardi).

Consulting activity

- 2019, April Statistical consulting for Sorgenia SpA

Other skills and Interests

I studied piano for some years and I love practice it. I had been a collaborator of the chess club Amici degli Scacchi OSS (Via S. Giovanni Bosco, 3 - 21040 Oggiona con Santo Stefano - VA) for two years. I did volunteer in kennel of Gallarate (VA). I worked as an electrician during summer holidays. I love walk, travel by train and experience different cultures.