# Bayesian Analysis of Lidar Signals Using Reversible Jump Markov Chain Monte Carlo Algorithms

**Sergio Hernández-Marín**

Thesis submitted for the degree of Doctor of Philosophy


Heriot-Watt University

School of Engineering and Physical Sciences

Department of Electrical, Electronic and Computing Engineering

March 2007

# Declaration

I hereby declare that the work presented in this thesis was carried out by myself at Heriot-Watt University, except where due acknowledgment is made, and has not been submitted for any other degree.

*Author:*

Sergio Hernandez-Marin

*Supervisor:*

Prof. Andrew M. Wallace

Date

# Abstract

Standard 3D ranging and imaging systems process only one return from an assumed single opaque surface. However, there are many situations when the laser return consists of multiple peaks, due for example to the footprint of the beam impinging on a target with surfaces distributed in depth, or on a target with semi-transparent surfaces.

In this thesis, we present a unified theory of pixel processing based on a Bayesian framework, applicable to pulsed laser ranging and depth imaging systems, which allows for a careful and thorough treatment of all types of uncertainties associated with the data. We use reversible jump Markov chain Monte Carlo (RJMCMC) techniques to evaluate the posterior distribution of the parameters and to explore spaces with different dimensionality. Further, we use a delayed rejection step to allow the generated Markov chain to mix better through the use of different proposal distributions. Finally, for full 3D image data, we incorporate spatial constraints on the number of peaks through a particular Markov random field, the Potts prior model. This allows uncertainty about the underlying spatial process. To palliate some inherent deficiencies of this prior model, we also introduce two proposal distributions, one based on spatial mode jumping, the other on a spatial birth/death process.

The methods we employ are demonstrated on simulated and real data, showing that the laser return parameters can be estimated to a high degree of accuracy. We show practical examples from both near and far range depth imaging. Two techniques of particular relevance for this work are Time-Correlated Single Photon Counting (TCSPC) and Burst Illumination Laser (BIL) imaging. In each case, the goal is a complete characterisation of the 3D surfaces viewed by the particular laser imaging system. For TCSPC data, we determine the number, positions and amplitudes of returns from a histogram, or histograms, of photon counts. For BIL data, we determine the same parameters from integrated intensities as a range gate is moved through an image sequence. In either case, a more informative multi-layered 3D image can be created through the spatial analysis.

To Bea
To my family

# Acknowledgements

My sincere thanks to my supervisors, Prof. A.M. Wallace and Prof. G.J. Gibson, for their support, valuable comments and for their insights and suggestions that helped me to become a better researcher. Their valuable feedback contributed greatly to this thesis.

I would like to thank the vision group members Angel, Arvind, Matt, Xun and Zsolt. Without them, the lab would have been very boring.

I would like to thank BAE Systems and SELEX Sensors & Airborne Systems for their support under a research scholarship.

Thanks to my friends Pepi and Rafa for their continuous support and encouragement. Also thanks to my friend Jaime for being my link between Valencia and Edinburgh.

Finally, my special thanks go to Bea for her support, encouragement in every aspect of both my personal and professional life, and invaluable help to overcome difficult times. You make me want to be a better man. This special thanks is also shared with my family for their continuous support and encouragement in difficult (and not so difficult) times. I would not be where I am right now without their support.

# Table of Contents

# List of Figures

# List of Tables

Chapter 1

# Introduction

*There must be a beginning of any great matter, but the continuing unto the end until it be thoroughly finished yields the true glory.*

*Francis Drake.*

Laser ranging and 3D imaging systems have a wide range of current applications. These include 3D metrology for the automotive and aerospace industries, usually performed to high accuracy and at relatively short range, from less than 1 to 25 metres. These also include the detection, classification and positioning of objects at distances of several hundred metres to several kilometres, for example in environmental mapping and vehicle tracking. The goal of our work is a complete characterisation of the 3D surfaces viewed by a laser 3D imaging system.

Current systems are able to measure 3D position (and by scanning or focal plane arrays to acquire 3D images) to high accuracy from single, opaque, near-lambertian or mirror surfaces of high reflectivity. However, the design and signal processing methodology are not suited to cope with a large signal dynamic range, for example from near and far surfaces, or from surfaces of low and high reflectance. In addition, current systems are not equipped to measure multiple returns from glass or other transparent surfaces, or from surfaces at different ranges in the path of a divergent beam, whether or not part of the same scene object. Moreover, such approaches may give suboptimal results because they neglect the correlations between the parameters of neighboring image pixels. Laser imaging should be considered as a spatial process since returns coming from the same surface are expected to have similar amplitude, position and number of peaks. There is often a wide range of plausible reconstructions given the data and what is known about the spatial process. This multitude of possibilities must be accounted for when making

inference since more data does not remove the uncertainty. Thus, plausible reconstructions should reflect both the data and prior knowledge about the process. Providing reasonable variability statements about the estimated process is a difficult task.

This thesis addresses these problems, investigating ways to find the number and positions of many weak and strong reflections in the returned signal, developing techniques based on Bayesian modelling of mixture distributions and Markov chain Monte Carlo techniques. The Bayesian paradigm is a natural framework for addressing these aspects of inference for laser range data. The Bayesian approach allows the use of real prior information, spatial or otherwise, if available through prior distributions. It may incorporate uncertainty in both the model and the prior specification, exposes multiple explanations for the set of plausible reconstructions and avoids unsatisfactory and difficult hypothesis testing for the number of returns. In general, the posterior distributions that arise in such analyses are intractable and computational methods such as Markov chain Monte Carlo (MCMC) are required. However, as the number of scatterers in the returned signal is also an unknown, we need to use a more general version of the MCMC algorithm called reversible jump MCMC (RJMCMC) which allows us to deal with varying-dimension problems. While we shall concentrate on the evaluation of our approach on two types of laser imaging systems, we believe that the proposed methodology is applicable to many forms of pulsed time-series data.

The structure of the thesis is as follows:

**Chapter 2** provides a review of existing laser ranging and 3D imaging systems and their advantages and drawbacks. We also introduce two techniques of particular relevance to this work, Time-Correlated Single Photon Counting and Burst Illumination Laser with sections describing their features. We conclude with some observations regarding the relevance of the reviewed literature to the direction of research presented in this dissertation.

**Chapter 3** provides an introduction to the concept of statistical inference and the Bayesian paradigm. Subsequently, we introduce Markov chain Monte Carlo (MCMC) and reversible jump Markov chain Monte Carlo (RJMCMC) techniques, including relevant

theory for Markov chains. Due to the similarity with our approach, mixture distributions are also reviewed in the context of a Bayesian framework as well as alternative computational tools to RJMCMC for their analysis. Finally, we introduce the topic of spatial statistics with particular emphasis on Markov random fields (MRF). The following two chapters build upon these ideas.

**Chapter 4** presents a new algorithm for 3D ranging and imaging in a Bayesian framework. The posterior distribution is explored using a RJMCMC algorithm. In this chapter, spatial constraints are not considered and the data is analysed on an independent pixel basis. The method is demonstrated on several synthetic and real data sets for single and multiple returns.

**Chapter 5** describes further improvements to the method described in Chapter 4 including the incorporation of spatial dependencies between neighbouring pixels. A MRF prior is used to assess uncertainty about the underlying spatial process whereas two new moves, "spatial mode jumping" and "spatial birth/death process" are included in the RJMCMC algorithm to alleviate some deficiencies of the MRF used. The method is demonstrated on several synthetic and real data sets for multiple returns.

**Chapter 6** concludes the thesis with suggestions for further work.

This work has generated the following journal articles and conference proceedings:

- S. Hernandez-Marin, A.M. Wallace, G.J. Gibson, Markov chain Monte Carlo algorithms for 3D ranging and imaging, Proc. of the 9th IAPR Conference on Machine Vision Applications (Tsukuba, May 2005) pp. 193-196.
- A.M. Wallace, G.S. Buller, R.C.W. Sung, R.D. Harkins, A. McCarthy, S. Hernandez-Marin, G.J. Gibson and R. Lamb, Multi-spectral laser detection and ranging for range profiling and surface characterization, J.Opt. A: Pure Appl. Opt. 7 (2005) S438-S444.
- M.R. Kitchin, S. Hernandez-Marin, A.M. Wallace and G.J. Gibson, Processing and registration of range images from time of flight laser systems, Proc. of the SPIE, Volume 5988, pp. 19-30 (2005).

- S. Hernandez-Marin, A.M. Wallace, G.J. Gibson, Creating Multi-layered 3D Images Using Reversible Jump MCMC Algorithms, Advances in Visual Computing, Lecture Notes in Computer Science, Springer-Verlag, Vol. 4292, p.p. 405-416 (2006).

- R.E. Warburton, A. McCarthy, A.M. Wallace, S. Hernandez-Marin, R. Lamb, G.S. Buller, Enhanced performance Photon-counting Time-of-Flight sensor, Optics Express, Vol. 15, Issue 2, pp. 423-429, (2007).

- S. Hernandez-Marin, A.M. Wallace, G.J. Gibson, Bayesian Analysis of Lidar Signals with multiple returns, IEEE Trans. on Pattern Analysis and Machine Intelligence, in press (2007).

- R.E. Warburton, A. McCarthy, A.M. Wallace, S. Hernandez-Marin, R. H. Hadfield, S. W. Nam, G.S. Buller, Subcentimeter depth resolution using a single-photon counting time-of-flight laser ranging system at 1550 nm wavelength, submitted.

- S. Hernandez-Marin, A.M. Wallace, G.J. Gibson, Spatial modelling of multi-layered Lidar images using reversible jump MCMC, submitted.

- A. McCarthy, S. Hernandez-Marin, R. J. Collins, V. Fernandez, A. M. Wallace and G. S. Buller, Covert Scanning of Low Signature Targets Using High Speed Photon-Counting, submitted.

- S. Hernandez-Marin, A.M. Wallace, G.J. Gibson, Bayesian Spatial Modelling of Lidar Images, in preparation.

Chapter 2

# Laser Ranging and 3D Imaging Systems

*A picture is worth a thousand words.*

*Napoleon Bonaparte.*

The generation of 3D images from optical range finding systems has attracted considerable research interest in recent years, perhaps owing to some advantages over conventional 2D electro-optical intensity imagery within the field of target recognition. Range information helps to discriminate objects since the contrast and reflectance/emittance data of a 2D image does not uniquely identify a target as it gives only limited information regarding the physical shape and size of an object in a scene. In contrast, 3D images provide geometry in terms of spatial coordinates by incorporating depth information. For instance, this geometric information can be used for automatic object recognition/registration between such range images and polygonal mesh models [87]. Interest in Ladar/Lidar systems has grown recently since they have become cheaper, smaller and simpler to construct than other approaches. Moreover, their measurement accuracy is also increasing steadily. The technology of 3D Ladar/Lidar systems has improved to such an extent that depth and reflectance images of objects/targets can now be acquired using scanned detectors or focal plane arrays, sometimes at video frame rates although further development is needed to compete in image quality with conventional acquisition systems such as video recording systems.

This chapter presents (by literature survey) in section 2.1 a review of the current techniques employed in 3D imaging and ranging. In section 2.2 particular emphasis is placed on Time-of-Flight (TOF) Ladar/Lidar techniques. In sections 2.3 and 2.4 two TOF techniques of particular interest to this work are presented. Finally, in section 2.5 some conclusions are extracted from the previous sections.

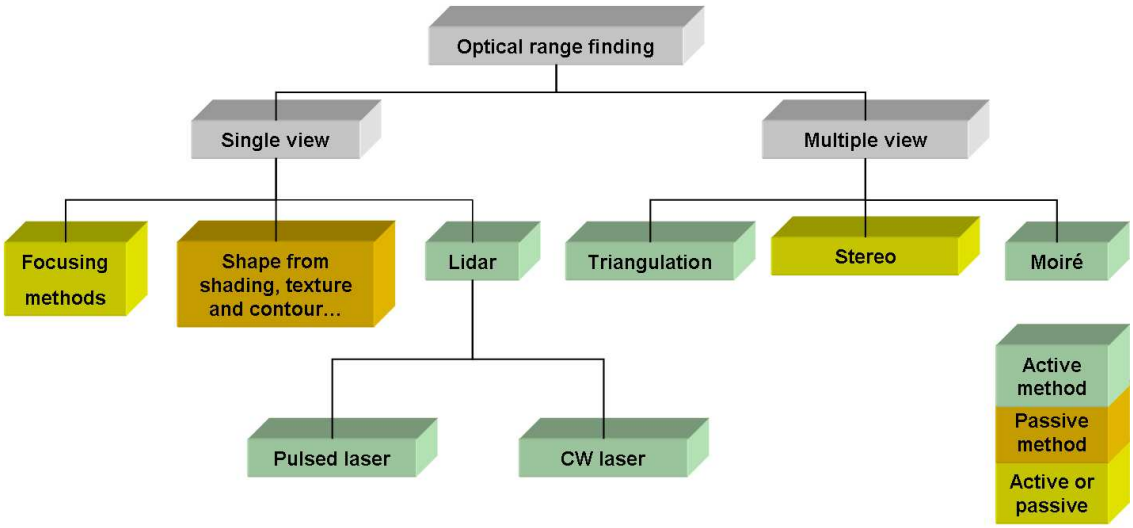



**Figure 2.1:** Optical range-finding techniques

## 2.1 3D Imaging Systems

Optical range-finding techniques for 3D imaging can be divided into several categories in terms of their operational principle. For instance, Amann *et al.* [4] define three categories: interferometry, time-of-flight and triangulation methods whereas Ailisto *et al.* [1] summarised these techniques according to a similar scheme to that of figure 2.1 However, many other classifications can be obtained using other general criteria. We partition optical 3D imaging systems using range-finding techniques into two main categories: *passive* and *active* 3D imaging systems.

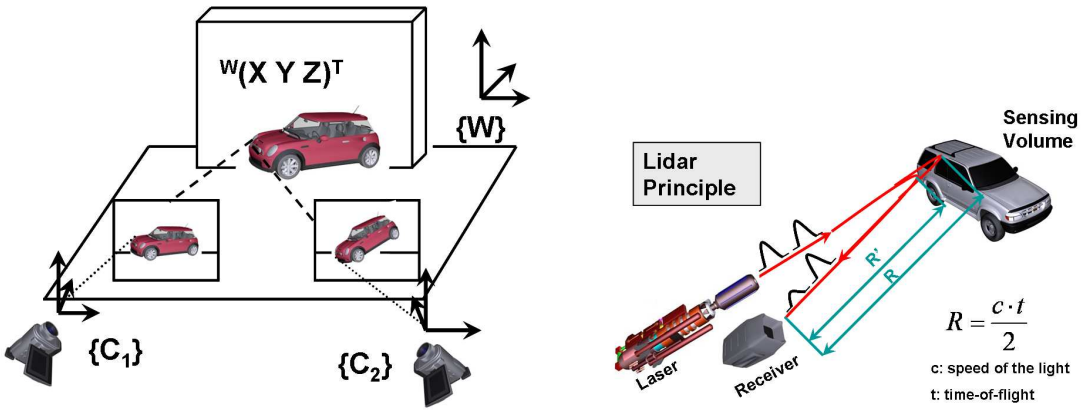
### 2.1.1 Passive 3D Imaging Systems

Practical *passive* 3D imaging systems rely on the acquisition and processing of two or more intensity images acquired from different viewpoints, using either the reflective properties of the scene objects in response to a natural source, *e.g.* the sun, or the self-emissive properties of the objects themselves, for example in an infra-red image. Provided corresponding points or features can be identified within the two or more images, it is possible to use the laws of projective (or other) geometry to recover 3D information from the scene. Passive techniques are important where there are circumstantial constraints that prevent the use of artificial energy sources of radiation. Passive techniques include *focusing methods*, *shape from X* and *(passive) stereo*, to cite a few.

Focusing methods include *depth-from-focus* and *depth-from-defocus*. In summary, these methods obtain depth information by actively controlling camera parameters. Depth-from-focus (see for instance [48]) determines distance to one point by taking many images of continuously improving focus. The best focus determines the pixel depth. However, the focus criteria as well as efficient peak detection from the focus criteria profile, are complicated. The depth-from-defocus method [50] uses the direct relationship between the depth, the camera parameters and the amount of blurring in images to derive the depth. Because the blurring in an image can be caused by either the imaging or the scene itself, at least two images taken under different camera configurations are generally required to eliminate the ambiguity. A fundamental problem in depth from defocus is the measurement of relative defocus between images. Passive versions of focusing methods are rarely used and active methods are preferred.

In computer vision, the techniques to recover shape are called *shape-from-X* techniques, where X can be shading, motion, texture, silhouette, etc. Shape-from-shading (SFS) [163] deals with the recovery of shape from a gradual variation of shading in the image. To calculate depth measurements, one studies the changes in brightness over a surface using constraints in the orientation of surfaces. Although this technique attempts to emulate the human capability to infer 3D shape from cues in single images, it is mathematically unstable. Moreover, the target shape cannot have texture since textures also cause luminance changes and shape-from-texture also has to be used although these techniques are not themselves not very robust. In general, methods based on these techniques are not appropriate for general 3D depth data acquisition since they are very dependent on the illumination and surface reflectance properties of objects present in the scene.

*(Passive) Stereo methods* [40] attempt to find matching image features, such as edges or points, between two or more general images about which nothing is known a priori to reconstruct the scene geometry. (Passive) stereo methods use two or more cameras to observe features on an object (see figure 2.2(a)). If the same feature is observed by two cameras, then the two lines of sight passing through the feature point on each camera's image plane will intersect at a point on the object. Typically, the 3D reconstruction is performed by triangulation of matched pairs. In general, these tasks are non-trivial since




(a) (b)

**Figure 2.2:** Some diagrams of different 3D imaging techniques (a) Passive stereo principle (b) Lidar principle

the correspondence problem is hard and computationally complicated. Obtaining high accuracy using these techniques is an ill-posed problem since no metric data is provided outside highly constrained scenarios. Therefore, these techniques are not appropriate when dealing with non-cooperative targets. Further, the necessary separation of the two (or more) cameras leads to problems of occlusion.

### 2.1.2 Active 3D Imaging Systems

Active 3D imaging systems use the projection of an energy source, such as a laser, microwave radar, millimeter waves or sonar, on to the scene. Active systems do allow the direct acquisition of 3D data, eliminate the need to both detect and correspond salient features in passively acquired images and usually can get dense data. In general, these methods are much more robust and accurate than passive techniques. The most common methods of active laser imaging include *Moire fringe methods*, *triangulation* of a laser source with a separated detector, and *light detection and ranging* (Lidar) techniques.

*Moire fringe methods* are based on the projection of a grid on to an object and an image is formed in the plane of some reference grating. Moire fringes are light/dark bands seen by superimposing two nearly identical arrays of lines and dots. The superimposition of the projected grid and the reference grid creates different patterns that give accurate descriptions of changes in depth. The first use of fringe projection for determining surface

topography dates from 1967 and was presented by Rowe and Welford [126] as an alternative to standard interferometry techniques. One of the main advantages of Moire methods is that they have a very high resolution and are therefore capable of producing very accurate depth data [108]. However, these methods have a very limited dynamic range. Indeed, as the angle from which the image is viewed increases, the fringe density becomes increasingly dense. Moreover, ambiguities can arise with depth discontinuities when identifying fringe patterns, that is, Moire methods can have phase discrimination problems when the surface does not exhibit smooth shape variations.

*Active stereo* methods are a variant of *passive* stereo methods in which one of the cameras is replaced by a light emitter. As in the passive case, finding the correspondences is difficult, even when taking into account epipolar constraints. Moreover, these techniques fail for surfaces with strong specular reflectance and suffer from low surface reflectance. To deal with the correspondence problem, methods based on *coded structured light* are envisaged (see for instance Salvi [127] for more complete discussion). *Coded structured light* methods are based on the projection of a coded light pattern on to the measuring space. The correspondence problem is solved locating such coded points in the image recorded by the camera avoiding the use of geometrical constraints.

Depth from triangulation [76] makes use of at least two known scene viewpoints. This technique is called triangulation because the laser spot, the camera and the laser emitter form a triangle. Corresponding features from the different viewpoints are identified, and rays are intersected to find the 3D position of each feature. In short, given the baseline of a triangle, i.e. the distance between two of its vertices and the angles at these vertices, the range from one of the vertices to the third one is computed as the corresponding triangle side. Triangulation is a mature technique, well suited to range-finding, profiling and 3D imaging for short range but is not usually considered a practical proposition to acquire 3D images of moving targets at long range. Since the resolution in these techniques is proportional to distance from target, it tends to become too coarse for longer ranges. Further practical problems include: there are practical problems due to the trade-off between the scale of the object measured, the stand-off distance and the accuracy of measurement; occlusion caused by the necessary separation of viewpoint between the transmitter

and receiver; and to poor and false returns caused by variable material reflectance. The requirements for the processing speed of the sensor, power dissipation or highly tuned precision moving parts [106] are also important disadvantages of these systems.

*Lidar* techniques [4], also known as optical radar and laser detection and ranging (LaDAR) systems techniques show a large dynamic range (from a few metres to several kilometres) and measurement uncertainties that are typically dependent of receiving a relatively strong signal from the target. These techniques generate angle-angle-range images where two dimensions come from the angular resolution of the imaging system and the third dimension results from performing a time-of-flight (TOF) range measurement for each pixel. In practice, the active light source and the receiver are located very close to each other. This facilitates a compact set up and avoids the occlusion problem found in triangulation techniques. To construct a 3D image a scanning system or a focal plane array must be used.

## 2.2 Time-of-Flight Systems

To get accurate depth values, Lidar techniques use the measurement of time-of-flight (TOF) [75] between a coaxial or near-coaxial transmitter and detector and the target (see figure 2.2(b)). TOF is a more practical proposition for active, target ranging and recognition. In order to measure time of flight, two different techniques can be employed *continuous wave* (CW) or *pulsed* laser range finding. The most common solutions in CW are to use *amplitude-modulated* CW (am/CW) or *frequency-modulated* CW (fm/CW).

The am/CW technique [135, 160] measures the TOF indirectly by measuring the shift in phase between an am/CW emitted beam and its reflection whereas the fm/CW technique [129,134] measures the TOF indirectly by measuring the beat frequency of a fm/CW emitted beam and its reflection. In the pulsed range-finding method, when a pulsed laser signal is directed towards and reflected from a target, the total TOF, $t$, determines the range from the transmitter/receiver. The range is found simply from $R = \frac{1}{2}\ ct$ where $R$ is the range, $c$ is generally the speed of the light (depending on the medium in which the pulse is propagating) and the term $\frac{1}{2}$ arises from the round trip of the pulse. Pulsed TOF

is in general fast (since no modulation and demodulation is required) and good accuracy can be obtained with single-pulse measurements. On the other hand, CW techniques are usually less expensive [1]. CW techniques require also a relatively powerful laser or light-emitting diode (LED). Moreover, the accuracy is also limited by the speed of the clock inherent in all modulation-based systems. However, fm/CW systems are generally more accurate than those based on conventional pulsed signals. To achieve (sub-)millimetric accuracy, pulsed TOF systems must have (sub-)picosecond time discrimination capability as well as low jitter[1] and low ray divergence. The main drawback of TOF ranging systems is that their accuracy is dependent on receiving a relatively strong signal from the target that can be detected and processed. The aim of this thesis is to exploit two techniques of TOF which have acquired particular relevance for the last years. These techniques are Time-correlated Single Photon Counting (TCSPC) and Burst Illumination Laser (BIL) imaging and will be explained in subsequent sections.

### 2.2.1 Detector Arrangements of TOF Systems

A straightforward method to acquire 3D information about a scene is to scan the collimated beam and the detector over the target. With every laser pulse a very small part of the object is illuminated and the TOF is stored. The amount of object illuminated depends on the divergence properties of the laser beam. In some detectors, the full waveform of the pulse is recorded or just the time of return, according to some threshold value determined by the experience of the user. Other systems record the first and last return, or even more returns, each of them representing different surfaces. The main current advantage of scanned systems over focal plane arrays is that they can achieve high angular resolution. On the other hand, single element detectors limit the speed of acquisition of the data and generally increase the weight, volume, power and cost of the receiver system. Pyburn *et al.* [115] show a practical implementation of a scanning system.

One alternative to scanning systems is a *burst illumination* or *gated viewing laser radar* [24, 44] that makes use (typically) of a fast-gated camera, an electron-bombarded charge-coupled device (CCD) or other camera, and "gates" the response from a divergent pulsed source impacting the target so that the returned signal is integrated over a narrow time

[1] Briefly, timing jitter is a fluctuation of the temporal position of pulses.

window (and hence depth of field). This has the advantages of providing both foreground separation and a crude measure of the range (again using the time-of-flight principle). If the time window is varied then it is possible to scan a depth "slice" through the target region to get a full 3D image.

Finally, in the last few years another alternative to the two previous systems has arisen. The principle is similar to burst illumination except that it uses a single short pulse (or flash) which illuminates a target area and an integrating focal plane array (FPA) detector using avalanche photodiodes (APD) in linear or geiger mode (known as single photon avalanch diodes (SPADs)). The time of flight is measured per pixel which delivers the third dimension (depth) of the surface of the object. Since the spatial position of each detector is known, the acquired depth value enables 3D imaging. The attractive feature of such systems is their potential to acquire entire 3D images on a single pulse at a high frame rate with a depth resolution of a few centimetres and small size, weight, and power requirements, using current technology. The sophistication of processing of the laser return is understandably limited by the need to combine the processing and detection circuitry. An example of an implementation of these systems can be found in the work of Halmos *et al.* [70].

Compared to CCDs, SPADs are faster, made with simpler technology, more robust and reliable, with high Quantum Efficiency [2] (QE), are intrinsically digital and are not affected by readout noise [3]. On the other hand, they have some disadvantages that have to be solved yet. For instance, *afterpulses* which can be thought as secondary false counts due to traps, *crosstalk*, a phenomenon in which a detector emits photons due to the electroluminiscence effect during the avalanche process that can migrate to neighbouring detectors, the *dark count rate*[4], the *dead time* (the time required for quenching the avalanche and restoring the operating bias) and device uniformity.

[2] The photon-to-electron conversion efficiency of a photoelectric detector
[3] Randomness introduced by the collection, amplification, and analog-to-digital conversion of signal data
[4] Pulses generated by the thermal characteristics of the detector and are not related to the light input

### 2.2.2 Problems Associated with TOF Systems

There are several factors that determine the timing resolution and hence distance accuracy of TOF ranging systems. These include the shape (rise time, full-width half-maximum) of the laser pulse, and the efficiency and response time of the photodetector. For signal conditioning and measurement, the linearity, or at least calibrated repeatibility of the ramp (or digital counter if used), minimisation of timing jitter and the resolution of the ADC and digital timing circuit must all be optimised. Assuming that these design considerations can be satisfied, existing TOF ranging systems are accurate to about $0.1 - 1mm$ in depth in restricted laboratory environments, depending on how the measurement is made [154]. However, this accuracy is dependent on receiving a relatively strong and consistent signal from the target. The intensity values of the signal depend on the laser reflectance which in turn depends on many factors, including: target range, angle of incidence, albedo [5], surface roughness, and shininess (how much energy the surface reflects specularly). While it is unlikely that all of these parameters can be determined from a limited number of reflectance measurements from an unknown obstacle, the model can be simplified for the purpose of target range estimation. There are two major types of noise that contribute to the reflected signal, making it hard to interpret. One is photon noise, which is the result of ambient (non-laser) photons seen at the photodetector. Photon noise is independent of range and reflectance properties of the target. Another type of degradation is attributable to electronic noise in the laser modulation and/or noise in the acquisition and signal processing electronics. The amplitude of this noise depends on signal amplitude.

One common problem that impacts on almost all laser rangefinders, and not only TOF systems, is the *mixed pixel* problem. This phenomenon is caused when the reflected energy comes from two or more reflecting surfaces. In this case, the estimated range can be between the range of different surfaces and therefore far from representative of the actual value. *Range-intensity crosstalk* is another common phenomenon whereby the time of detection of the returned signal is influenced by its amplitude and hence the reflectance properties and range of the target. Lower signal intensities decrease the signal-to-noise ratio and therefore increase the variance in measured range. *Temporal mixing* appears

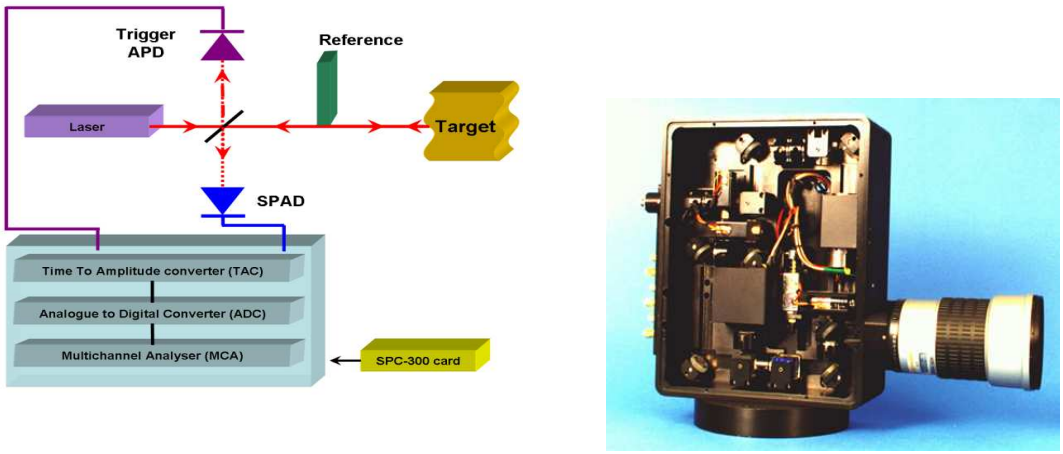
[5] ratio of the amount of electromagnetic energy reflected by a surface to the amount of energy incident upon it

when the laser is combined with a scanner and occurs within a single laser sample. This phenomenon is related to the emergence of amplitude jumps at edges between darker and lighter surfaces and leads to large errors in phase estimation. This phenomenon affects especially am/CW systems. A further, often unacknowledged problem, is *temperature sensitivity*. Changes of temperature causes laser frequency shifts that in turn cause the measured intensity of the signal to drop. This problem can be alleviated either by temperature control or by use of a reference channel for calibration.

## 2.3 Time-Correlated Single Photon Counting

Time-correlated single photon counting (TCSPC) [109] is a statistical sampling technique with single photon detection sensitivity, capable of sub-picosecond timing resolution. Typical applications of TCSPC include fluorescence lifetime measurements and fluorescence correlation spectroscopy (FCS) [10], optical tomography [9] and optical cryptography [82]. In the last few years TCSPC has also been applied successfully to 3D ranging and imaging [3, 7, 97, 154]. This technique offers several advantages when compared with previous methods for 3D ranging and imaging based on time-of-flight, i.e. very accurate time (and hence distance) resolution, very high sensitivity, and the ability to cope with variation in amplitude of the reflected return of several orders of magnitude. Although the TCSPC system we analyse in this thesis is based on pulsed TOF, it has been shown in recent work [117] that this technique can be used on chirped AM ladar.

Figure 2.3 shows a schematic diagram and a photograph of the first TOF-TCSPC sensor [97]. The laser source is a passively Q-switched AlGaAs laser diode (developed at the Ioffe Institute, St. Petersburg) which emits $10 - 20ps$ pulses of energy $\sim 10pJ$ at $850nm$ and a repetition frequency of up to $25MHz$. The major part of this laser signal is directed toward the target by a polarising beam splitter. The back-scattered light is captured by the input/ouput lens, a conventional telephoto lens for a single lens reflex camera. A fraction of the laser pulse is diverted to a reference channel, in this case one optical fibre from a set with appropriate lengths (delays). The received signals from the target and reference are combined and transmitted back through the series of wave plates and beam splitters to be imaged onto the actively-quenched single photon avalanche diode (SPAD).




(a) (b)

**Figure 2.3:** Schematic diagram (a) and photograph (b) of the TOF-TCSPC sensor

The conventional avalanche photodiode (APD) acts as a trigger to a single photon counting (SPC-300) electronics module. The temporal location of this reference pulse provides the START input for a time-to-amplitude converter (TAC). The signal from the SPAD provides the STOP input to the TAC, and it is these returns, from both the target and the fibre-optic delay (reference), that are used to build up a photon detection histogram of the type shown in figure 2.4. When scanned, it is possible to build up a full depth image of sub 100$\mu m$ resolution.

The SPAD (also known as Geiger-mode APD) is a key element of our system; the response consists of a fast rise and a slow, wavelength-independent, exponential tail. The devices used in this system were manufactured at the Politecnico di Milano and are representative of the state of the art. Measurement accuracy is affected by statistical fluctuations in the avalanche build up time which cause peak-width variations, the maximum electric field in the active junction which affects the time resolution, and thermally generated carriers in the depletion layer which result in an inherent dark count. The laser pulse width, the jitter of the detection system and size of the laser spot are also factors that determine the width of the pulse of figure 2.4. The tail is due to minority carriers, photo-generated in the neutral region beneath the junction, that succeed in reaching the depletion layer by diffusion. Figure 2.4 represents the number of detected photons (vertical axis) versus time (horizontal axis). The horizontal axis is defined as sampled bins (temporal channels); in

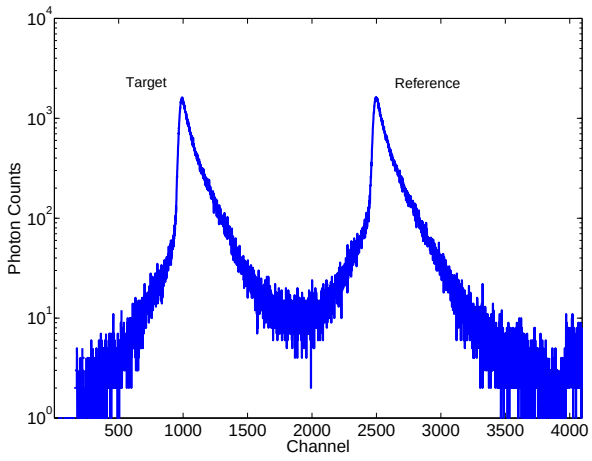
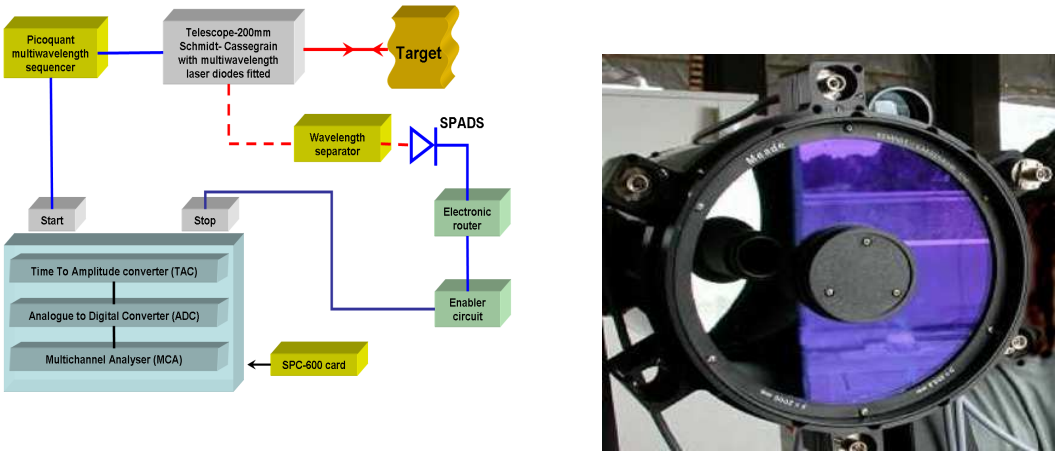



**Figure 2.4:** Photon count histogram built from repeated laser pulses

this example each bin corresponds to $2.44ps$. The two peaks correspond to the accumulated target (left) and reference (right) single photon returns. It is important to emphasise that each recording on the vertical axis of figure 2.4 represents a single photon event, i.e. the first received photon from either the target or the reference channel in response to a burst of laser energy in the form of a pulse. The origin of the background noise in figure 2.4 is either dark counts or ambient light. The average count rate is intentionally much less than one per pulse, ensuring (with correction for pulse pile-up[6]) equal probability of recording a photon arriving at any time during the interval. The number of photons collected depends on the laser repetition rate, the acquisition time and the properties of the optical system and target surface. Using the difference between the target and reference signals recorded by the SPAD to define the distance, rather than using the APD, enables us to balance any temporal drift in the SPAD signal channel.

Recently, the TOF-TCSPC technique has been expanded to detect remote targets at several kilometres in a multi-spectral system using six wavelengths in the visible to near infra-red wavebands [153], as illustrated schematically in figure 2.5. The sensor is designed to operate at ranges of several kilometres, and has six transmitting laser diodes of wavelengths $631nm$, $687nm$, $781nm$, $842nm$, $913nm$ and $972nm$ arranged in a ring around a receiving Schmidt-Cassegrain telescope that focuses the broadband return onto

[6] An effect that would lead to an over-representation of early photons in the histogram. It is due to dead times of detectors and electronics after a photon event. If the probability of detection was bigger than one, the system would very often register the first photon but miss the following one or more photons




(a) (b)

**Figure 2.5:** Schematic diagram (a) and photograph (b) of the telescope and transmitting diodes of the multi-spectral TOF-TCSPC system

an optical fibre mounted on a translation stage. The wavelengths are separated by an optical routing module into six distinct narrowband channels. Each channel is focussed in turn on to a channel-specific single photon counting avalanche diode (SPAD). The time multiplexing of the several channels is performed by a multi-wavelength sequencer. In this case the receiving SPADs are commercial Perkin-Elmer modules. The main advantage of a multi-spectral system is that it augments the limited reflectance information that can be obtained by integrating a single laser wavelength return by combining distance measurement with an assessment of the surface colour signature. For instance, it can be used in object classification since natural and man-made objects have different spectral signatures [90]. Johnson *et al.* [85] have also developed a multispectral system but instead of using a scanned single element they used of a focal plane array.

In recent years interest has grown in making 3D focal plane ladars using APDs in linear mode or SPADs in Geiger mode. Focal plane ladar systems are arrays of 3D sensing elements, usually on a single monolithic substrate and integrated with timing circuits at each pixel to give parallel range measurement. The main advantages and disadvantages of these systems have been described in section 2.2.1. Lincoln labs [7, 77, 94] have built arrays of up to 32 by 32 Geiger-mode APD elements (by 2003), connected to CMOS timing circuitry, using both silicon and InGaAs to extend the range beyond the visible

into the short-wave infra-red spectrum. For example, the GEN-III system uses a frequency doubled, passively Q-switched Nd:Yag laser operated at $532nm$; the return is focused on to a 32 by 32 Geiger-APD array and gives $15cm$ resolution in depth at a rate of $1000-5000$ $frames/s$. Raytheon [70], have used 32 by 32 HgCdTe arrays again with receiver on-chip circuitry (ROIC) to form 3D images in the wavelength $1.1-1.6\mu m$, operating in the linear (below breakdown) mode. The approach makes use of hybrid technology which adds noise in the fabrication process and difficulty in fabrication process.

Rockwell [150] has evaluated 64 by 64 arrays (which allows simultaneous range and intensity imaging) of both InGaAs ($1.5\mu m$) and HgCdTe ($3-5\mu m$) and 256 by 256 arrays (only range imaging), hybridized to a silicon detector, combined with ROIC consisting of a matched filter and comparator and achieved a poor resolution of $0.5m$ range depth discrimination. Zappa *et al.* [161] have tested a 60-element SPAD array (called SPADA) of $20\mu m$ diameter silicon SPADs with low dark-counting rate and low trapping effects that might lead to an improved focal plane array specification in future, and if combined with the type of processing developed previously for single elements give better range resolution. Niclass *et al.* [106] proposed a two-dimensional 32 by 32 array of single photon elements fabricated in standard CMOS technology which did not include any analog components and achieved millimetric depth resolution at low levels of light intensity. However, the measurements were carried out serially and it was designed for short-range applications. The terminology "flash ladar" is also used for the multiple-slit streak tube imaging systems developed by Arete [55]. The streak tube provides an alternative to solid state devices such as the SPAD in giving time-resolved measurements. These systems are usually pulsed at a single frequency and give a measure of reflectance as well as range, but there have been examples of continuously modulated focal plane ladars, *e.g.* that mix the received signals with the transmitted signals using a sawtooth chirp signal [135].

### 2.3.1 Time-of-Flight Depth Imaging Using TCSPC

The TCSPC system described here uses a single detector with a mechanical scanning device in which the estimation of the range is performed on a pixel-by-pixel basis. Classical approaches can be envisaged to analyse this data. For instance, the peak detector could be a simple threshold comparator as in the work of Steinvall and Carlsson [137] who used two

different simple algorithms to detect the return pulse: peak detection and leading-edge detection. Leading-edge detection is one which is often preferred because it has in general a smaller range walk than peak detection. However, for a leading-edge detector, in the event of a plane surface and oblique incidence, pulse stretching will cause a range error.

When the laser pulses are repeated many times, a histogram of the photon arrival times is built up. Thus, rather than simply measuring the peak positions, the histogram data is processed to obtain an accurate estimate of the time lag between the reflected target and reference signals in the histogram. The distance between the target and reference is $\frac{1}{2}c\tau$ where $c$ is the speed of the laser light and $\tau$ is the time lag. For instance, Umasuthan *et al.* [149] obtained accurate estimates of the peak position by iterative fitting of an operating model to either the raw histogram data or to the autocorrelation function and minimisation of an error metric. The operating model made use of Lorentzian functions and a (Hermite or Lagrange) polynomial function to make the operating model symmetric. The different parameters were determined by maximum likelihood estimation (MLE). The peak maximum can be obtained calculating the centroid value of the operating model. Since the number of terms in the polynomial function is finite, the polynomials have oscillations of large magnitude at high index values[7]. To avoid this phenomenon, they truncated the peak return. However, if a histogram contains multiple responses of quite different magnitudes the implementation of truncation function of a signal of high magnitude can obscure a genuine but much smaller response. Pellegrini *et al.* [109] calculated the distance between the reference and target peaks of each histogram using the autocorrelation method suggested by Umasuthan *et al.* [149]. However, they used a four piecewise exponential as an operating model. Although this approach avoids the problem of the polynomial functions, if fails when dealing with small returns.

In recent years, Wallace *et al.* [156] have developed a deterministic algorithm to deal with multiple returns originated from different target surfaces or semitransparent surfaces. They applied a non-parametric, bump hunting procedure that provided an initial estimate of the number, amplitude and positions of the suspected returns. This used Gaussian filters of progressively decreasing standard deviation. Once the initial estimate was obtained,

[7] This is similar to the Gibbs phenomenon arisen when approximating, for instance, a square function by its Fourier series

they computed a maximum likelihood estimate (MLE) of the set of parameters. Again, the main problem of this algorithm is that it may fail when the returns are hidden by the background noise.

### 2.3.2 TCSPC Imaging Advantages

Key advantages of TCSPC imaging systems include the following.

- Complete depth and surface reflectance profile of an object can be obtained (if the target is repeatedly pulsed).
- Very accurate time resolution: picosecond timing accuracy.
- Very high sensitivity: single-photon sensitivity accuracy.
- Sensitivity accuracy measurements at long range ( 17 km on corner cubes [156]).
- Time resolution is not limited by the width of the detector impulse response.
- No occlusion due to the coaxial alignment of the incident and return optical paths.
- No large return signal required.
- Operated at eye-safe wavelengths.

## 2.4 Burst Illumination Laser

Gated viewing of a distant target under laser illumination is usually referred to as "burst illumination laser" (BIL). A single laser pulse diverges and illuminates the scene within a time (distance) window determined from prior knowledge of the scene characteristics, *e.g.* a spot, lidar ranging system. Thus, only those objects that fall within a finite volume of the search space, the "range gate", are exposed. All returns from foreground/background objects are rejected whereas the atmospheric effects are diminished in the image capture process (see figure 2.4) which allows the extraction of the object of interest. For instance, a time gate of $20ns$ with $10\mu s$ delay after the laser pulse is launched will reveal a $3m$ deep range at a distance of $1.5km$.

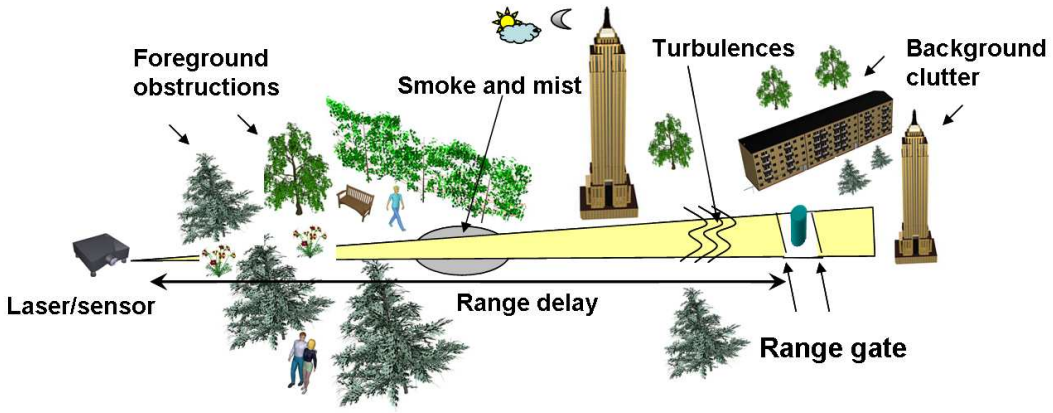



**Figure 2.6:** Basic principles of a BIL system

The images acquired by such as system are "flat" (2D intensity images), in the sense that they are usually illuminated by a coaxial light source so that the lighting is shadowless, but such images are also normally characterised by atmospheric turbulence effects and speckle due to the use of the coherent light source. Indeed, long-range target observation implies long path propagation of the laser beam through the atmosphere. Sometimes, it is necessary to explore forward propagation as the wavefront progresses towards the object and backward propagation as the light is reflected back to the receiver. The reason is that the paths for transmission and reception are different and therefore the interaction between signal and optical path is different in each case. In general, many studies agree that the main interactions with the atmosphere are: turbulence, atmospheric scattering and, when using systems with coherent detection, beam jitter. These facts lead to illumination of the target in a non-homogenous way and degradation of the observed image. Therefore, there is an increasing need to model the effects of the atmosphere in order to calculate the operating ranges of the systems. Churoux *et al.* [34] provided a model to evaluate the performance of BIL imaging through a turbulent atmosphere. This atmospheric turbulence, induced by thermal convection, creates scintillation in the target plane which has to be modeled statistically. Scintillation spatially modulates the illuminating amplitude of the target. In this model, atmospheric turbulence is incorporated in a more general transmission model as a Modulation Transfer Function (MTF) whereas atmospheric scattering acts only as a transmission coefficient. However, this model does not consider the effects of the speckle.

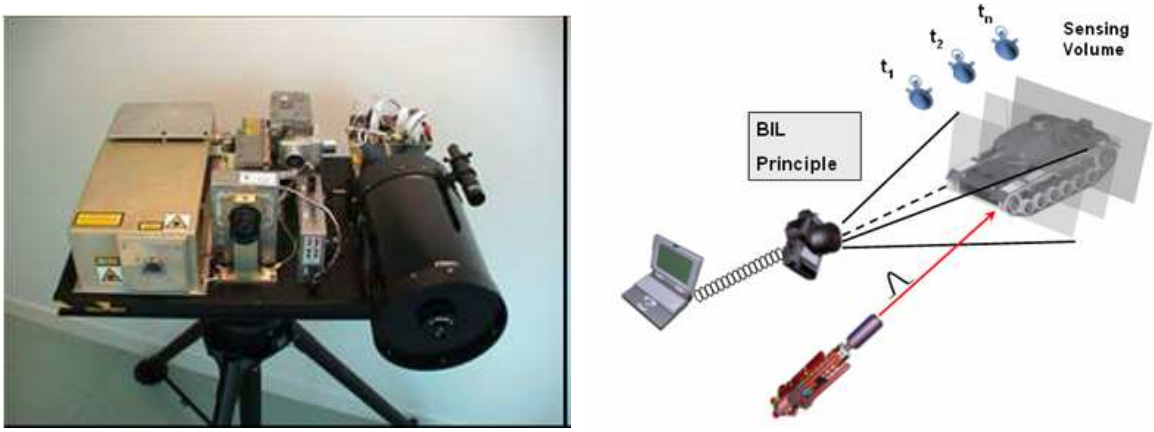
(a) (b)

**Figure 2.7:** Single frame BIL images (a) Short range (b) Long range

The basis of target speckle in the image plane of an optical system was discussed by Goodman [61] but he did not tackle the problem of a turbulent atmosphere. Speckle, an image artifact inherent in laser-illuminated imaging systems, results from interference patterns caused by the coherent illumination. Edouart *et al.* [47] developed a model which takes into account the target speckle and the turbulence perturbation. According to their work, the speckle pattern is characterised by the intensity correlation function which quantifies the size and the contrast of the granularities. The model aims to give this intensity correlation function in the image plane of a laser illuminating system of the combined effect turbulence-speckle. In more recent work, Duncan *et al.* [46], based on trials of their system, indicated an experimental value of the refractive index structure constant under the which the effects of the speckle become the limiting factor in the system performance against atmospheric turbulence. Laser speckle degrades target identification performance but can be reduced by averaging successive images [44].

Figure 2.7 shows two examples of a single frame of a BIL image acquired by the Selex Sensors and Airborne Systems (formerly BaE SYSTEMS Avionics) prototype system, shown in figure 2.8(a) and described in [8], at a relatively short range of a few hundred metres (figure 2.7(a)) and at several kilometres (see figure 2.7(b)). The intensity values are plotted on a log scale for convenience, as the divergent beam is non-uniform, causing fall-off in intensity at the edges of the image. This makes a linear intensity image more difficult to interpret. Other factors that influence the amount of reflected energy are the bidirectional reflection distribution function (BDRF) [8] and the angle towards the sensor.

[8]The BRDF gives the reflectance of a target as a function of illumination geometry and viewing geom-




(a) (b)

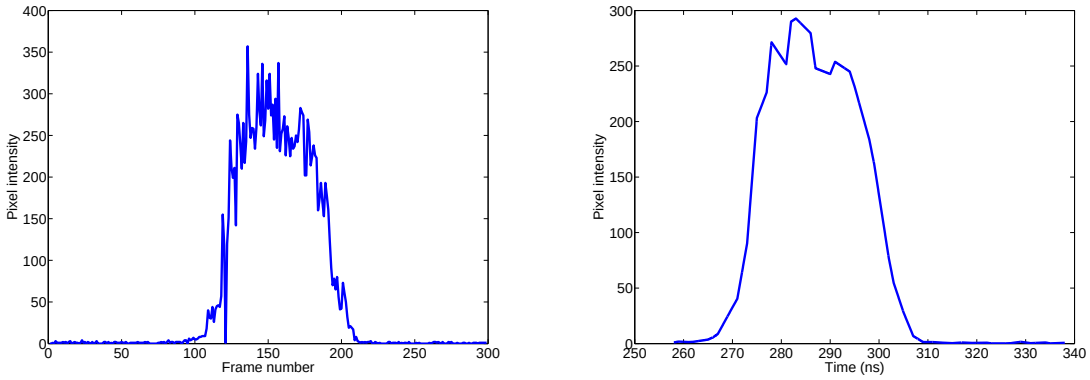
**Figure 2.8:** (a)Photograph of the BIL system (Courtesy of Selex S&AS) (b)3D imaging principle with a BIL system

### 2.4.1 Time-of-Flight Depth Imaging Using BIL

BIL imaging uses a pulsed laser which instantaneously illuminates a slice of space. A gated camera views this slice of space. A 3D image is created from multiple flashes by varying the delay between the transmit pulse and the fixed "range gate" continuously or in small increments as shown in figure 2.8(b). Successively moving the gate over a scene in steps along the range axis produces an intensity variation in each individual pixel (see figure 2.9(a)) creating an intensity profile. Since the return obtained is very noisy due to the effects of the atmosphere and the non-uniform properties of the laser, it has been shown that a multiple frame temporal averaging improves the signal-to-noise ratio and the intensity profile. The smoothing effect on the signal is easily seen in figure 2.9(b).

The resulting pixel intensity values for any given delay time are proportional to the convolution of the camera gain and the pulse return after the round trip. Any measured pixel intensity value can therefore be considered as a noise sample from the resulting curve of this convolution. The range is therefore evaluated separately for each pixel and is estimated by calculating the position of the maximum value of the convolution function from a set of noisy observations. One simple approach to determine range is to use a correlation detector of the returned signal and an instrumental profile which implies that only one

etry. The BRDF depends on wavelength and is determined by the structural and optical properties of the surface.




(a) (b)

**Figure 2.9:** (a)Variation of pixel intensity as function of frame number (b) Variation of the same pixel as a function of time (ns)

range measurement can be taken on a pixel. For instance, Murray *et al.* [105] implemented an algorithm called the "bin-balancing matched filter" (BBMFTM) which followed this principle. This method relied on the concept that balancing the integrated signal value across the bins provides a precise estimate of the pulse arrival time. Under the assumption of a-priori knowledge of the shape of the illumination pulse, this algorithm generated a processing kernel from the stored out-going pulse shape, applied the sequence to the time sequence data, and output range estimates for each target return appearing under each pixel. In order to use this algorithm, they assumed that the pulse shape of the ladar return was similar to the transmitted pulse shape. This does not always hold when the effects of atmosphere and noise are considered.

Other approaches include that of least-squares parameter fitting of a parametric model to the data [5]. This method requires that the parametric model is known a priori or approximated. However, the intensity profile is difficult to model due to factors such as the speckle, the atmosphere and pulse-shape distortions, *e.g.* jitter, asymmetries (different rising and falling time). Other methods include data feature positioning [5]. Methods in this category include the gate-edge detection method algorithm and the weighted-average method [24, 25]. The gate detection method calculates range values by determining the position of the rise or fall transitions of the gate in the data using local constant fraction thresholding. This algorithm is very sensitive to the noise and suffers severely from range-intensity crosstalk. In the weighted average method, the range is calculated as the

weighted average of the depth of the gating (intensity) profile. The main drawback of this technique is that the range value is "shifted" if, for instance, the returned pulse saturates the detector element. Nevertheless, there have been attempts to model the depth of gating (intensity) profile using rectangular pulses to improve the range accuracy by correcting the average range [23]. However, none of these methods solves the problem of multiple returns within the instantaneous field of view (IFOV) of individual pixel elements from different or semitransparent surfaces. Indeed, when using these techniques, the assumption of a single return will shift the range value and therefore give a wrong estimation of the range. In conclusion, we believe that a fully stochastic approach to the problem of BIL imaging will give more accurate results than the previous methods. A stochastic approach will be better able to represent the uncertainties associated with the model and to determine the unknown number of multiple returns.

### 2.4.2 BIL Imaging Advantages

The key advantages of BIL imaging systems can be summarised as follows [33, 34]

- Reduced atmospheric scattering rejection in adverse weather conditions.
- The small dependence on the environment radiance variations (active system).
- Good angular resolution due to short wavelength.
- Reduced scattering effect linked to short wavelengths.
- Reduced line of sight stabilization requirements during imaging acquisition.
- Complete depth and surface reflectance profile of an object can be obtained.
- No need for a scanning device.
- Long range application ($\sim 20\ km$)
- Low amount of clutter due to control of the depth of field.

## 2.5 Conclusions

In this chapter, we have provided a review of existing laser ranging and 3D imaging systems. Particular emphasis has been placed on two Time-of-Flight (TOF) systems, Time-correlated Single Photon Counting (TCSPC) and Burst Illumination Laser (BIL) imaging, that offer many advantages in terms of accuracy and acquisition speed compared to other techniques. These techniques have been described alongside current signal and image processing algorithms used to extract depth information. In all cases, the depth resolution can be improved by post-processing of the data, provided that the time and circuit/software complexity can be tolerated. However, current processing techniques are inadequate for processing these data. For example, in the event of a far or non-cooperative target, the return signal may be small, comparable to the background noise, and fall below a threshold level. In that case, current techniques may fall to detect it. Furthermore, if the return is composed of multiple peaks coming from different surfaces, discriminating the peak of interest or inferring the unknown number of peaks is not trivial. In addition, one would like to quantify the uncertainty in any range estimate. Current algorithms do not provide such information as only point estimates of the range are provided. Moreover, a stochastic approach may also allow processing of spatial data by incorporating constraints between pixels in a neighbourhood.

Finally, similarities can be drawn between the processing of the photon count histograms when repetitively pulsing a single SPAD element, between the photon count data accumulated on single pixels in Geiger-mode APD arrays, and on the integrated pixel response profile to the laser flash in a gated system as the stand-off distance (i.e. gate delay) is varied. These can be used to provide a unified theory of pixel processing.

Chapter 3

# Bayesian Methodology

*Statistical thinking will one day be as necessary for efficient citizenship as the ability to read or write*

*H. G. Wells.*

In chapter 2 we gave a general description of 3D ranging and imaging systems. As new technologies emerge, the amount of data to be processed increases considerably. Moreover, the dimensionality of the problem also grows if one incorporates spatial and spectral constraints in the problem. Although the inclusion of these constraints will lead to a more accurate analysis of the scene of interest, the complexity of the problem makes the search for a satisfactory solution with the techniques discussed in Chapter 2 discouraging. Instead we shall adopt a Bayesian approach which, we hope, can give satisfactory answers to these problems. Bayesian inference is a powerful and increasingly popular statistical approach, which allows one to deal with complex problems in a conceptually simple and unified way whilst providing a quantitative assessment of uncertainty in the results inferred.

This chapter is completely devoted to the Bayesian framework although some comments to non-Bayesian techniques are given for the sake of clarity. In this chapter we detail in a general way the principles of the Bayesian paradigm and how it is used to carry out intensive computations. Then, we describe Markov chain Monte Carlo simulation methods as a statistical tool to deal with inference problems that were formerly intractable. Subsequently, mixture distributions in the context of a Bayesian approach are also reviewed due to similarity with our approach. Finally, a brief introduction to spatial modelling in Bayesian context is also given as most environments of interest contain spatial patterns by models that capture dependencies among elements.

## 3.1 The Bayesian Approach

There are many different definitions for the process of statistical inference. For example, in 1962 Savage stated:

> *By [statistical] inference I mean how we find things out - whether with a view to using the new knowledge as a basis for explicit action or not - and how it comes to pass that we have often acquire practically identical opinions in the light of evidence*

whereas Bernardo and Smith [11](p. 6) view statistical inference as a particular form of a decision problem:

> *[Statistical Inference]...a decision problem where an action corresponds to reporting a probability belief distribution for some unknown quantity of interest.*

In general inference can be seen as the process whereby we alter our beliefs regarding the world on the basis of observations. In the computer vision and image processing communities we can find many situations where statistical inference plays a very important role such as detection and tracking of targets, image restoration, 3D reconstruction and so forth. A key element of statistical inference is probability theory which will allow us to formulate a model that we hope is adequate to describe the situation of interest and to quantify uncertainty in the conclusions we extract from observations. It is in this latter aspect -the extraction of conclusions- that the two main philosophies of inference, the frequentist approach and the Bayesian approach, differ from each other[1].

In the work presented in this thesis we consider primarily a Bayesian approach to the problem described in the previous chapter, but will also use frequentist techniques to quantify performance obtained with the Bayesian techniques. Since the foundations of

[1] In http://www.statisticalengineering.com/frequentists_and_bayesians.htm, Charles Annis presents in a very peculiar point of view the difference between both philosophies. "*A frequentist is a person whose long-run ambition is to be wrong 5 % of the time whereas a Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.*"

our methodology is based on Bayesian inference, we devote a large part of this chapter to explain its philosophy and methodology. Nevertheless, for the sake of clarity, we examine briefly alternative approaches using the frequentist framework and we discuss the advantages and disadvantages of both approaches. A comprehensive general introduction to frequentist techniques can be found in [28] whereas we refer the reader to [11] for a fully comprehensive introduction to the Bayesian approach.

The Bayesian approach considers the uncertainties associated with all unknowns whether they are observed or unobserved and all forms of uncertainty are expressed in terms of probability distributions. The theory underlying this approach begins with the formulation of a mathematical model, with parameter vector $\theta$, for the physical process from which observations, $\boldsymbol{y}$, (are assumed to) arise. In brief, this model specifies the distribution of $\boldsymbol{y}$ given $\theta$. In the Bayesian paradigm the information brought by the data is combined with prior information on the unknowns parameters of the model and summarised in a probability distribution called the posterior distribution which is the object of all Bayesian inference. This posterior distribution represents the totality of knowledge about $\theta$ after observing $\boldsymbol{y}$ and is obtained through Bayes theorem as

$$\pi(\theta|\boldsymbol{y}) = \frac{\pi(\theta)L(\boldsymbol{y}|\theta)}{\int \pi(\theta')L(\boldsymbol{y}|\theta')d\theta'} \tag{3.1}$$

where $\pi(\theta)L(\boldsymbol{y}|\theta)$ is the joint probability distribution of the random quantities $y$ and $\theta$, $L(\boldsymbol{y}|\theta)$ the likelihood function, $\pi(\theta)$ the prior distribution on the parameter $\theta$ and $\int \pi(\theta')L(\boldsymbol{y}|\theta')d\theta'$ is a normalising constant known as the marginal density of $\boldsymbol{y}$. Whereas the likelihood function is a key element in both Bayesian and frequentist statistics, the prior distribution is the major source of controversy. Indeed, in both approaches, the likelihood principle, i.e. all the information that the data tell you about the parameters is embodied in the likelihood, can be respected. However the fact that the prior represents subjective beliefs in the Bayesian approach attracts criticism since these priors can influence and bias the conclusions. The influence of the prior on the conclusions we obtain will be explained in the following chapters. Note that the incorrect term "subjectivist" is often employed to label Bayesians. The choice of the mathematical model is also subject to subjectivity and is an intrinsic part of both approaches.

Once the posterior $\pi(\theta|\boldsymbol{y})$ is obtained, any features of interest such as moments, quantiles, etc. can be considered for a Bayesian approach. All these quantities can be expressed in terms of posterior expectations of functions of $\theta$. For instance, suppose we want to obtain a point estimate of $\theta$, to that we can calculate the posterior mean $E[\theta|\boldsymbol{y}]$, the posterior mean, the posterior mode, ..., or if we are interested in a measure of uncertainty we might calculate the posterior standard deviation $Var[\theta|\boldsymbol{y}]$. Further, inference on individual parameters or subsets of parameters can be obtained from marginal densities. In our case, we use the information obtained from the posterior distribution to infer the parameters which better explain either a photon count or intensity histogram. However, exploring posterior distributions can be very complicated particularly in high-dimensional spaces if they do not have a closed-form[2] solution and generally cannot be solved analytically. Solutions to this problem include numerical integration methods, *e.g.* Gaussian quadrature [53], Laplace approximations whereby analytic approximations are performed (see [143] and [130]) or the use of conjugate prior densities. These are prior distributions for which the posterior distributions derived are also members of the original prior family. For example, the gamma distribution, $\Gamma(\alpha, \beta)$ is the conjugate prior for the parameter $\lambda$ of an exponential distribution, $Exp(\lambda)$. Given $n$ independent and identically distributed samples $x = (x_1, \cdots, x_n)$, the resulting posterior distribution has the same algebraic form as the prior (with different parameters), $\Gamma(\alpha+n, \beta+\sum_{i=1}^{n} x_i)$. Nevertheless, the use of this kind of prior does not allow great flexibility on the use of the available information(see [121] for more information on these topics).

The main advantages and disadvantages of the Bayesian approach can be summarised as follows. On the one hand, the Bayesian approach facilitates representing and taking fuller account of uncertainties related to model and parameter values and provides a natural method for updating beliefs in response to new information and therefore to incorporate prior knowledge into the analysis. On the other hand, the main drawbacks are the potential sensitivity of the results to the choice of the prior and difficulty in extracting meaningful information from the posterior distribution and translate it into conclusions regarding the situation of interest. Whereas the the prior is usually chosen by convention, experience of the scientist or by analysis of information from synthesis studies (see [86] for a review on

[2]Succinctly, we say that an equation has a closed-form solution, if and only if, at least one solution can be expressed analytically in terms of a bounded number of certain standard functions.

the selection of prior distributions), powerful simulation algorithms such as Markov chain Monte Carlo(MCMC) techniques are needed to explore complex posterior distributions. MCMC techniques have enabled many Bayesian problems that were previously intractable to be solved. Essentially, they allow posterior distributions to be explored by simulation and conclusions and inferences to be drawn directly from the sampled values.

## 3.2 Stochastic Computational Methods for Exploring Bayesian Posteriors

In this section, we introduce Markov chain Monte Carlo(MCMC) algorithms as a computational tool for Bayesian inference which enable to explore complex posterior distributions as those arisen when dealing with lidar data. We begin by describing its constituent parts: Monte Carlo methods and Markov chains. We then describe the general form of MCMC given by the Metropolis Hastings, and a special case: the Gibbs sampler. Some examples of these algorithms are also provided. Finally, we describe other alternative samplers needed to understand other approaches to the lidar data analysis problem.

### 3.2.1 Monte Carlo Methods

Monte Carlo methods entail experimental simulations of mathematical or physical problems based on sampling random values. Methods based on this principle of random sampling are required when the dimensionality and/or complexity of a problem make straightforward numerical solutions impossible or impractical. In Bayesian inference, these methods are often needed to carry out the integration required when computing marginal densities. Compared to other techniques of numerical integration, such as the trapezoidal rule or Simpson's method, Monte Carlo techniques are more effective since they do not suffer from the *curse of dimensionality*, i.e. the number of computations required does not increase exponentially with the dimensionality of the integral. The fundamental disadvantage of simple Monte Carlo integration is that its accuracy increases only as the square root of the number of sampled points. However, Monte Carlo techniques have other attractive features that we summarise below and are studied more extensively in texts such as [54].

(i) The strong law of large numbers ensures that the approximation becomes better as the number of simulated values increases.

(ii) The central limit theorem can be applied at any stage of the simulation process to calculate the approximation error.

The only requirement for using this technique is that the physical (or mathematical) system has to be described by a probability density function (pdf), $\pi(\cdot)$. Once the pdf are known, the Monte Carlo simulation can proceed by drawing samples $\{X_t,\ t = 1, \cdots, n\}$ from $\pi(\cdot)$. Many simulations are performed and then we can estimate the population mean of a function of interest, $f(x)$, as an average over the number of observations as follows:

$$E[f(X)] \approx \frac{1}{n} \sum_{t=1}^{n} f(X_t) \tag{3.2}$$

If the samples $\{X_t\}$ are independent, the law of large numbers holds and the approximations can be made as accurate as desired by increasing the number of simulated values. In many practical applications, one can predict the statistical error (the "variance") in this average result, and hence an estimate of the number of Monte Carlo trials needed to achieve a given error is obtained.

Monte Carlo analysis can also be used to provide accurate and reliable statistics for the parameters of a model that has been fitted to an experimental dataset. We can use this approach to identify a set of parameters which maximise the posterior distribution of interest. For example, suppose we want to identify $\max_{\theta \in \Theta} \pi(\theta)$; the most basic approach to its resolution is to sample from a uniform distribution on $\Theta, u_1, \cdots, u_m \sim U_\Theta$, and approximate the solution as $h_m^* = max(\pi(u_1), \cdots, \pi(u_m))$. This method converges in probability (as $m$ goes to $\infty$), but it may be very slow and it does not consider any specific feature of $\pi$.

### 3.2.2 Markov Chains

In general, drawing samples independent $\{X_t\}$ from $\pi(\cdot)$ is not feasible since $\pi(\cdot)$ can be very complex. However, the $\{X_t\}$ need not necessarily be independent. Markov chains allow handling of complicated distributions such as those arising in the large dimensional

models of the type we anticipate when processing lidar or BIL data when several scattering target surfaces are present. Markov chains theory, allow us to construct a Markov chain with the property that its limiting, invariant distribution, is the target distribution $\pi(\cdot)$. Thus, under some conditions, the values of the Markov chain generated can be considered as a dependent samples generated from $\pi(\cdot)$ and can be used in equation 3.2.

A Markov chain, $\{X_t\}$, on a general state space $E$ is a sequence of random variables $\{X_t,\ t = 1, \cdots, n\}$ which evolves in discrete time constructed from a *transition kernel* $K$.

**Definition** The *transition kernel* [121], $K$, is a function defined on a sampled space $\Omega$ and equiped with a Borel *sigma*-algebra $\mathcal{B}$ such that

(i) $\forall\ x \in \Omega$, $K(x, \cdot)$ is a probability measure.

(ii) $\forall\ A \in \mathcal{B}(\Omega)$, $K(\cdot, A)$ is measurable.

A chain is time-homogeneous, if the transition probabilities are time-independent. That is, the conditional probabilities of state $(t+1)$ giving the state $(t)$ has a prescribed value independent of time instant. In this study, we consider time-homogeneous chains since the standard convergence properties do not apply to time-heterogeneous chains. The distribution of a time-homogeneous Markov chain on $E$ is specified by its initial distribution and its *transition kernel*. The transition kernel embodies the Markov property in time, that is, given the present state of the chain, $X^{(t)}, (t \geq 0)$, the future state is conditionally independent on the past states. More formally[3],

$$K(X^{(t+1)} \in A | X^{(t)} = x, X^{(t-1)} \in A_{t-1}, \cdots, X^{(0)} \in A_0) = K(X^{(t+1)} \in A | X^{(t)} = x) \quad (3.3)$$

for all sets $A_0, \cdots, A_{t-1}, A \subset E$ and $x \in E$. The evolution of the Markov chain through one time step, $t > 1$, is described by

$$K^t(x, A) = \int_E K^{t-1}(y, A) K(x, dy)$$

where $K^{(1)}(x, dy) = K(x, dy)$ and

$$K(x, A) = \int_A K(x, dy) \quad (3.4)$$

[3]To simplify we write $K(x, A) = K(X^{(t+1)} \in A | X^{(t)} = x)$

**Definition** A Markov chain is said to be $\varphi - irreducible$, if for any point $x$ and any measurable set $A$ such that, given a nonzero measure $\varphi$ on $E$, $\varphi(A) > 0$, there exists an integer $n \in \mathrm{N}$ such that $K^n(x, A) > 0$.

We are interested in finding conditions under which the $t^{th}$ iterate of a transition kernel converges to the limiting distribution $\pi$ as $t \to \infty$. To that, we need only verify that the following conditions hold.

1. Invariance. The invariant(or stationary) distribution is one that satisfies
$$\pi(B) = \int_E K(y, B)\pi(dy), \forall B \in \mathcal{B}(E) \tag{3.5}$$
2. The Markov chain is irreducible. This means that it is possible to reach all points in the state space in the support of $\pi$ within a finite number of transitions.
3. The Markov chain is aperiodic. That is, the state space cannot be partitioned into disjoint subsets $E_1$,$E_2$,$\cdots$,$E_m$ for which the Markov chain is constrained to cycle through these subsets from iteration to iteration.

In other words, the invariant probability distribution of a Markov chain is the unique and limiting distribution $\pi$ such that if $X^t \sim \pi$, then $X^{t+1} \sim \pi$, providing that the Markov chain is *irreducible* and *aperiodic*.

One way to obtain an invariant distribution is to use reversibility. The reversibility condition is stronger than equation 3.5. Note that reversibility is a sufficient but not necessary condition for the Markov chain to converge to the invariant distribution. A Markov chain is said to be reversible if it satisfies the *detailed-balance condition*, which is expressed in general form as
$$\pi(dx)K(x, dy) = \pi(dy)K(y, dx) \tag{3.6}$$
The term "balance" is used because it equates the rates of moves through states, and detailed because it does this for every possible pair of states.

The property of *irreducibility* is a first measure of the sensitivity of the Markov chain to the initial conditions. Irreducibility also ensures that most of the chains involved are

recurrent. *Recurrence* is the property that the average number of visits to an arbitrary set $A$ is infinite, at least from almost all starting values and is a sufficient condition to imply convergence of average probabilities, that is, to ensure that a law of large numbers holds for a Markov chain. Moreover, if the chain is *Harris recurrent* then the chain has the same limiting behavior for every starting value.

Given the previous definitions, we state the following theorems from Tierney [142] which establish the links between Markov chains on continuous state spaces and the theory of MCMC methods. Indeed, the first theorem gives conditions under which the strong law of large numbers hold whereas the second one gives conditions under which the distribution of the chain converges to the unique invariant distribution.

**Theorem 1** *Suppose $\{X_n\}$ is an irreducible Markov chain with transition kernel $K$ and invariant distribution $\pi$, then $\pi$ is the unique invariant distribution of $K$ and for all $\pi$-integrable real-valued functions $h$,*

$$\frac{1}{M}\sum_{i=1}^{M} h(X_i) \rightarrow \int h(y)\pi(dy) \text{ as } M \rightarrow \infty \text{ almost surely}$$

**Theorem 2** *Suppose $\{X_n\}$ is an irreducible, aperiodic Markov chain with transition kernel $K$ and invariant distribution $\pi$. Then, for $\pi$-almost all $x \in E$*

$$||K^n(x,\cdot) - \pi(\cdot)|| \rightarrow 0 \text{ as } n \rightarrow \infty$$

*where $||\cdot||$ is known as the total variation distance defined as $||\mu_1 - \mu_2|| = 2\sup_{A\subset E} |\mu_1(A) - \mu_2(A)|$, $\mu_1$ and $\mu_2$ are probability distributions.*

To provide a central limit theorem, stronger conditions are needed and requires that the chains used are ergodic, i.e., the chains must be irreducible, aperiodic and Harris recurrent. These stronger conditions are usually stated in terms of rate of convergence. In general, geometric ergodicity and uniform ergodicity are used to state the existence of a central limit theorem (see for instance [60]). Note that if the central limit theorem holds for the invariant distribution, it also holds for all initial distributions.

### 3.2.3 Markov Chain Monte Carlo Algorithms

Given data $y$, Bayesian inference may be performed using Markov chain Monte Carlo (MCMC) methods which involve the construction of a Markov chain, $\{X^{(n)}\}$, with lim-

iting distribution our target distribution of interest $\pi$ and use these simulated values for obtaining features of $\pi$ of interest. For example, we might be interested in estimating the expectation of some function $f_{int}$:

$$E_\pi[f_{int}(X)] = \int f_{int}(x)\pi(dx) = \mu \tag{3.7}$$

Then, the MCMC estimate of $\mu$, $\hat{\mu}$, will be the *sample path average*

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f_{int}(X_i) \tag{3.8}$$

Note, that the use of a chain from a MCMC algorithm with limiting distribution $\pi$ is similar to use of an independent and identically distributed sample (iid) from $\pi$ in the sense that the ergodic theorem guarantees the convergence of the empirical average. Two of the simplest recipes to this are the *Metropolis-Hastings algorithm* and the *Gibbs sampler*.

#### 3.2.3.1 The Metropolis-Hastings Algorithm

Originally developed by Metropolis *et al.* [101] as an approach to statistical mechanical problems and then generalized by Hastings [73], the Metropolis-Hastings algorithm is an easy method of constructing an homogeneous Markov chain whose limiting distribution is our distribution of interest. Metropolis-Hastings algorithms are based on a Markov chain whose dependence on the predecessor is divided into two parts: a proposal and an acceptance of the proposal. The proposal suggests an arbitrary next step in the trajectory of the chain and the acceptance makes sure the appropriate limiting direction is maintained by rejecting unwanted moves within the chain.

The key idea is to find a transition kernel $K$ of a Markov chain whose invariant distribution is our target distribution $\pi$. One way to obtain an invariant distribution is to use reversibility as explained in previous sections. Suppose the current state of the Markov chain is $x$ and we propose a move $x'$, a Markov chain is said to be reversible if satisfies the detailed-balance condition, that is,

$$\pi(x)K(x, x') = \pi(x')K(x', x) \tag{3.9}$$

In the Metropolis-Hastings paradigm, the kernel is composed of two elements: an arbitrary transition probability, $q(\cdot, \cdot)$, known as the *proposal distribution* and $\alpha(\cdot, \cdot)$ known as the

*acceptance probability.* The probability of a transition from state $x$ to state $x'$ is

$$p(x, x') = q(x, x')\alpha(x, x'), \forall\ x \neq x' \tag{3.10}$$

whereas the probability of remaining in the same state is expressed as

$$r(x) = 1 - \int q(x, x')\alpha(x, x')dx' \tag{3.11}$$

Thus, the formal specification of the Metropolis-Hastings kernel is

$$K(x, x') = p(x, x') + r(x)\delta_x(A) \tag{3.12}$$

where $\delta_x(\cdot)$ is the mass function.

Using equations 3.4, 3.9 and 3.12, it is easy to show that the Markov chain generated is invariant since

$$\begin{aligned}
\int K(x, A)\pi(x)dx &= \int \left\{ \int_A p(x, x')dx' \right\} \pi(x)dx + \int r(x)\delta_x(A)\pi(x)dx \\
&= \int_A \left\{ \int p(x, x')\pi(x)dx \right\} dx' + \int_A r(x)\pi(x)dx \\
&= \int_A \left\{ \int p(x', x)\pi(x')dx \right\} dx' + \int_A r(x)\pi(x)dx \\
&= \int_A \{1 - r(x')\}\pi(x')dx' + \int_A r(x)\pi(x)dx \\
&= \int_A \pi(x')dx'
\end{aligned} \tag{3.13}$$

The procedure of the Metropolis-Hastings algorithm can be explained as follows, at each time $t$, we propose a new state $x'$ for the Markov chain drawn from a proposal distribution $q(\cdot, x)$. Then, the candidate point is accepted with a probability $\alpha(x, x')$. Different possibilities exist for the choice of the acceptance probability. For instance, Hastings [73] proposed the use of the following general expression

$$\alpha(x, x') = \frac{s(x, x')}{1 + \frac{\pi(x)q(x', x)}{\pi(x')q(x, x')}} \tag{3.14}$$

where $s(x, x')$ is a symmetric function such as $\alpha(x, x') \leq 1$. However, Peskun [110] showed that it is optimal to choose the function $s(x, x')$ as large as possible to reduce the asymptotic variance of estimates, the expression of the acceptance probability is

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')} \right\} \tag{3.15}$$

which is the expression that we use throughout this thesis. Other examples of acceptance probability expressions can be found in [147] and the references therein. An important feature of this algorithm is that is based on ratios of densities and therefore the target density need to be known only up to a normalising constant. Moreover, there are no restrictions in the choice of the proposal distribution, $q(\cdot, \cdot)$, other than irreducibility since, independently of it, the stationary distribution of the chain will be our target distribution. Simple conditions of convergence of this algorithm are given by Roberts and Smith in [124] whereas Roberts and Tweedie [125] show results on geometric convergence and central limit theorems for multidimensional Metropolis-Hastings algorithms.

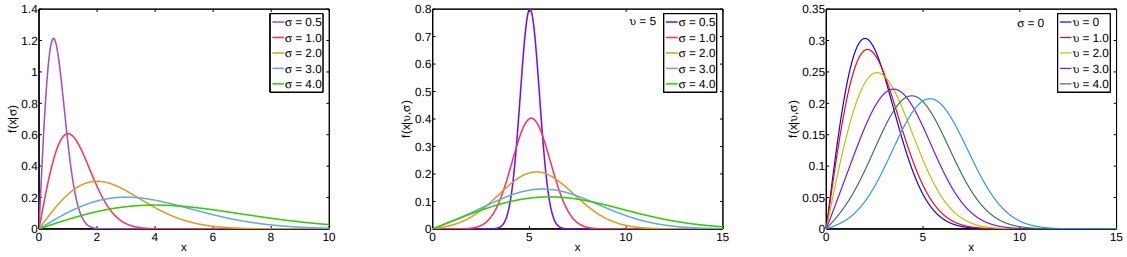
**Algorithm 1 (Metropolis-Hastings)**

1. *Initalize* $x_0$ *and set* $t = 0$
2. *For* $t = 1$ *to maximum number of iterations*
   3. *Sample* $x'$ *from* $q(\cdot, x_t)$
   4. *Sample* $u$ *from* $U(0, 1)$
   5. *If* $u \leq \alpha(x_t, x')$ *we accept the current move and set* $x_{t+1} = x'$ *otherwise we stay at the current state and set* $x_{t+1} = x_t$

It is important to note that a sequence $(X(t))$ produced by a Metropolis-Hastings algorithm is not an independent and identically distributed (iid) sample of the target distribution. Therefore, there is a certain degree of correlation between observations but, by virtue of the ergodic theorem, the sample averages will, for large sample size, provide good estimators of the expectation.

**A Toy Example**

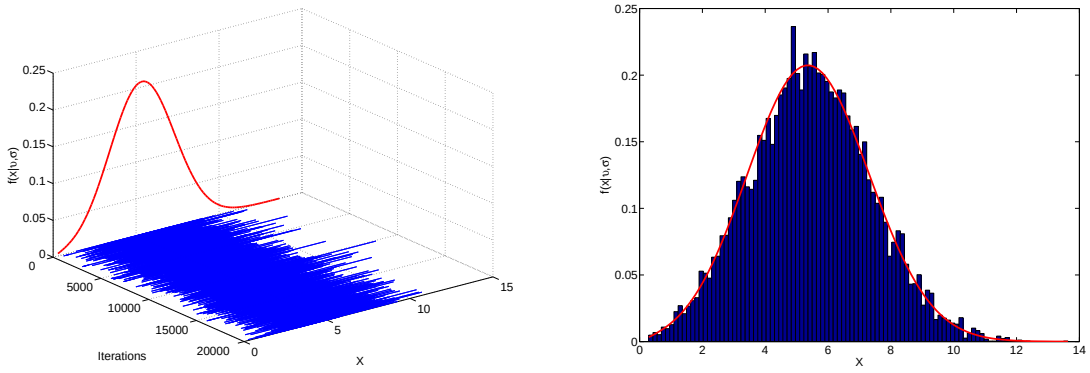
One of the the applications of the Metropolis-Hastings algorithm is to generate a sequence of samples from a probability distribution that is difficult to directly sample from. To show the performances of the algorithm, we show an example on how to sample from a Rice distribution. Rice distributions (figures 3.1(b) and 3.1(c)) are used in wireless communications as a mathematical models for the distortion that a carrier-modulated telecommunication signal experiences over certain propagation media. Concretely, these distributions arise in situations in which radio signals reach the receiving antenna by two or more paths. If there is a dominant signal arriving with many weaker multipath signals, the




(a) (b) (c)

**Figure 3.1:** (a) Rayleigh pdf (b) Rice pdf, $\upsilon$ fixed (b) Rice pdf, $\sigma$ fixed

fading envelop distribution is Ricean. If there is no dominant signal the Rice distribution degenerates to a Rayleigh distribution (figure 3.1(a)).




(a) (b)

**Figure 3.2:** Metropolis-Hastings algorithm, a toy example (a) Rice pdf (red) and Markov chain generated (blue) (b) Normalized histogram of samples generated and Rice pdf (red)

Suppose that we want to sample from the Rice distribution, to model the noise in a communications channel. The probability density function (pdf) of the Rice distribution is given by

$$f(x|\upsilon,\sigma) = \frac{x}{\sigma^2}\ exp\left(\frac{-(x^2+\upsilon^2)}{2\sigma^2}\right) I_0\left(\frac{x\upsilon}{\sigma^2}\right)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind. When $\upsilon = 0$ we obtain the Rayleigh distribution. Sampling from this distribution is very difficult since the expression includes Bessel functions with complex terms. However, sampling from Rayleigh distributions is easy and we can use this fact to obtain samples from the Rice distribution with parameters $\upsilon = 5$ and $\sigma = 2$ using algorithm 1. To that we initialise $x_0 \sim U(0, 10)$ and set

$t = 0$. We run the Metropolis-Hastings algorithm for a maximum number of iterations of 20000, use as a proposal distribution $q = f(x(t-1), 0),\ t > 1$ and obtain the acceptance probability given by equation 3.15. Figure 3.2(a) shows a plot of the Rice pdf (red) and the Markov chain generated (blue) in which it can be observed that the sampler explores the whole support of the function. Figure 3.2(b) shows how well the histogram of the samples obtained fits the Rice pdf.

#### 3.2.3.2 The Gibbs Sampler

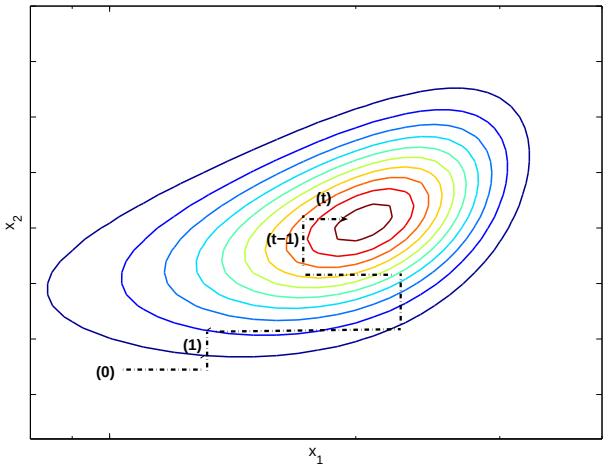



**Figure 3.3:** Illustration of diferent moves of a Gibbs sampler for a bivariate target distribution. Updates to $x_1$ implies horizontal moves whereas updates to $x_2$ implies vertical moves

The Gibbs sampler is another method of constructing a Markov chain with a given stationary distribution. The Gibbs sampler is, formally, a special case of the single-component Metropolis-Hastings algorithm in which the components, *e.g.* the parameters of the distributions, are updated one by one. Formerly used in statistical physics and called the Heat-Bath Algorithm (also Glauber dynamics), it was rediscovered by Geman and Geman [59] in their seminal work on Bayesian image analysis, it received its name for the use of Gibbs distributions on latices. In practice, this algorithm is not restricted to these distributions and other distributions can be used if we are able to obtain their full conditional distributions. Briefly, suppose the random variable $X$ can be divided into components $\{X_1, X_2, \cdots, X_m\}$ of possible differing dimension for some $m > 1$, the *full conditional distribution* $\pi(X_i|X_{-i})$ is the distribution of the $i^{th}$ component of $X$ given the remaining

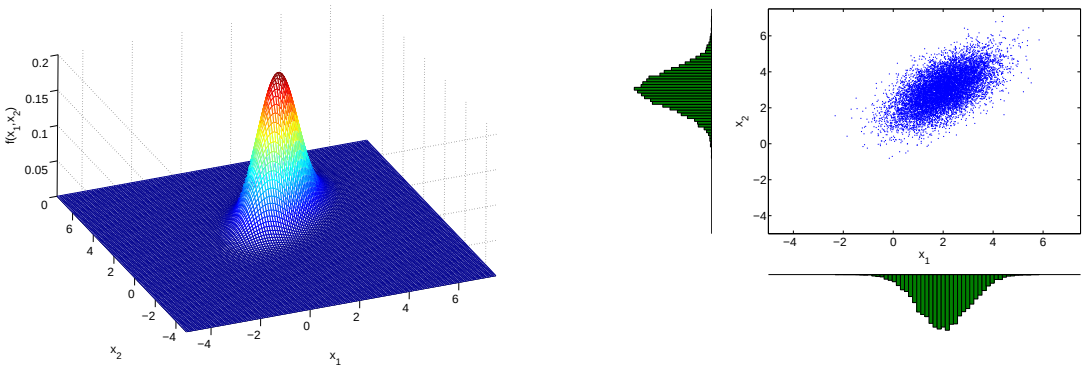
components. The Gibbs sampler draws samples from the full conditional distributions in turn (see figure 3.3) to generate a Markov chain whose distribution converges to the stationary distribution. The implementation of the Gibbs sampler is summarised in pseudocode as follows

**Algorithm 2 (Gibbs sampler)** *Given* $x^{(t)} = (x_1^{(t)}, \cdots, x_m^{(t)})$, *generate*

*1.* $X_1^{(t+1)} \sim \pi_1(x_1|x_2^{(t)}, \cdots, x_m^{(t)})$
*2.* $X_2^{(t+1)} \sim \pi_1(x_2|x_1^{(t+1)}, x_3^{(t)}, \cdots, x_m^{(t)})$
$\vdots$
*m.* $X_m^{(t+1)} \sim \pi_1(x_m|x_1^{(t+1)}, \cdots, x_{m-1}^{(t+1)})$

Formally, the Gibbs sampler is equivalent to the composition of $m$ single-component Metropolis-Hastings algorithms with acceptance probability equal to 1 and proposal distribution equal to the full conditional distribution. In practice, even though each component updtate is reversible, the systematic-scan of algorithm 2 can result in a non-reversible Markov chain but the invariance property still holds. Alternatives to the systematic Gibbs sampler to produce a reversible Gibbs sampler using a symmetric scan or random sweep Gibbs sampler can be found in [121]. The demonstration of the convergence of the Gibbs sampler to the stationary distribution is quite simple and can be found in [131] whereas geometric convergence is shown in [123]. The use of the Gibbs sampler implies limitations compared to the Metropolis-Hastings algorithm since sometimes it may not be easy to obtain the full conditional distributions of the target distribution. Moreover, the Gibbs sampler does not apply to problems where the parameter space has components of differing dimensions, because the irreducibility condition on such spaces does not hold. There are analogies of the Gibbs sampler and the EM (Expectation-Maximization) approaches [39] used to obtain likelihood functions with missing data. Briefly, the EM method computes the expected value of the log-likelihood of the complete data using the current estimate of the parameter value and then maximizes this log-likelihood to obtain the parameter value. The Gibbs sampler can be thought of as a stochastic analog to the EM algorithm in which random sampling replaces the expectation and maximization steps.

### A Toy Example




(a) (b)

**Figure 3.4:** Gibbs sampler. Toy example (a) Bivariate normal distribution (b) Scatter plot of bivariate normal by Gibbs sampler and histograms

One of the most important probability distributions for modelling jointly random variables in statistical applications is the multivariate normal distribution. To cite one example, one would like to sample from a multivariate distribution in tracking applications to model the process noise in Kalman filters [120]. Although there are many procedures to sample from a multivariate distribution by using the Cholesky decomposition of the covariance matrix and the Box-Muller transform, we will obtain samples from this distribution using Gibbs sampler. Moreover, to simplify the example we restrict ourselves to bivariate normal distributions.

Suppose we want to sample from a bivariate normal distribution with joint probability density given by

$$f(x_1, x_2) = K^{-1} \cdot \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 - \left( 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \, \sigma_2} \right) \right] \right\}$$

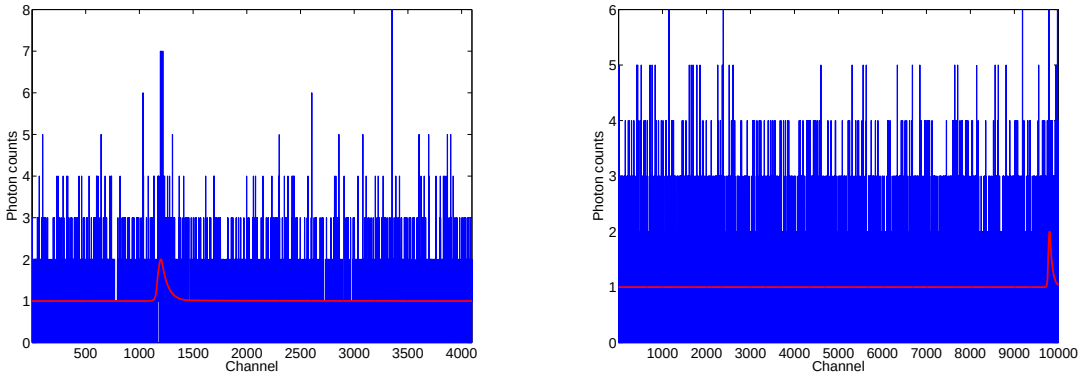
where $K = 2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}$ is a constant, $\rho$ is the correlation between $x_1$ and $x_2$, $\mu_1$, $\mu_2$, $\sigma_1$ and $\sigma_2$ are the mean value and the standard deviation of $x_1$ and $x_2$ respectively. Assume the parameter values are $\mu_1 = 2$, $\mu_2 = 3$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.6$. We run algorithm 2 for 10000 iterations using the following full conditional distributions.

$$x_1 | x_2 \sim \mathcal{N}(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1\sqrt{1-\rho^2})$$

$$x_2|x_1 \sim \mathcal{N}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2\sqrt{1-\rho^2})$$

with $(x_1^0, x_2^0) = (0, 0)$. Figure 3.4(a) shows the bivariate normal distribution from which we want to sample. Figure 3.4(b) shows a scatter plot of the different values obtained for $(x_1, x_2)$.

#### 3.2.3.3 Practical Considerations for the use of MCMC

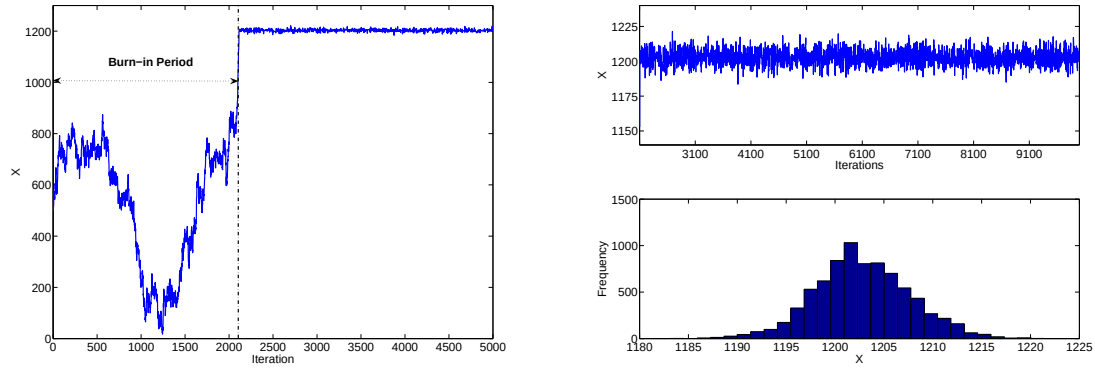



(a) (b)

**Figure 3.5:** Histograms of photon counts simulated to study the convergence the influence of the burn-in period and the mixing behaviour of Markov chain generated. The red colour represents the actual simulated signals whereas the blue color represents the signal degraded by noise

Irreducibility on a Markov chain implies convergence almost sure to the stationary distribution for almost every starting value if the Markov chain generated is recurrent or for every value if the chain is Harris recurrent. Having said that, every starting value (except those null set of initial values if the chain is not Harris recurrent) we choose (at random or otherwise) to initialise our Markov chain is a valid one and the chain will eventually reach the stationary distribution. Ideally, we would sample $X^{(0)}$ from the stationary distribution of the Markov chain, however, this is not always possible and in general we sample from the prior distribution or even use other algorithms to give initial estimates using our prior knowledge of the state space.

It is common practice to discard the first $n$ values generated of the chain, referred as the *burn-in period*, in order to reduce the influence of the starting values in equation 3.8. A rapidly *mixing* chain, i.e. a chain that moves rapidly around the support of our distribu-

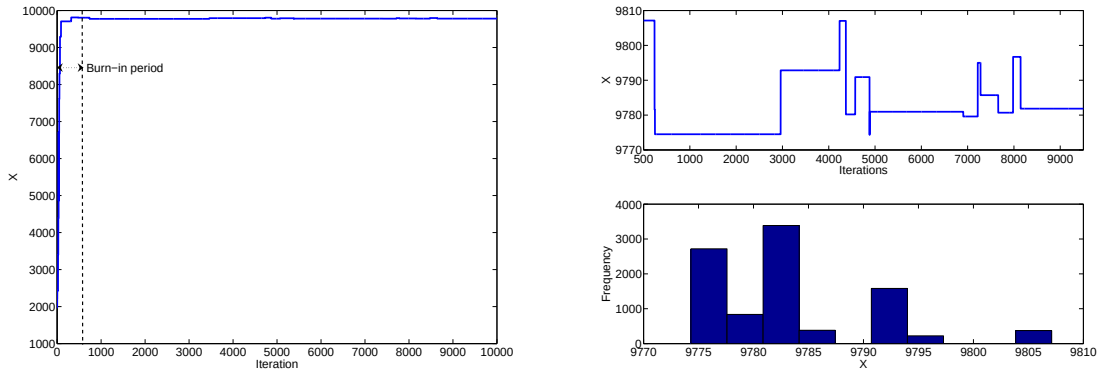



(a) (b)

**Figure 3.6:** Example of convergence of a Markov chain. (a) 10000 iterations from the Metropolis-Hastings algorithm that show a lengthy burn-in period (b) After this burn-in period, the Markov chain obtained shows a very good mixing behaviour as can be stated from the generated time-series trace of the position parameter and the histogram of this trace.

tion of interest, will quickly converge to the stationary distribution from extreme starting values whereas special care has to be taken for slow-mixing chains, to avoid an "excessive" burn-in period. Figures 3.6 and 3.7 show two different examples of two different time-series traces obtained when trying to find the position of the peaks of figures 3.5(a) and 3.5(b) respectively. In figure 3.6(a), the chain needs a burn-in period larger than the chain of figure 3.7(a). However, the mixing of the former chain after this burn-in period (figure 3.6(b)) is better than the later (figure 3.7(b)) which will require a larger number of iterations than the first one. As can be seen from the time-series trace plots, the length of the burn-in $n$ depends on the starting value and the rate of convergence of the transition kernel to the stationary distribution whereas the rate of convergence depends on the mixing of the chain. On the other hand, the rate of convergence affects the total length of the chain $N$ required to give adequate representation of the distribution.

In general, there are three types of convergence for which assessment is necessary [28]: *convergence to the stationary distribution* in which theoretical stationarity is only achieved asymptotically, *convergence of averages*, i.e. the chain should be run long enough to obtain adequate precision when using equation 3.8, and *convergence to independent identically distributed (iid) samples*. The asymptotic property of convergence of the Markov chain to the stationary distribution may motivate a sceptical attitude to be raised towards the




(a) (b)

**Figure 3.7:** Example of convergence of a Markov chain. (a) 10000 iterations from Metropolis-Hastings algorithm that show a quickly burn-in period (b) After this burn-in period, the Markov chain obtained shows a very bad mixing behaviour as can be stated from the generated chain and the histogram. Thus, a longer simulation is needed

use of MCMC methodologies since a-priori an infinite number of iterations are needed. However, the algorithm remains valid as we consider a single realization (or *path*) of the chain after the burn-in period when we begin considering that our samples come from the stationary distribution. To avoid this burn-in period, new methods have arisen such as perfect simulation (also known as coupling from the past) [114] which allow us to sample directly from the stationary distribution of certain Markov chains without visiting each state in the chain. However, it is not very clear how to use this methodology apart from a very few concrete examples in which the state spaces is discrete, although there has been improvements to generalise the algorithm to continuous spaces [104]. Convergence of averages, is fundamental in the implementation of MCMC algorithms and requires a good mixing of the Markov chain as explained previously. If the mixing speed of the chain is fast enough, the exploration of the posterior distribution by the chain would be more or less lengthy. Finally, to reduce the correlation between the consecutive samples of the Markov chain estimates, it is necessary to use either batch sampling techniques or thinning techniques which use subsampling of the Markov chain in order to generate iid samples. The size of the batch depends on the mixing properties of the Markov chain generated.

It is therefore important to make sensible choices of values $n$ and $N$. This is a hard issue and a range formal tools, known as *convergence diagnostics*, has been devised to assess

them. The two most popular methods are those of Gelman and Rubin [57], based on the analysis of the output of multiple chains, and Raftery and Lewis [116] based on a long run single chain. The former method is intended to assess the convergence of the Markov chains to the stationary distribution whereas the latter also provides a way of bounding the variance of estimates of quantiles of functions of parameters. More information on *convergence diagnostics* can be found in the review article by Cowles and Carlin [35]. In this thesis, we have taken different approaches to assess the convergence and mixing behaviour of the different Markov chains generated. On the one hand, we have used graphical outputs and Gelman and Rubin methodology for situations where the dimension of our state space is strictly 1. On the other hand, graphical outputs are considered when the dimension of our parameter space is variable and greater than 1.

### 3.2.4 Other Samplers

There are many other samplers that can be used instead of the Metropolis-Hastings algorithm or the Gibbs sampler that use the output of a Markov chain as a basis for inference. For instance, these two kernels can be combined to build the so-called *hybrid samplers* in which some components are updated using Gibbs kernels and others using Metropolis-Hastings steps. The list is extensive and many of them can be found in specialised literature such as [60]. However, we describe succinctly two of these algorithms for the sake of clarity of other algorithms described in later sections.

*Simulated Annealing* (see for instance [60]) is a technique that is used to optimise a function on a finite set with very large size. The main assumption to use this algorithm is based on the fact that MCMC methods may fail to converge to the stationary distribution (within a reasonable time) if movement through state space is inhibited by regions of low probability. Since movements from one region of high probability to other region requires the passage through states of very high probability, such event is unlikely to occur in a simulation run of limited length. Simulated Annealing is a technique that modifies the stationary distribution $\pi$ of the Markov chain to improve mixing. In this way, the new modified stationary distribution is defined as $\pi^*(\cdot) = \pi(\cdot)^{\frac{1}{T}}$, where $T$ is known as the temperature of the system. Increasing the temperature parameter will flatten $\pi$ and will allow a better exploration of the state space. Note that the temperature parameter $T$ is

changed along the simulation according to a pre-specified schedule. Large changes on the value of $T$ will make the chain generated to converge to a local maxima of the posterior distribution whereas small changes on $T$ will make impracticable the use of the algorithm. The main drawback of the Simulated Annealing algorithm is that the standard convergence properties do not apply to time-heterogeneous Markov chains. The simulating annealing algorithm converges towards a point (presumably the mode of $\pi$) whereas the standard MCMC algorithm converges in distribution.

The *Langevin-Hastings* (see for instance the work in [133]) algorithm is an alternative to the random-walk Metropolis algorithm in which the proposed value is a random perturbation of the current value of the chain. The Langevin-Hastings algorithm adjusts the proposal distributions by considering a time-discretised Langevin diffusion, that is, using the gradient of the log density of $\pi$. The main drawbacks of this approach are the non-recurrence of the algorithm and its erratic behaviour when the target distribution is not sufficiently smooth or has zeros in the state-space.

## 3.3 Reversible Jump Markov Chain Monte Carlo Algorithms

The MCMC approach described in earlier sections is effective when the dimension of the state space is fixed, *e.g.* a Euclidean space of dimension, say, $n$, but in some cases the dimensionality is in some sense object of inference. For example, when the number of components of a mixture distribution is unknown [119] or segmentation of functional magnetic resonance imaging (FMRI) data [158] to cite some of them. In our case, we are interested in problems where the number of returns (as well as the different parameters of these returns) in our lidar signal is unknown. Therefore, we can use this approach to determine the suitable number of components of the mixture distribution, in which a probability distribution is represented as a linear superposition of component distributions. Unfortunately, standard MCMC algorithms are not applicable and new algorithms are needed in order to explore all these different subspaces.

Grenander and Miller [68] tried to solve this problem in the specific field of pattern recognition. In their paper, they proposed an algorithm that they termed *jump diffusion*.

Basically, this algorithm comprises two different moves:

- The first one allows movement between models; it is called jump because the algorithm jumps from one model of dimension $n$ to other of dimension $m$ on a random exponential time.
- The second move is within-model diffusion according to Langevin stochastic differential equations.

However, Green [68](pag. 589) stated that the number of jump transitions discussed in Grenander and Miller work is limited and that this algorithm suffers from the drawbacks explained in section 3.2.4. Consequently, Green described the basis of a new, more general framework that is described in rigorous way in [63]. In this work, Green proposed a technique called *reversible jump Markov chain Monte Carlo* (RJMCMC). This technique is an extension of the popular Metropolis-Hastings algorithm designed to allow jumps between different dimensional spaces. Furthermore, these jumps must be controlled by the acceptance ratio which appears in traditional Metropolis-Hastings algorithms. Basically, this approach incorporates two concepts: the dimension-matching parameter, which helps to verify the so-called "dimension-matching" condition, and the "state transformation" function which deterministically relates the current and the proposal state and the dimension-matching parameter. Green imposed the stronger condition of reversibility to make the kernel being invariant. The two concepts previously mentioned allow the chain to be reversible. The diffusion-jump algorithm previously commented can be considered as a particular case of the RJMCMC algorithm.

Since its first appearance, the RJMCMC algorithm has been employed and discussed in many contexts such as hidden Markov models [122], mapping of quantitative trait loci [138], texture segmentation [45] and many other contexts. To describe the algorithm, we use the measure-theoretical approach used in [63,64] which allows a real understanding of the generality of the method and the variety of applications. For a more accessible explanation of the algorithm, we refer the reader to the article by Waagepetersen and Sorensen [152].

### 3.3.1 General Formulation

Let the probability measure on a random variable $x$, $\pi(dx)$, denote a *target distribution* of interest. In general, this target distribution is the posterior distribution for the parameters $\theta$ with values in a space $\chi$. Usually $\chi$ will be a subset of $\Re^k$ and $\pi(dx)$ will have a Radon-Nikodym density $\pi(x)$ with respect to a $\sigma - finite\ measure\ \mu$. Therefore, the posterior distribution can be written as explained in [142] as

$$\pi(dx) = \pi(x)\mu(dx) \tag{3.16}$$

As explained in previous sections, the task is to construct a Markov chain on a general state space with $\pi(dx)$ as its limiting distribution. Once the convergence towards this limiting distribution is achieved, we form a realisation of the chain $\{X^{(1)},\ X^{(2)},\ ...,\ X^{(N)}\}$ and we can treat it as it was a random sample from $\pi(dx)$. If the chain is ergodic, the ergodic theorem holds and it is possible to use these realisations of the chain to build histograms, estimate moments and even calculate probabilities in a frequentist way. In order to allow the Markov chain to converge towards the limiting distribution, the transition kernel, $K(\cdot,\cdot)$, constructed has to verify the conditions described in section 3.2.2. Invariance of the transition kernel is guaranteed if reversibility with respect to $\pi$ holds. Through detailed balance, the reversibility condition holds:

$$\int_A \pi(dx)K(x,B) = \int_B \pi(dx')K(x',A) \tag{3.17}$$

for all Borel sets $A,\ B \subset \chi$.

Since we are dealing with complex models, the Metropolis-Hastings paradigm is more appropriate than the Gibbs sampler. When the dimension of the state space increases or $\pi$ has a complex form, it is very difficult to sample from distributions unless standard distributions or conjugate priors are used. On the other hand, if the dimension of the state space is not fixed, the procedure used in the Gibbs sampler is not meaningful since the interpretation of the different variables can be different depending on which subspace we are.

As in the standard Metropolis-Hastings algorithm, the procedure begins drawing a new state $x'$ from a proposal measure $q(x, dx')$. This new state is accepted with some probability determined by the so-called acceptance probability $\alpha(x, x')$. In the measure theory approach, the transition kernel of the Metropolis-Hastings paradigm can written as:

$$K(x, B) = \int_B q(x, dx')\alpha(x, x') + I_{(x \in B)} \int_\chi (1 - \alpha(x, u))q(x, du) \tag{3.18}$$

where $I_{(.)}$ is the indicator function, $q(x, dx')\alpha(x, x')$ the probability of performing a move and $\int_\chi (1-\alpha(x, u))q(x, du)$ is the probability of rejecting the move, also known as the *atom* of the Markov chain.

Substituting equation 3.18 in 3.17

$$\begin{aligned} &\int_A \pi(dx) \int_B q(x, dx')\alpha(x, x') + \int_{A \cap B} \pi(dx) \left\{ \int_\chi (1 - \alpha(x, u))q(x, du) \right\} \\ &= \int_B \pi(dx') \int_A q(x', dx)\alpha(x', x) + \int_{B \cap A} \pi(dx') \left\{ \int_\chi (1 - \alpha(x', u'))q(x', du') \right\} \end{aligned} \tag{3.19}$$

Since the *atom* contributes the same quantity in both sides of equation 3.19, Green [64] stated that

$$\int_{(x,x') \in \ A \times B} \pi(dx)q(x, dx')\alpha(x, x') = \int_{(x,x') \in \ A \times B} \pi(dx')q(x', dx)\alpha(x', x) \tag{3.20}$$

Assuming now that $\pi(dx)q(x, dx')$ is dominated by a symmetric measure $\nu$ on $\chi \times \chi$. By virtue of the Radon-Nikodym theorem, $\pi(dx)q(x, dx')$ has a density (Radon-Nikodym derivative) $f$ with respect to $\nu$:

$$\pi(dx)q(x, dx') = f(x, x')\nu(dx, dx') \tag{3.21}$$

Thus, equation 3.20 becomes

$$\int_{(x,x') \in \ A \times B} \alpha(x, x')f(x, x')\nu(dx, dx') = \int_{(x,x') \in \ A \times B} \alpha(x', x)f(x', x)\nu(dx', dx) \tag{3.22}$$

Since $\nu$ is symmetric, the equality of equation 3.22 holds if

$$\alpha(x, x')f(x, x') = \alpha(x', x)f(x', x) \tag{3.23}$$

Peskun's theorem shows (see [110, 142]) that the following equation satisfies equation 3.23

$$\alpha(x, x') = \min\left\{1, \frac{f(x', x)}{f(x, x')}\right\} \tag{3.24}$$

which is a ratio of Radon-Nikodym derivatives with respect to a common dominating measure and can be rewritten in an informal way as

$$\alpha(x, x') = \min\left\{1, \frac{\pi(dx')q(x', dx)}{\pi(dx)q(x, dx')}\right\} \tag{3.25}$$

where $\alpha(x, x')$ is known as the acceptance probability.

Despite of the fact that this algorithm is expressed in an abstract way through the use of measures instead of probabilities, the general expression adopted in equation 3.24 is more flexible since it is not limited to the usual Lebesgue measure. As an example, we refer to the work of Srivastava *et al.* [133] and Lacoste *et al.* [89]. Srivastava *et al.* used a Jump-diffusion algorithm in a Lie space to estimate simultaneously both the target orientations and their numbers in problems of recognizing targets from their observed images. On the other hand, Lacoste *et al.* addressed the problem of unsupervised extraction of line networks from remotely sensed images using a point process to exploit their topological properties and a simulated annealing using RJMCMC algorithm to perform optimisation under a Poisson measure.

### 3.3.2 A More Constructive Formulation of the RJMCMC

Although the measure theoretical formulation described in the previous section was necessary to demonstrate the validity of the method, RJMCMC can be formulated in a way that bypasses much of the measure theory. Applying the method to real problems necessitates no direct consideration of measure theory. To this end, Green [64] explored the actual mechanics of the transitions in terms of random numbers, whilst still retaining a relatively general framework. In this way, the dominating measure and Radon-Nikodym derivatives can be generated implicitly. Take the case where $\chi \subset \Re^d$ and that the target distribution $\pi(dx)$ has a density $\pi(x)$ with respect to a *d-dimensional* Lebesgue measure $\nu_d$. As stated in [63], it may be useful to generate the proposed new state as a deterministic function, $x' = h(x, u)$, $h : \Re^d \times \Re^r \rightarrow \Re^{d'}$, of the current state $x \in \Re^d$ and a vector of $r$ random components $u$ from a known joint density $g$ with respect to a $m$-dimensional Lebesgue

measure, $\nu_m$. Hence, the left-hand side of equation 3.22 can be rewritten as:

$$\int_{(x,x')\in\ A\times B} \pi(x)g(u)\alpha(x,x')\nu_d(dx)\nu_m(du) \tag{3.26}$$

In a similar way, the reverse move (from $x'$ to $x$) can be generated using another deterministic function, $x = h'(x', u')$, $h' : \Re^{d'} \times \Re^{r'} \rightarrow \Re^d$, where $x' \in \Re^{d'}$ is the proposed new state and $u'$ is a vector of $r'$ random components $u'$ from a known joint density $g'$ with respect to a $m'$-dimensional Lebesgue measure, $\nu'_{m'}$. If the transformation from $(x, u)$ to $(x', u')$ is a diffeomorphism we can apply the change of variable theorem to equation 3.26 and affirm that the equality in equation 3.22 holds if

$$\pi(x)g(u)\alpha(x,x') = \pi(x')g'(u')\alpha(x',x)\left|\frac{\delta(x',u')}{\delta(x,u)}\right| \tag{3.27}$$

being the last factor the Jacobian arising from the change of variable from $(x, u)$ to $(x', u')$ and is not really an inherent component of the RJMCMC algorithm. As Green commented, one of the advantages of using the different deterministic functions is the flexibility given to the RJMCMC methodology allowing different ways of expressing the proposals.

Using equation 3.27, the expression of the acceptance probability becomes

$$\alpha(x,x') = \min\left\{1, \frac{\pi(x')g'(u')}{\pi(x)g(u)}\left|\frac{\delta(x',u')}{\delta(x,u)}\right|\right\} \tag{3.28}$$

In this case, the abstract form involving the ratio of measures is therefore reduced to a "simple" ratio between ordinary joint probabilities.

All these assumptions hold only if $\pi$ is absolutely continuous with respect to the measures which rules in the respective spaces and if the transformation from $(x, u)$ to $(x', u')$ verifies

$$d + r = d' + r' \tag{3.29}$$

which is known as the "dimension-balancing" condition. As long as the "dimension-balancing" is satisfied, $x$ and $x'$ can have different dimensions.

### 3.3.3 Model Choice Problem

In addition to learning about the parameters of a model, we might be interested in comparing different models. Usually, when comparing different models, the parameter dimension

is fixed but sometimes it is necessary to have a MCMC sampler that jumps between subspaces of different dimensions. Such spaces of different dimension correspond to the parameter spaces of the different models. When dealing with a model choice problem, the dimension of the parameter space is itself subject to inference and we shall be interested in jointly estimate the parameter and the dimension of the parameter space. A typical example of this problem is that of the analysis of mixture distributions where both the number of components and the mixture component parameters are unknown. In this case, different models differ only on the number of components.

In general, this problem can be tackled using the result given by equation 3.24 and finding an appropriate measure in the joint state space of the parameters and the dimension of the parameter space. However in many cases, this measure is hard to find and it is not evident. Green [63] proposed an strategy to cope with this problem by using the approach of the deterministic proposals exposed in the previous section combined with set of moves which allow to jump between the different parameter subspaces. We describe now the formulation given by Green that can be used in a generic way.

Suppose now that our state space $\chi$ is no longer a subset of $\Re^d$, but rather a general state countable union of spaces $\chi_k$, $k \in \mathcal{K}$, with possibly different dimensions $n_k$, so that $\chi = \bigcup_K \chi_k$. In order to explore the whole space, multiple type of moves have to be set. A move type $m$ might use transitions between $\chi_i$ and $\chi_j$, $i, j \in \mathcal{K}$. Suppose now that we index the models by $k$ in a $\mathcal{K}$ and that $\chi \subset \Re^d$. We redefine the state space as

$$\chi = \bigcup_{k \in \mathcal{K}} (\{k\} \times \Re^{n_k}) \tag{3.30}$$

where $n_k$ is the dimension of the space $\Re$ that verifies $n_k \leq d$ and depends on the model index $k$. Again we are interested in $\pi$, the distribution over $\chi$.

To explore this state space using reversible jump we need multiple moves. A particular move $m \in \mathcal{M}$ proposes to take $x' = (k', \theta'_{k'})$ from $x = (k, \theta_k)$ where $\theta_k$ is a vector of parameters whose length depends on $k$. The new detailed balance equation can be written as:

$$\int_{(x,x') \in A \times B} \pi(dx) q_m(x, dx') \alpha_m(x, x') = \int_{(x,x') \in A \times B} \pi(dx') q_m(x', dx) \alpha_m(x', x) \tag{3.31}$$

for each different move $m$, and where $q_m(x, dx')$ is the joint distribution of move type $m$ and destination $x'$. Each move is a transition kernel reversible with respect to $\pi$ and the combination of all these moves, will create an ergodic chain. In this case the transition kernel is a mixture over a number of move types

$$K(x, B) = \sum_M \int_B q_m(x, dx')\alpha_m(x, x') + r(x)I(x \in B), \forall\ x \notin B \tag{3.32}$$

where $r(x)$ is the probability of rejecting the move.

With a similar reasoning to that of section 3.3.1 and using equations 3.32 and 3.31, the acceptance probability is determined as

$$\alpha_m(x, x') = \min\left\{1, \frac{\pi(dx')}{\pi(dx)}\frac{q_m(x', dx)}{q_m(x, dx')}\right\} \tag{3.33}$$

which can be written more explicitly as:

$$\alpha_m(x, x') = \min\left\{1, \frac{\pi(x')}{\pi(x)}\frac{j_m(x')}{j_m(x)}\frac{g'_m(u')}{g_m(u)}\left|\frac{\delta(x', u')}{\delta(x, u)}\right|\right\} \tag{3.34}$$

where $j_m(x)$ is the probability of choosing move type $m$ when at $x$. This term appears due to the random nature on the choice of the different moves. In this case, the "dimension-balancing" condition depends on $m$, $d_m + r_m = d'_m + r'_m$ as well as the deterministic functions $h_m$ and $h'_m$ and the Jacobian term. Note that when at $x = (k, \theta_k)$ all the moves may not be available and therefore, there is a non-zero probability that no move is attempted (aperiodicity).

In some cases, the value $r_m$ (or $r'_m$) is set to be zero and therefore there is no need to generate the vector of random values $u$ (respectively $u'$). For instance, in a simple birth/death process where we only allow the birth or the death of a single element. Indeed, in the birth case the "dimension-balancing" condition becomes

$$\begin{aligned} &d_m + r_m = d'_m,\ r_m = 1 \\ &and\ h : \Re^{d_m} \times \Re^1 \rightarrow \Re^{d'_m} \end{aligned}$$

whereas in the death case is

$$\begin{aligned} &d_m = d'_m + r'_m,\ r'_m = 1 \\ &and\ h' : \Re^{d'_m} \times \Re^1 \rightarrow \Re^{d_m} \end{aligned}$$

Thus, if $r'_m = 0$ (birth process) the acceptance probability can be simplified to the following expression:

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')}{\pi(x)} \frac{j_m(x')}{j_m(x) g_m(u)} \left|\frac{\delta(x')}{\delta(x, u)}\right|\right\} \tag{3.35}$$

Respectively, if $r_m = 0$ (death process), the acceptance probability is

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')}{\pi(x)} \frac{j_m(x') g_m(u')}{j_m(x)} \left|\frac{\delta(x', u')}{\delta(x)}\right|\right\} \tag{3.36}$$

### 3.3.4 Practical Considerations for the use of RJMCMC Algorithms

The majority of the considerations commented in section 3.2.3.3 to MCMC can be applied to RJMCMC. However, since the state space that we are exploring is more general than the state spaces explored with MCMC new considerations have to be taken into account when using RJMCMC algorithms. One of the most important difficulties is the construction of efficient proposals. In general, moves that imply change of dimension result in low acceptances rates compared to other type of moves in which the dimension of the state space is fixed. In general, these low convergence rates lead to slow convergence of the algorithm. Brooks *et al.* [21] propose different mechanisms to alleviate the problem of choosing a more efficient proposal distribution based upon an analysis of the acceptance probabilities and generalising the RJMCMC algorithm using the so-called dual space approach that allows to retain some degree of memory of the process. Alternative proposal mechanisms can be found in [2]. Recently, interest has grown in constructing automatic generic RJMCMC samplers which adapt their proposal distributions to the specific problem of interest [64]. Although very promising results with synthetic data have already been obtained [72], further work is needed to apply these algorithms to more general problems.

Another problematic area when using RJMCMC is to determine whether or not a particular realisation of a Markov chain has converged to the stationary distribution. Indeed, as the state space becomes more general it is difficult to implement theoretically and practically diagnosis of convergence. One reason for this is that some models will be visited infrequently and so the assessment of convergence within such models is almost impossible. Brooks and Giudici [20] give some directions to address this problem and present a variation of the Gelman and Rubin method which applies to variable dimension problems.

However much work has to be done in this topic which is still the subject of research.

### 3.3.5 The Delayed Rejection Algorithm

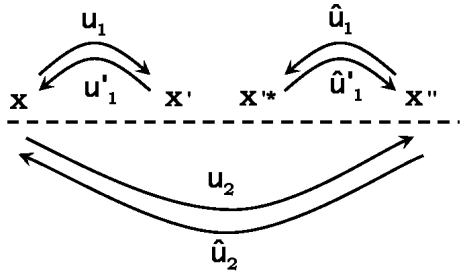



**Figure 3.8:** Schematic adapted from [65] representing the variables in play in a delayed rejection algorithm

The delayed rejection algorithm was first introduced by Tierney and Mira [144] as an extension of the Metropolis-Hastings algorithm. As in the standard Metropolis-Hastings algorithm, the delayed rejection proceeds by proposing a candidate value and accepting it according to the acceptance probability of equation 3.15. However, if the candidate is rejected instead of retaining the same value and advancing in time, we propose a new candidate in the second stage (see figure 3.8). This process is not limited to two stages and can be expanded to more stages [102]. These algorithm presents several advantages compared to the standard Metropolis-Hastings algorithm. On a theoretical point view, the algorithm is improved in the sense of Peskun [110] since increasing the rate of accepted values will reduce the autocorrelation of the Markov chain and the sample path averaging will be more efficient. On the other hand, the proposals in the following stages are allowed to use the previous rejected value in the new proposal distribution. Therefore, the algorithm is adaptive in the sense that we can use information from previously rejected values to explore the state space more efficiently.

Green and Mira [65] extended the delayed rejection algorithm to more general spaces within the framework of the RJMCMC algorithm relaxing some of the conditions given by Tierney and Mira [144]. Although there are detractors [72] of this methodology who cite the increase of the computational cost [72], we believe that the advantages reported using a delayed rejection step can outweight the drawbacks associated with it. The delayed rejection algorithm will permits us temporary local adjustments of the proposal distribu-

tion. To explain the algorithm, we shall follow the notation and structure of Green and Mira [65].

With respect to figure 3.8, assume we are at $x$, we have proposed a move[4] $x' = h_1^+(x, u_1)$ and this move has been been rejected with probability $1 - \alpha_1(x, x')$, where $\alpha_1$ is given by equation 3.28. Once, $x'$ is rejected, we propose a new value $x'' = h_2^-(x, u_1, u_2)$ which may depend on the previous rejected value. We have to find now the correct acceptance probability for the second stage, that is, $\alpha_2(x, x'')$. As in the standard RJMCMC approach, detailed balance at the second stage ensures that the Markov chain generated is reversible and then invariant. Assuming that $\Pi(A, B, C)$ is the probability in equilibrium of being at $A$, proposing a move to $B$ and rejecting it and then proposing and accepting a second move to $C$ for all Borel sets $A, B, C \subseteq \chi$. The detailed balance conditions requires that

$$\Pi(A, \chi, C) = \Pi(C, \chi, A), \ \forall \ A, C \tag{3.37}$$

The left-hand side of equation 3.37 is given by the following expression

$$\Pi(A, B, C) = \int\limits_{(x,x',x'') \ \in \ A\times B\times C} dx du_1 du_2 \pi(x) g_1(u_1) g_2(u_2) \{1 - \alpha_1(x, x')\} \alpha_2(x, x'') \tag{3.38}$$

whereas the right-hand side is given by

$$\Pi(C, B^*, A) = \int\limits_{(x,x'^*,x'') \ \in \ C\times B^*\times A} dx'' \ d\tilde{u}_1 \ d\tilde{u}_2 \ \pi(x'') \tilde{g}_1(\tilde{u}_1) \tilde{g}_2(\tilde{u}_2) \{1 - \alpha_1(x'', x'^*)\} \alpha_2(x'', x) \tag{3.39}$$

where $x'^* = h_1^-(x'', \tilde{u}_1)$, $x = h_2^-(x'', \tilde{u}_1, \tilde{u}_2)$ and $\tilde{g}_1$, $\tilde{g}_2$ are the densities from which $\tilde{u}_1$ and $\tilde{u}_2$ are drawn. The state $x'^*$ is known as the virtual state and prevents the rejected state to be the same in both directions and therefore does not require the existence of symmetric state spaces. The virtual state is the key difference with the algorithm developed by Tierney and Mira [144] and allows its use in variable-dimension problems.

Suppose that we have a diffeomorphism from $x \times [0, 1] \times [0, 1]$ to $x'' \times [0, 1] \times [0, 1]$ which maps $(x, u_1, u_2)$ to $(x'', \tilde{u}_1, \tilde{u}_2)$. Using the change of variable theorem equation 3.39 becomes

$$\Pi(C, B^*, A) = \int\limits_{(x,x'^*,x'') \ \in \ C\times B^*\times A} dx \ du_1 \ du_2 \ \pi(x) \tilde{g}_1(\tilde{u}_1) \tilde{g}_2(\tilde{u}_2) \{1 - \alpha_1(x'', x'^*)\} \alpha_2(x'', x) \ J \tag{3.40}$$

[4] As in the RJMCMC case, $h(\cdot, \cdot)$ is a deterministic function of the current state and some random vector. The subscripts + and - represents in our case increasing or decreasing the dimension of the space.

where $J = \left| \frac{\delta(x'', \tilde{u}_1, \tilde{u}_2)}{\delta(x, u_1, u_2)} \right|$ is the Jacobian arising from the change of variable theorem.

Using equation 3.38 and 3.40, the detailed balance condition is ensured if

$$\pi(x)g_1(u_1)g_2(u_2)\{1-\alpha_1(x,x')\}\alpha_2(x,x'') = \pi(x)\tilde{g}_1(\tilde{u}_1)\tilde{g}_2(\tilde{u}_2)\{1-\alpha_1(x'',x'^*)\}\alpha_2(x'',x)\ J \tag{3.41}$$

and the acceptance probability of the second stage can be determined as

$$\alpha_2(x,x'') = \min\left\{1, \frac{\pi(x'')}{\pi(x)}\frac{\tilde{g}_1(\tilde{u}_1)\tilde{g}_2(\tilde{u}_2)}{g_1(u_1)g_2(u_2)}\frac{\{1-\alpha_1(x'',x'^*)\}}{\{1-\alpha_1(x,x')\}}J\right\} \tag{3.42}$$

The term $\alpha_1(x'', x'^*)$ is obtained using a similar equation to equation 3.28 in which $x'' = h_1^+(x'^*, \tilde{u}_1')$ with $\tilde{u}_1'$ is drawn from a suitable distribution $\tilde{g}_1'$.

## 3.4 Mixture Distributions

In this section, we briefly review mixture distribution in a Bayesian context due to the similarities with our approach (see chapter 4). In many applications, the process under observation cannot be modeled by classical probability distributions such as normal, gamma, Poisson, etc. The probability structure may be too complicated to be explained with these simple distributions as in the case of the TCSPC and BIL data. In such situations, one solution for explaining the underlying phenomenon is to use finite mixtures of probability distributions [99,145]. Applications of mixture distributions include disease mapping [66], clustering of galaxies [139] and segmentation of medical images [158], to cite some of them. The analysis of mixture models is a parametric procedure which relies in approximating the true density of interest by using a summation of standard densities as follows

$$f(x|\theta) = \sum_{j=1}^{k} p_j\ f_j(x, \theta_j) \tag{3.43}$$

where $f_j$ is a standard density with parameter set $\theta_j$, $p_j$ is a weight such as $0 \leq p_j \leq 1$, $\sum_{j=1}^{k} w_j = 1$, $k$, $k < \infty$, is the number of mixture components.

Mixture models do not always possess an individual significance or physical interpretation for the process modeled. However, there are situations in which we might be interested in determining the heterogeneity of a population of samples as in the case of clustering in disease studies [88] or Bayesian classification of Neolithic tools [37] to cite some examples.

In the Bayesian paradigm, given the sample $x = \{x_1, \cdots, x_n\}$, the posterior distribution distribution associated to a mixture model can be expressed as

$$\pi(p, \theta|x) \propto \prod_{i=1}^{n} \left\{ \sum_{j=1}^{k} p_j \ f_j(x_i, \theta_j) \right\} \pi(p, \theta) \tag{3.44}$$

where is the $\pi(p, \theta)$ joint prior distribution of the parameters and weights. If the sample size is very large, the number of terms involved makes the computation of the normalising constant and the posterior distribution impractical.

To cope with this problem one of the most famous solutions is that of using the inherent missing data structure of equation 3.43 (see for instance [41] or [49]) to facilitate inference. In this methodology, we associate with every observation $x_i$ from a mixture distribution of $k$ components an indicator variable $z_i \in \{1, \cdots, k\}$, known as the missing data, that indicates the component of the mixture from which $x$ is drawn. By demarginalisation of equation 3.43,

$$Z_i \sim \mathcal{M}_k(1; p_1, \cdots, p_k), \quad x_i|z_i \sim f_{z_i}(x|\theta_{z_i}) \tag{3.45}$$

where $\mathcal{M}_k(1; p_1, \cdots, p_k)$ is the multinomial distribution with $k$ possible outcomes and a single observation. If we consider the complete model structure, $(x_i, z_i)$, the mixture structure disappears since the likelihood takes the following expression

$$L(p, \theta|x_1, \cdots, x_n, z_1, \cdots, z_n) = \prod_{j=1}^{k} \prod_{i; z_i = j} p_j \ f_j(x_i, \theta_j) \tag{3.46}$$

Since this is a population model, there is non-zero probability that no sample is allocated to any given component. The missing data structure is also exploited in the frequentist framework in applications of the EM algorithm [39].

### 3.4.1 On the Analysis of Mixture Distributions with an Unknown Number of Components

In applications where the components have a physical interpretation, inference for $k$ may be of interest in itself. Inference for the number of components $k$ may be seen as a specific example of the very common problem of choosing a model from a given set of competing models. Taking a Bayesian approach to this problem has the advantage that it provides

not only a way of selecting a single "best[5]" model, but also a coherent way of representing model uncertainty. Classical Bayesian approaches to this problem typically involve the estimation of the marginal posterior distribution of the different parameters given a fixed $k$ and then comparing these marginal posterior distributions according to some criteria such as Bayes factor (see for instance [27]). However, these methodologies can suffer from underestimating the real dimension of the state space since not all the possible $k$ are explored.

Richardson and Green [119] proposed a fully Bayesian approach to the problem of estimating normal mixture distributions in which the number of components and the mixture component parameters are estimated jointly. Since moving from states spaces of different dimension is difficult, the proposed the use of Green's RJMCMC algorithm [63] which allow them to jump between different dimensionality subspaces. In order to explore better the state space, they also proposed the use of a mixture of kernels to perform moves between subspaces of different dimension. Namely these moves are birth/death of a component in the mixture distribution and split of one component into two different components or merging of two components into one. These different moves are performed in tandem to ensure the property of reversibility of the Markov chain. When the dimension of the state space is fixed, Richardson and Green considering updating the different parameters by means of Gibbs sampling or Metropolis-Hastings algorithms.

### 3.4.2 Inference Difficulties on Mixture Distributions

The analysis of mixture distributions with an unknown number of components using RJMCMC offers numerous advantages. To cite some them, it allows us to treat $k$ as random value instead of a fixed one, to explore better the mixture fully with a single analysis, to obtain posterior distributions of the parameters of interest instead of just "best estimates" and therefore to produce a fully Bayesian analysis of the phenomenon of interest. However, this methodology has also some drawbacks as mentioned in the discussion of the paper by Richardson and Green [119].

Mixture models are considered to belong to the group of inverse problems, where the

[5]The term best does not refer to the actual exact model that fits data but the model which is closer to reality among all the available models

value of some (or all) the model parameters is provided indirectly by data. The use of data augmentation in mixture models is a delicate issue since there is always a non-zero probability that no data are allocated to any given component of the mixture, that is, the data do not brings any information about any of the parameters of the mixture. As a result of this, the likelihood function may become unbounded and improper [6] posterior distributions may arise when using improper prior distributions [84].

Another important issue to be considered when dealing with mixture models is the so-called "label-switching" problem [119, 140]. This problem is due to the property of invariance under permutation of the indexes of the components in the mixture. In other words, the different components are not identifiable marginally because they are exchangeable. Jasra *et al.* [84] review different techniques that try to cope with the "label-switching" problem.

Finally, convergence, interpretation of the outcome of the Markov chains generated and the choice of the prior are also delicate. The mixture model explored with an RJMCMC algorithm inherits the problems of convergence assessment previously mentioned. Assessing the convergence of the Markov chains is difficult and, in general, the user relies on a-priori knowledge of the problem and subjective assumptions. The interpretation of the results is also tricky and heavily influenced by the "label-switching" problem previously described. Further, there can be great sensitivity to choice of the prior.

### 3.4.3 Alternatives to RJMCMC

The RJMCMC approach to the analysis of mixture distributions with an unknown number of components has been proved to be successful in different applications such as Normal mixture distributions [119], Poisson mixture distributions [151] and more recently to multivariate mixtures of normals [38] to cite some examples. However, this methodology has also generated some detractors notably due to the complexity in choosing appropriate proposal distributions, the not straightforward implementation of the algorithm and evaluating the Jacobian term. This section comments on some of the alternative approaches that also make use of Bayesian ideas and MCMC methods to explore distributions on state

[6] A distribution $\pi(x)$ is said to be improper if $\int \pi(y)dy = \infty$

spaces where the unknowns do not have a fixed dimension. These methods are known as transdimensional MCMC methods and include the RJMCMC methodology.

Stephens [139] argues in favor of using continuous-time simulation for variable dimension MCMC known as birth-death MCMC (BDMCMC) as an alternative to Green's [63] Reversible Jump Principle. In this work, the different parameters of the mixture are considered as a sample from a spatial (marked) point process[7] and simulate by constructing a continuous time Markov birth-death process. Note that this method also constructs an ergodic Markov chain with stationary distribution the posterior distribution of interest with an unknown number of components. The BDMCMC methodology can be applied in contexts where the parameter of interest may be considered as a point process, and where the likelihood of these parameters may be calculated in an explicit way. The main differences of the BDMCMC algorithm compared to the RJMCMC are that the algorithm operates in continuous time, uses neither missing data nor constraints to avoid the label-switching problem. Stephens also argues that the continuous-time alternative is simpler to implement than reversible jump, however, Cappe *et al.* [26] disagree with Stephens arguing that this simplicity is only a consequence of the birth-or-death move and that if more difficult moves such as the split-or-merge move are implemented the Jacobian term will also appear. Further, these difficult moves are computationally more expensive than in the RJMCMC. Moreover, Cappe *et al.* also show that under some weak regularity conditions there exists a sequence of RJMCMC processes converges to the BDMCMC process.

In recent years, interest has arisen in applying the same procedures that have been successful in dynamic approaches to the so-called batch problems [32]. These are problems where interest is in the joint posterior distribution of static parameters conditional on a single set of observations. For example, Chen and Liu [31] who propose a *mixture Kalman filter* that uses a random mixture of Gaussian distributions to approximate a target distribution. The main problem of Kalman filters is that they cannot be used on non-linear and non-Gaussian systems and different alternatives to the problem are needed. Further, they do not take into account all the salient statistical features of the processes under

[7] A marked point process is a point process defined by a density function with respect to the Poisson measure with some random parameters known as marks associated to each point. For instance, in image analysis these parameters might define some geometric properties of an underlying object. More information on point processes can be found in [12] and [103].

consideration, thereby often leading to poor results.

Sequential Monte Carlo methods have been proposed as a competitive alternative to MCMC methods (see [43] for a very detailed information on sequential Monte Carlo methods). In dynamic problems, when new data become available, there is a new posterior distribution of interest which render MCMC methods computationally impractical for real time applications. In general, sequential Monte Carlo methods approximates the different posterior distributions obtained by a cloud of weighted particles and updated sequentially as each new information is obtained.

Chopin [32] proposed the use of *sequential particle filters* for static model to cope with the problems of Kalman filtering previously mentioned. Although the results obtained seem promising, this algorithm requires the a priori knowledge of the number of components of the mixture distribution. In more recent works, Fearnhead [51] shows how particle filters can be used in problems where the number of components in a mixture is also unknown and show results in the specific application of mixture Dirichlet process models. Having resolved the problem that the Kalman filter presented, it remains to solve other specific problems associated with the use of sequential Monte Carlo methods. Indeed, since we are approximating the posterior distribution with a set of particles, Monte Carlo error accumulates over time. Moreover, MCMC methods converge asymptotically in distribution whereas sequential Monte Carlo methods converge asymptotically in probability, that is, pointwise. In general, sequential Monte Carlo methods suffer from progressive "impoverishment[8]" or "degeneracy" [6] introduced by the reweighting step in which the different particles are updated.

Another interesting approach is that of Petris and Tardella [111] who introduced a geometric approach to transdimensional MCMC and showed that it can be used as an alternative to the RJMCMC algorithm when dealing with nested models. The main idea of this algorithm is the transformation of the mixture of the posterior densities into a global density which gives some kind of geometric intuition of the jumps between the different subspaces. However, as they also stated, their approach is limited and computationally intensive.

[8] A phenomenon that appears in the particle filter approach where after a few iterations, all but one particle will have negligible weight

Recently, Nobile and Fearnside [107] have presented a new MCMC sampler called the *allocation sampler* that makes draws from the joint posterior distribution of the number of components and the allocation variables. They argue that the main advantage of this sampler is that remains essentially the same regardless of the dimension of the data, that is, can be exploited in multivariate scenarios and the family of mixture components. However, the price of this algorithm is that the mixture weights and component parameters must be able to be integrated out of the model in closed form. This algorithm can also be considered a version of the RJMCMC algorithm.

## 3.5 Spatial Statistics

In many applications, the data observed in the phenomenon under study can be referenced spatially as is the case when completely charaterizing 3D surfaces (see [83] for a list of other applications). Spatial statistics is concerned with the study of these data and associated stochastic model and processes. Since data that are spatially close may be more likely to be similar, the spatial model incorporates this assumption into the generating mechanism. Following Cressie's [36] notation, a spatial process can be represented as $\{Z(s) : s \in D\}$, where $s \in \Re^d$ is a generic location and $Z(s)$ is a random vector at location $s$. We devote this section to the special case when $D$ is a fixed (regular or irregular) collection of countably many points of $\Re^d$ called *lattice*. Examples of irregular lattices include disease mapping applications [88] or agricultural field experiments [18] whereas regular lattices include location of archeological sites [19], image analysis and remote sensing [91]. Note that these lattice structures can be represented by a graph in which the vertices or nodes represent the sites of the random variables and the edges some kind of relationship between the different random variables.

In general, one wants to capture to an extend the variability and the interactions of the different random variables present on the lattice. With such a spatial structure, one may consider modeling these random variables using their joint probability distribution to make inferences on the underlying process. However, the choice of a multivariate probability distribution is not easy and there is no generic theory for selecting a global model that considers all the possible relationships among the different random variables. One of the

most used multivariate distributions is the multivariate normal distribution which models associations through its covariance matrix. Part of the success of this distribution is due to the existence of multiple methods to sample from it (see example in section 3.2.3.2). In other situations, sampling requires the use of simulation techniques such as MCMC but one still has the problem of specifying an appropriate model.

It seems therefore more natural to use conditional distributions to specify local interactions which are easier to explain and then define the joint probability distribution, up to a constant, through these conditional distributions. Although the relationship between conditional distributions and joint probability distributions is evident in time series, there is no natural order of the different nodes in spatial applications. The link between conditional and joint distributions is not clear, there is no obvious method to infer the joint probability from the associated conditional probabilities. Another disadvantage is that the natural specification of an equilibrium of statistical process is in terms of the joint probability rather that the conditional distribution of the variables. However, Hammersley and Clifford demonstrated in an unpublished paper that one class of stochastic models known as Markov random fields defines consistent joint probability distributions from conditional probability distributions. This result is known as the Hammersley-Clifford theorem and Besag [13] provided a short proof of a special case of this theorem which is adequate for almost all statistical applications. For instance, the Gibbs sampler of Geman and Geman [59] was motivated by the results of this theorem.

### 3.5.1 Markov Random Fields

Let $S = \{s_i : i = 1, \cdots, n\}$ be the locations on a lattice and $\mathbf{Z} = \{Z(s_1), \cdots, Z(s_n)\}$ be a stochastic process where $Z(s_i)$ takes a value in a finite set $\zeta_i$. The joint probability function of the random variables is

$$Pr(\mathbf{z}) = Pr(Z(s_1) = z(s_1), \cdots, Z(s_n) = z(s_n)) \tag{3.47}$$

where $\mathbf{z} = \{z(s_1), \cdots, z(s_n)\}$ and takes values in $\zeta = \prod_{i=1}^{n} \zeta_i$ which is the set of all possible configurations of $z$. $Pr(\cdot)$ denote probabilities since the data are discrete (note that these will be densities when generalising to continuous data). Suppose $S$ is the set of nodes of a graph, the sites in $S$ are related one to another via a neighbourhood system. The

*neighbourhood set*[9] of a site $i$ is defined as

$$\delta_i = \{j : j \sim i\} \tag{3.48}$$

A neighbourhood structure can be defined using different criteria. For instance in image processing the neighbourhoods are defined on the basis of adjacency of picture elements (pixels) [60] whereas in disease applications [88] is usual to base neighbourhod on a distance metric such as the Euclidean metric. Different structures of lattices give different neighbour systems. For instance, on a square lattice, there are for 4 neighbours for every interior site in a first-order neighbourhood, 8 neighbours in a second-order and so on, whereas in a hexagonal lattice [15], the first-order neighbourhood corresponds to 6 neighbours and the second-order to 8. A *clique*, $c$, is a set of sites that consists either of a single site or of sites that are all neighbours of each other (see [91] for detailed figures of cliques on a square lattice and [146] on hexagonal lattices).

The stochastic process is said to be a *Markov random field* if

(i) satisfies the positivity condition, i.e., $Pr(\mathbf{z}) > 0 \; \forall \; \mathbf{z} \in \zeta$

(ii) $Pr(z(s_i)|z(s_j),\ i \neq j) = Pr(z(s_i)|z(s_j),\ j \in \delta_i)$

The second condition represents the Markovian property of the system and is called (in Bayesian jargon) the full conditional distribution. To construct a desirable Markov random field, one must designate the neighbours $\delta_i$ of each site. To preserve (ii) the neighbour relationship must be symmetric.

We are interested now in defining the conditions from which we can specify the joint probability distribution of a MRF. This can be performed using the theoretical work that shows the equivalence between MRF and Gibbs distributions [13].

A Gibbs distribution represents the most general form that the joint probability distribution can take and is described mathematically as

$$Pr(\mathbf{z}) = K^{-1} e^{-\frac{1}{T} U(\mathbf{z})} \tag{3.49}$$

[9] $i \sim j$: site $j$ is neighbour of $i$

where $K = \sum_{\mathbf{z} \in \zeta} e^{-\frac{1}{T}U(\mathbf{z})}$ is a normalizing constant called the *partition function*, $T$ is a constant known as the *temperature* of the system, in general considered to be equal to 1 and $U(\mathbf{z})$ is the *energy function* given by

$$U(\mathbf{z}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{z}) \tag{3.50}$$

and is the sum of clique potentials $V_c$ over all possible cliques $\mathcal{C}$. The value of $V_c$ must be finite, depends on the local configuration on the clique $c$ and can be modeled for desired system behaviour. The Gibbs distribution is said to be *homogeneous* if the clique potentials are independent of the relative position of the cliques and *isotropic* if they are independent of the orientation of the cliques.

A MRF is characterized by the Markovian property by means of which the local properties of the random variables are modeled whereas Gibbs distribution model the global properties of such random variables. Succinctly, the Hammersley-Clifford theorem [13] identifies Markov random field processes with those having the Gibbs property as long as the positivity condition holds. Thus, it is possible to specify the joint probability by means of conditional distributions.

#### 3.5.1.1 Pairwise Difference Priors

In the Bayesian paradigm, information about an underlying spatial process is included in the prior probability, $\pi(\mathbf{z})$, for a random vector[10] $\mathbf{z} = \{z_1, \cdots, z_n\}$. Although in theory the prior distribution should capture global and problem-specific knowledge about $\mathbf{z}$, in practice it is often sufficient to model only the local properties of $\mathbf{z}$. The most used expressions for priors in spatial applications are those which involve only pairwise interactions among the different variables. Although these interactions have been largely criticised in favour of structures with higher-order interactions (see for instance [146] and the references therein) they give acceptably good results for less complexity. A pairwise difference Markov random field prior [17, 19] can be expressed as follows

$$\pi(\mathbf{z}|\gamma) \propto exp\{-\sum_{i \sim j} w_{ij} \Phi(\gamma(z_i - z_j))\} \tag{3.51}$$

[10] to simplify the notation we make $z(s_i) = z_i$

where $\gamma$ is a scale parameter, $\phi(\cdot)$ is an even function, $w_{ij}$ are specified nonzero weights. A range of choices of $\Phi$ can be found in the literature. Examples include the pairwise prior of Geman and McClure [58] used to detect discontinuities in linear image restoration problems or the work of Besag, Green, Higdon and Mengersen [17] in which different pairwise priors are used for a range of applications including tomography and analysis of variety trials. Since in most of the applications the minimal state space is not bounded, $\pi$ is improper. Thus, an additional requirement(see [17]) is that the full conditional distributions of equation 3.51

$$\pi(z_i|z_j, i \neq j) \propto \gamma \; exp\{-\sum_{j \in \delta_i} w_{ij}\Phi(\gamma(z_i - z_j))\} \tag{3.52}$$

are well defined.

#### 3.5.1.2 A Toy Example

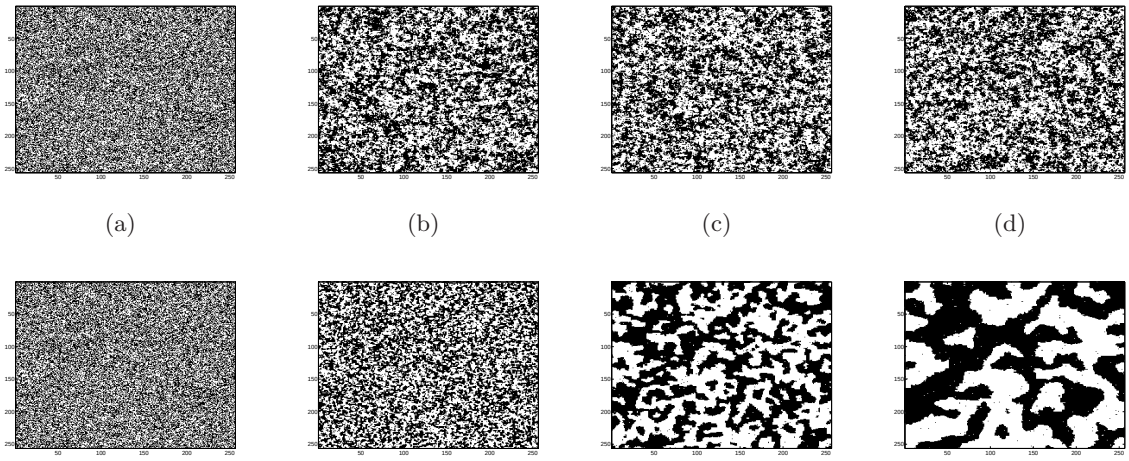
The *Ising model* is a simple model used in statistical mechanics (see [79] for a more physical introduction of the model). Introduced in the 1920's in statistical physics of condensed matter by Ising, the Ising model tries to imitate behaviour in which individual magnetic particles modify their behavior so as to conform to the behavior of other particles in their vicinity. The Ising Model is a pairwise difference prior with $\Phi(u) = -I[u = 0]$ [79], that is,

$$\pi(z|\beta) \propto exp\{\beta \sum_{i \sim j} w_{ij} I[z_i = z_j]\} \tag{3.53}$$

where the parameter $\beta > 0$ statistically favours identity of neighbours and $z \in \{0, 1\}^n$. Note that if $\beta = 0$ equation 3.53 becomes a uniform distribution.

We have simulated the Ising model using a Gibbs sampler, to that, the pixels are updated in turn using algorithm 2 (pag. 41) in a random scan fashion instead of a raster scan. Figure 3.9 illustrates the sampling process for two different values of the parameter $\beta$. As it can be seen from the figure, the parameter $\beta$ determines the strength of the "clustering" of the elements. As $\beta$ increases the dependence between neighbourhing pixels increases.

In many cases such as segmentation and classification problems in image processing, the Ising model is not useful since one requires the use of discrete variables with more than two possible states. For these cases, a more general Ising model called the *Potts model* is



(a) (b) (c) (d)

(e) (f) (g) (h)

**Figure 3.9:** A toy example. Realisations from the Ising model for $\beta = 0.35$ (a) Initial configuration (b) after $10^6$ iterations (c) after $2 \cdot 10^6$ iterations (d) after $10^7$ iterations, and for $\beta = 0.75$ (e) Initial configuration (f) after $10^6$ iterations (g) after $2 \cdot 10^6$ iterations (h) after $10^7$ iterations

used.

As final comment, we would like to mention that one of the main disadvantages of sampling from Ising model with Gibbs (and standard Metropolis-Hastings) samplers is that they can perform poorly since the updates are local. To solve this problem auxiliary variables methods [16] such as the *Swendsen-Wang algorithm* [141] that can lead to improved efficiency has been proposed. Briefly, these methodologies are used to generate a single Markov chain that make substantial changes to several components jointly, mix faster and are easier to simulate. We refer the readers to the work of Higdon [80] for a review of the Swendsen-Wang algorithm and other auxiliary variable algorithms.

### 3.5.2 Notes on Inference for Lattice Models

This section reviews various ways to perform inference for lattice models. Since this thesis is more concerned with the problem of inference in statistical image analysis, we shall focus on the classical techniques performed to face them. Initially, we suppose that a set of records $\mathbf{y}$ is generated by stochastic degradation of a true pixel image $x$. As usual, a Bayesian approach to image analysis begins with the construction of a desirable prior probability distribution, $\pi(\mathbf{x})$, which will include our beliefs regarding the scene of interest.

This prior will include information on local characteristics of the pixels under the assumption that nearby pixels tend to be similar. For instance, in 3D image reconstruction it is normal to use a smoothness prior [96] representing an assumption that physical properties do not change abruptly. The next stage is that of constructing a likelihood function, $L(\mathbf{y}|\mathbf{x})$, based on physical considerations of the image system that models the reality as accurately as possible. By Bayes' theorem, we combine the prior and the likelihood to form the posterior distribution of any image, $\pi(\mathbf{x}|\mathbf{y}) \propto \pi(\mathbf{x})L(\mathbf{y}|\mathbf{x})$, and then we base any inferences about $x$ using this posterior distribution.

If one is only interested in point estimates, one of the most famous methods is that of finding the *maximum a posteriori* (MAP) [95] in which the estimate chosen is that which maximises the posterior probability distribution. However, even for small-size images, it is difficult to obtain the MAP estimates and more efficient optimisation algorithms are needed. In cases where the posterior distributions has a single maximum, Besag's *iterated conditional modes* (ICM) algorithm [14] can be used. ICM is a deterministic algorithm related to the Gibbs sampler in which instead of a random choice of pixel value selects the mode of the relevant conditional distribution. In this way, the posterior distribution never decreases but in cases in which the posterior distribution is multimodal, this algorithm can converge to a local maximum. Morever, this algorithm is very dependent of the initial values.

In their seminal paper, Geman and Geman [59] proposed using simulated annealing (SA) to obtain the exact MAP estimate. SA is a sequential procedure made up of a Gibbs sampler and annealing scheme (see section 3.2.4 for more details on the sampler). Besag [14] commented that the ICM algorithm is a zero-temperature version of SA. One of the main advantages of this algorithm is that it can be parallelised and therefore the performance can be improved. Although this important feature is often mentioned, parallelisation is rarely performed (see for instance [136] and the references therein).

In classification problems in image analysis it is common to use the *marginal posterior modes* (MPM) estimates [36] in which, instead of calculating the MAP of correct classification of the entire scene, we minimize the expected number of missclassifications under the model by choosing, for each pixel, the class that has maximum posterior probabil-

ity given the data. In general, it is very rare that the marginal conditional distributions are available in closed form and we depend on numerical techniques to obtain accurate estimates. Finally, we can use MCMC methods to obtain estimates of the posterior distribution in image analysis applications [60]. MCMC techniques not only give point estimates of a true image but also represent the uncertainty in such estimates through the posterior distribution.

#### 3.5.2.1 Parameter Estimation

In many applications, the (hyper-)parameters of the prior distribution are also subject to inference. For instance, one would like to estimate the parameter $\gamma$ of equation 3.51 to assess the strength of the correlation among the different elements of the lattice. However, estimation of these parameters is not always straightforward and classical statistical analysis such as *maximum likelihood estimation* (MLE) requires an enormous amount of computation even for a single parameter. The Bayesian approach is unproblematic either. The posterior distribution takes the form

$$\pi(x, \gamma|y) \propto L(y|x)\pi(x|\gamma)\pi(\gamma) \tag{3.54}$$

which involves a (hyper-)prior distribution, $\pi(\gamma)$, for the parameter of interest. In both cases, the normalising constant (the partition function in MRF) of $\pi(x|\gamma)$ is needed since the parameter $\gamma$ is considered as random and therefore we cannot get rid of it and must be used in equation 3.54. This normalising constant is in general intractable and complicates a fully Bayesian approach.

To cope with this problem, several methods has been proposed. Besag [14] used a maximum *pseudolikelihood* estimator which approximates the likelihood function with a product of conditional likelihoods of each variable given its neighbours. The advantage of this methodology is that does not require the calculation of the "full" normalising constant. From a Bayesian perspective, there has also been approaches to the calculation of the normalising constant such as the *acceptance ratio method* and the *thermodynamic integration approach* which are effective methods originated from theoretical physics. For a review of these methods, we refer the readers to the work of Gelman and Meng [56] and Higdon [79].

Chapter 4

# Bayesian Analysis of Lidar Signals

*If reason (I mean abstract reason, derived from inquiries a priori) be not alike mute with regard to all questions concerning cause and effect, this sentence at least it will venture to pronounce, that a mental world, or universe of ideas, requires a cause as much, as does a material world, or universe of objects; and, if similar in its arrangement, must require a similar cause. For what is there in this subject, which should occasion a different conclusion or inference? In an abstract view, they are entirely alike; and no difficulty attends the one supposition, which is not common to both of them.*

*David Hume.*

## 4.1 Introduction

We consider the processing of time-of-flight laser radar (Ladar) signals to obtain accurate range measurement and 3D reconstruction of surfaces at ranges from a few metres to several kilometres. Normally, such systems process only an assumed single return from a single opaque surface. In recent years, interest has grown in situations in which the laser return consists of multiple peaks, for example due to the footprint of the beam impinging on a target with surfaces distributed in depth or with semi-transparent surfaces [70, 78, 128, 155]. The goal of our work is a complete characterisation of the 3D surfaces viewed by the laser imaging system. Although we do apply our analyses to 3D image data, the algorithms we employ in this chapter consider each pixel or scanned element in isolation, i.e. independently of adjacent elements in image space, as a first step in a fuller analysis that takes account of between pixel dependencies.

We make a number of contributions. First, we present a unified theory of pixel processing

that is applicable to both Time-Correlated Single Photon Counting (TCSPC) and Burst Illumination Laser (BIL). To that end, we adopt a Bayesian statistical approach based on reversible jump Markov chain Monte Carlo (RJMCMC) techniques [63, 119] to assess the number, positions and amplitudes of the returned signals from target surfaces. The main advantages of the Bayesian approach are that it facilitates representation, takes a fuller account of the uncertainties related to models and parameter values, and provides a natural method for updating beliefs in response to new information and therefore to incorporate prior knowledge into the analysis. Second, we incorporate a delayed rejection step [65] in a RJMCMC algorithm that permits temporary local adjustments of the proposal distribution. In short, the delayed rejection step allows the Markov chain to mix better through the use of different proposal distributions and therefore to explore better the posterior distribution derived from the Bayesian approach. In Section 4.2, we introduce the TCSPC and BIL approaches and discuss related work. In Section 4.3 we present our models which describe TCSPC and BIL data. In Section 4.4 we describe our research methodology using a Bayesian framework. In Section 4.5 we explain how RJMCMC can be used to approximate the full joint posterior distribution obtained in the previous Section. In Sections 4.6.1 and 4.7 we present the effectiveness of the algorithm developed through series of different experiments using synthetic and real data. Finally, in Section 4.8 we present some conclusions.

## 4.2 Background and Related Work

We employ two principal approaches to range measurement and depth imaging, by TCSPC [3, 154], and by range gating, or repeated BIL [23, 46]. With reference to figure 4.2, a pulsed laser is directed towards the target; the time of flight (TOF) of the laser to and from the target gives a measurement of distance since the speed of light in the medium is known. However, in the TCSPC method, the detector (a single photon avalanche photodiode (SPAD)) records the arrival of the first photon; repeated laser pulses result in a histogram of photon arrival times that can be processed to determine the target range and signature. In the BIL method, the detector (in our case an electron-bombardment charge-coupled device (CCD)) integrates the received laser light over a time period (typi-

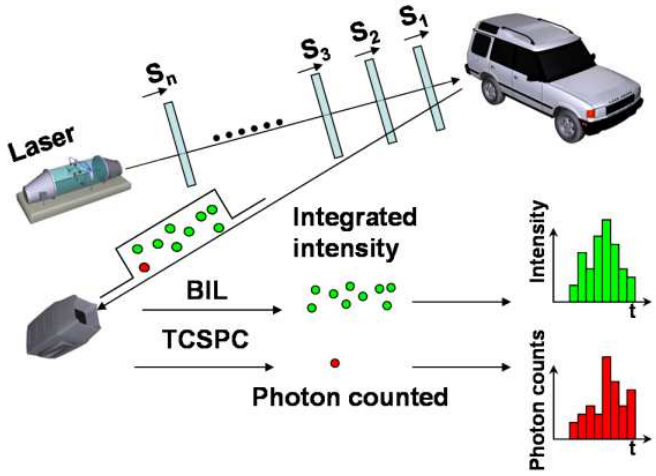



**Figure 4.1:** TCSPC and BIL imaging principles

cally a few *ns*). Thus, the pulsed laser-CCD detector pair records the laser return from a distant slice of space. The range data is created from the multiple flashes by variable range gating. A histogram of integrated intensities is created with the time axis determined by the range gate delay. In either case, this pattern of recorded photon arrival times or gate delays is analysed to find the distribution of depth and reflectance in the 3D scene.

To create a full 3D image (or "angle-angle-range" image) in the TCSPC method, the laser is normally scanned but there have been advances in the development of arrays of SPADs (or geiger-mode APDs) [3, 7]. In a BIL system, the CCD records a full 2D slice in 3D space at each range gate so the 3D image is built from the repeated laser pulses as the range gate changes. The BIL images are acquired by a sensor of similar specification to that described by Duncan et al. [46], and consist of multiple frames that, in our case, have been typically acquired at thirty different range gate delays. The relative merits of the respective approaches are discussed in the references given. It is found generally that the TCSPC approach has greater sensitivity and accuracy of measurement than the range gated method whilst the creation of the full 3D image is much faster in the BIL case. As mentioned previously, current classical processing methods only consider a single return. In general, this single return corresponds to either the first pulse or the last pulse received. The peak detector is a simple threshold comparator or a matched filter. The system used therefore requires a reasonably high signal-to noise ratio to be able to discriminate the return from the background noise. When the number of returns is increased, the approach

chosen is that of implementing first/last pulse logic in the receiver electronics. In first pulse logic, this involves processing the first pulse detected whereas in last pulse logic, this involves sampling and holding all the returns after the first but only recording the final return [70]. In practice, this fails because the pulse of interest is not obvious and more complex processing of the returns is needed. Schilling et al. [128] showed that the use of the entire multiple returns signal rather than a subset thereof provides more information about the scene of interest. For instance, in a defence application the problem may be to detect a target under camouflage netting or foliage. In geosciences and remote sensing the problems may be to produce a digital terrain model of a landscape or to estimate biomass parameters and the canopy structure of forested terrains [113].

Stochastic algorithms are computationally more intensive than methods such as matched filtering and threshold detection and can take significantly longer to reach a solution. Nevertheless, they may give estimates of the range to higher resolution than the classical algorithms. Bayesian methods to deal with the problem of mixtures with an unknown number of components are described in section 3.4.2, so we shall consider here other stochastic methods. Typical processing methods calculate the Maximum Likelihood Estimates (MLE) of the parameters in the model within a frequentist framework [154]. One of the most used tools for finding the MLE is the Expectation-Maximization (EM) algorithm [39]. The main problem is that this algorithm can converge to a local maximum or a saddle point of the likelihood. Moreover, this methodology is sensitive to the initial estimates. Previous statistical approaches to deal with the problem of multiple returns include non-parametric estimation using kernel-density estimation techniques and parametric estimation using the method of moments [145]. The former approach can fail to distinguish whether or not a return is composed of a single pulse broadened after passing through the atmosphere or several pulses which are very close together. Therefore, these methods do not give any physical interpretation to the components of the mixture and may make inefficient use of data. They may fail to give accurate estimates when dealing with small samples and, in general, the maximum likelihood estimation gives a better representation of the true quantities. To identify the number of components in a mixture distribution, McLachlan [100] proposed to bootstrap the distribution of the likelihood ratio test statistic, however, this technique can identify too many components. Wax and

Kailath [157] proposed detecting the number of signals in a multichannel time-series using information theoretic criteria for model selection such as Akaike (AIC) and Schwartz and Rissanen's Minimum Description Length (MDL). In practice, they calculated for every possible number of signals the MLE estimates of the parameters of the signals and then they calculated either the AIC or MDL for each model. The number of signals was determined by minimizing either the AIC or MDL criterion. A Bayesian variant of the AIC is the Bayesian Information Criterion [22] which assumes equal priors on each model and non-informative priors on the parameters. Since the maximum number of signals must be fixed a priori, there is a risk of underestimation of the actual number of peaks and therefore of not exploring the space of solutions completely. In general, the MLE of the parameters are obtained using the EM algorithm. Celeux and Diebolt [29] proposed the use of a stochastic EM (SEM) that can be seen as a random perturbation of the traditional EM algorithm. The SEM algorithm replaces the E step by incorporating a stochastic step to avoid entrapment in a local maximum of the likelihood function. A hybrid approach is found in Wallace et al. [156]; they used a deterministic non-parametric "bump-hunting" procedure that provided the number, amplitudes and positions of the suspected returns. Once the initial estimates were obtained they computed the MLE of the set of parameters that best explained the data. This methodology may fail when the number of returns increases or the returns are occluded by background noise.

## 4.3 Model

The exact functional forms of the photon count and the intensity histograms are unknown. In the case of photon count data we employ a nearest representation or operating model to describe the underlying distribution. This operating model is described in [109] and represents the parametric form of the expected temporal variation of the photon count distribution as shown in equation. 4.1.

$$f_{op}(i;p) = \beta \begin{cases} e^{(\frac{-(t_1-t_0)^2}{2\sigma^2})} e^{(\frac{(i-t_1)}{\tau_1})} & i < t_1 \\ e^{(\frac{-(i-t_0)^2}{2\sigma^2})} & t_1 \leq i < t_2 \\ e^{(\frac{-(t_2-t_0)^2}{2\sigma^2})} e^{(\frac{-(i-t_2)}{\tau_2})} & t_2 \leq i < t_3 \\ e^{(\frac{-(t_2-t_0)^2}{2\sigma^2})} e^{(\frac{-(t_3-t_2)}{\tau_2})} e^{(\frac{-(i-t_3)}{\tau_3})} & i \geq t_3 \end{cases} \tag{4.1}$$

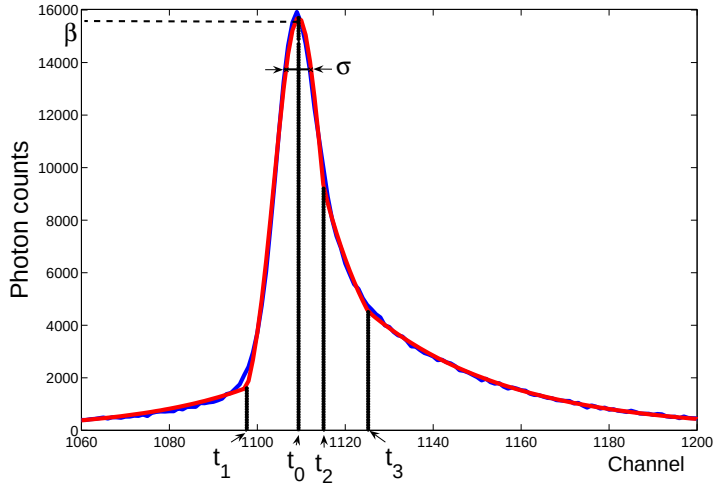



**Figure 4.2:** Histogram of photon counts of the real data (blue) and initial fit to obtain the shape parameters (red). The different parameters of the "operating" model are also shown.

where $p = (\beta, \sigma, t_0, t_1, t_2, t_3, \tau_1, \tau_2, \tau_3)$ is the parameter set, $\beta$ is an amplitude factor, $t_0$ is the time of the peak maximum, and $t_1$, $t_2$ and $t_3$ are the points at which the changeovers between functions occur as shown in Fig. 4.2. In this study, we assumed that the shape parameters of the returned pulse are fixed and known from an instrumental response and therefore the only parameters subject to inference are the amplitude $\beta$ and the time of the peak maximum $t_0$. For a single peak, the new parameter set is therefore $p' = (\beta, t_0)$.

In the case of the intensity histogram, we used a look-up table of an instrumental function, $w_{op}(i, \beta, t_0) = \beta \cdot w(i, t_0)$, instead of an operating model. In practice this instrumental function is acquired from the response of the BIL system to a Lambertian reflecting surface at a similar range to the object of interest as shown in Fig. 4.3. Over this distribution, only two parameters were considered to be variable: the amplitude $\beta$ and the time of the peak maximum, $t_0$.

However, there are many cases of several returns that may be overlapping in position and possibly weak in comparison with the noise or the background level, either because they have been observed at a distance or because the reflectance is low, *e.g.* due to different alignment of the target and the sensor or the presence of obscurants on the target. Assuming several signals can be present in the same histogram and that these signals will be observed against a finite background level, $B$, whose expected value is considered as

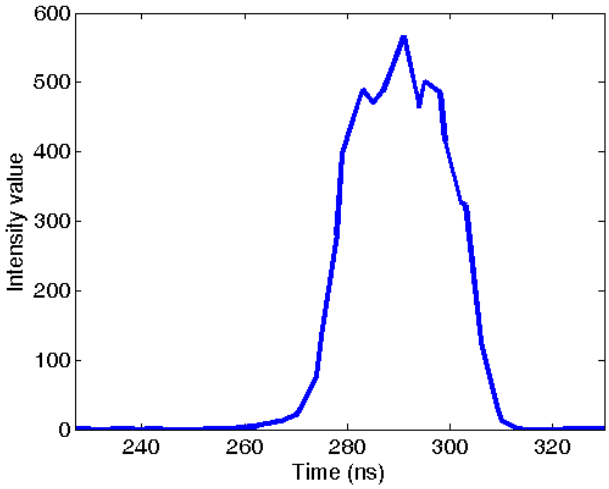



**Figure 4.3:** Example of a BIL intensity profile

a constant across all the channels, the observed photon (intensity) histogram, $\boldsymbol{y}$, can be considered as a sample of a non-normalized statistical mixture distribution with intensity

$$\mathcal{F}(i;k,\phi) = \sum_{j=1}^{k} f_{system}(i;\beta_j,t_{0_j}) + B \tag{4.2}$$

where $k$ is the number of peaks and $\phi$ is the set of parameters of each signal and the background: $\phi = (\boldsymbol{\beta}, \boldsymbol{t_0}, B)$ with $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)$ and $\boldsymbol{t_0} = (t_{0_1}, t_{0_2}, ..., t_{0_k})$. $f_{system}$ is either $f_{op_j}(i, \beta_j, t_{0_j})$ or $w_{op_j}(i, \beta_j, t_{0_j})$ depending on the data we are analysing.

If the time resolution is sufficiently fine, the value $y_i$ recorded in each channel $i$ can be considered as a random sample of a Poisson distribution with intensity $\mathcal{F}(i;k,\phi)$ which depends on the model parameters:

$$P(y_i|k,\phi) = e^{-\mathcal{F}(i;k,\phi)} \frac{\mathcal{F}(i;k,\phi)^{y_i}}{y_i!} \tag{4.3}$$

Equation 4.3 represents the physics of sensor operation, where $y_i$ is the number of photons (respectively the intensity value in a BIL histogram) in channel $i$. Furthermore, assuming that the observations recorded in each channel $i$ of the histogram are conditionally independent given the value of the parameters, the joint probability function of $\boldsymbol{y}$ is defined as:

$$L(k,\phi|\boldsymbol{y}) = \prod_{i=1}^{i_{\max}} e^{-\mathcal{F}(i;k,\phi)} \frac{\mathcal{F}(i;k,\phi)^{y_i}}{y_i!} \tag{4.4}$$

where $\boldsymbol{y} = \{y_1, y_2, \ldots, y_{i_{max}}\}$.

## 4.4 Bayesian Inference

The objective of this study is inference about the unknowns $k$ and $\phi$ in order to obtain accurate estimates of the position, amplitude, number of peaks and background of the returned signals. These unknowns are regarded as drawn from appropriate independent prior distributions. The position parameter, $t_{0_j}$, is drawn from a flat uniform distribution with support in the interval $[0, i_{max}]$, to reflect our complete prior ignorance of the true position of the peaks. Likewise, for the number of peaks, we take a uniform distribution on $[0, ..., k_{max}]$ with $k_{max}$ chosen to be suitably large. Converting the results obtained with this prior to those corresponding to other priors on $k$ is straightforward [119]. For the prior distribution of scale parameters $\beta_j$ and $B$ we assume a gamma distribution which is a natural choice in the statistics community for parameters whose support is in $\Re^+$. The full joint prior distribution of $k$ and $\phi = (\boldsymbol{\beta}, \boldsymbol{t_0}, B)$ can be modelled as

$$\boldsymbol{f}(k, \phi) = \frac{1}{k_{max}} \left( \frac{1}{i_{max}} \right)^k f_G(B|c, d) \prod_{n=1}^{k} f_G(\beta_n|a, b) \tag{4.5}$$

where $f_G$, is the probability density function (PDF) of a gamma distribution with mean $a \cdot b$ (respectively $c \cdot d$) and variance equal to $a \cdot b^2$ ($c \cdot d^2$ respectively).

In the Bayesian paradigm, the information brought by the data, the likelihood, is combined with prior information and summarised in a probability distribution called the posterior distribution, our target distribution. This posterior distribution captures knowledge about the parameters given the data. Thus, using equationa 4.4, 4.5 and Bayes' theorem the target distribution can be expressed as

$$\pi(k, \phi|\boldsymbol{y}) = \frac{L(k, \phi|\boldsymbol{y}) \boldsymbol{f}(k, \phi)}{\int L(k, \phi|\boldsymbol{y}) \boldsymbol{f}(k, \phi) \delta(k, \phi)} \propto L(k, \phi|\boldsymbol{y}) \boldsymbol{f}(k, \phi) \tag{4.6}$$

In general, equation 4.6 is intractable and difficult to analyse. Powerful simulation algorithms, such as MCMC, allow posterior distributions to be explored and conclusions and inferences to be drawn directly from the sampled values.

## 4.5 RJMCMC Methodology

The theory underlying MCMC techniques was explained in Section 3.2. MCMC techniques enable simulation from an unknown distribution by embedding it as the limiting distribution of a Markov chain and simulating samples from that chain until it approaches equilibrium. When the state space is a set of a countable number of subspaces of different dimension, standard MCMC algorithms are not applicable and other algorithms are needed to explore all these different subspaces. The reversible jump MCMC approach (see [63, 119]) explained in Section 3.3 copes with this problem. This is an extension of the MCMC algorithm designed to allow jumps between subspaces of different dimension.

We followed a similar approach to that described in [119] by constructing a Markov chain whose transitions involve changes to the number, positions and amplitudes of the peaks in the return signal. However, in contrast to [119], we do not associate each photon in the histogram with a particular component in a mixture model but treat the histogram as a discrete representation of a spatially heterogeneous Poisson process whose intensity is a linear superposition of the scaled and shifted templates of the form specified by equation 4.1. The transitions of the Markov chain we generate involve moves of various types. The different moves we allow are:

1. Updating the positions $t_0$,
2. Updating the amplitudes $\beta$,
3. Updating the background $B$,
4. Random birth or death of a peak,
5. Random splitting of a peak into two peaks or merging of two peaks into a single peak.

One complete pass over these five moves is called a sweep.

### 4.5.1 Fixed-Dimension Moves

Moves of type (1), (2) and (3) allow the posterior distribution to be explored in a state space with fixed dimension, i.e., in this particular case when the number of peaks does not vary, $k = k'$. To simplify notation, we will write $\pi(k, \phi|\boldsymbol{y}) = \pi(\phi|\boldsymbol{y})$ when $k$ is fixed. Furthermore, to improve the rate of accepted values, the different parameters are updated in isolation obtaining three different acceptance probabilities, one for each parameter. Initially, when $k = 0$, moves (1) and (2) are not available since there is only background noise.

The task is to construct a Markov chain on a general state space with $\pi(\phi|\boldsymbol{y})$ as its limiting distribution. Once convergence towards this limiting distribution is achieved, we form a realisation of the chain $\{\Phi^{(1)},\ \Phi^{(2)},\ ...,\ \Phi^{(N)}\}$ and treat it as a correlated sample from $\pi(\phi|\boldsymbol{y})$. By means of the ergodic theorem, it is possible to use this realisation of the chain to build histograms, estimate moments, calculate probabilities, etc. These realisations are drawn from an appropriate transition kernel which embodies the Markov property. As explained in chapter 3, to construct a suitable transition kernel, the Metropolis-Hastings algorithm and the Gibbs sampler (and variants of them) can be used to obtain these realisations. However, we are dealing with complex models where full conditional distributions do not belong to any known standard family of distributions and therefore it is not possible to draw samples from them directly. The Metropolis-Hastings approach is more appropriate to this problem yielding the following acceptance probability

$$\alpha(\phi, \phi') = min \left\{ 1,\ \overbrace{\frac{\pi(\phi'|y)}{\pi(\phi|y)}}^{Target\ ratio} \times \overbrace{\frac{q(\phi, \phi')}{q(\phi', \phi)}}^{Proposal\ ratio} \right\} \tag{4.7}$$

Equation 4.7 will determine whether a candidate value is accepted or rejected and it is a key element in the generation of the Markov chain. There are no restrictions in the choice of the proposal distribution, $q(\phi, \phi')$. However persistent rejection may indicate that locally the proposal distribution is badly calibrated to the target. Although this rejection is necessary to preserve the stationary distribution, if it is too high, the correlation between consecutive samples of the generated chain will increase. In particular, for random-walk type samples, this is very important for modeling purposes.

As explained in section 3.3.5, the delayed rejection step is a strategy considered to improve the performance of the RJMCMC methods to cope with this problem and is contemplated as an alternative to careful off-line tuning of state-dependent proposals. In our case, this methodology is used to better explore the posterior distribution of the different parameters since the range of values with which we work is very large.

Tierney and Mira [144] and later Green and Mira [65], for a variable dimension approach, proposed that whenever a candidate move is rejected, we make another attempt to move using a second proposal instead of turning to the next transition. The proposed distribution at this second stage is allowed to depend on the previous rejected value as well as the current position of the chain. Although the acceptance probability of the new candidate has to be adjusted to preserve the stationary distribution, this methodology does not destroy the Markovian property of the sampler. This process can be repeated indefinitely.

If the dimension is fixed, the methodology is the following: When at $\phi$ we propose a new state $\phi'$ from a density $q_1(\phi, \phi')$. This move is accepted using a probability similar to that of equation 4.7. If the move is rejected, a second proposal, $\phi''$, is made from a density $q_2(\phi, \phi', \phi'')$ and the new acceptance probability is $\alpha(\phi, \phi'') = \min\{1, A\}$, where $A$ is

$$A = \frac{\pi(\phi'')q_1(\phi'', \phi')q_2(\phi'', \phi', \phi)\{1 - \alpha_1(\phi'', \phi')\}}{\pi(\phi)q_1(\phi, \phi')q_2(\phi, \phi', \phi'')\{1 - \alpha_1(\phi, \phi')\}} \tag{4.8}$$

In this chapter, we use only the delayed-rejection algorithm when the dimension of the space of solutions is fixed as is the case with the parameter updating step. The mathematical framework to obtain this acceptance probability is explained in section 3.3.5.

#### 4.5.1.1 Updating the Positions

To update a position, our Markov chain explores the space of solutions using a random walk proposal. The proposal distribution is therefore

$$q(\boldsymbol{t'_0}, \boldsymbol{t_0}) = \prod_{j=1}^{k} q_{\mathcal{N}}(t'_{0_j} | t_{0_j}, \sigma_1) \tag{4.9}$$

where $q_{\mathcal{N}}$ denotes a normal PDF whose mean is equal to the current value of $t_o$ and with a standard deviation $\sigma_1$ which is chosen according to the size of the data vector (TCSPC

or BIL). The acceptance probability is $\alpha_1(\boldsymbol{t_0}, \boldsymbol{t'_0}) = \min\{1, A_1\}$ with $A_1$ equal to

$$A_1 = \frac{L(\boldsymbol{y}|k, \beta, t'_0, B)}{L(\boldsymbol{y}|k, \beta, t_0, B)} \tag{4.10}$$

Since $\sigma_1$ is in general sufficiently large to handle efficient exploration of the range of $t_o$, we need to use a delayed rejection algorithm in order to improve the mixing of the chain and therefore to explore in some way more locally to refine the results. This second stage also used a normal PDF as proposal with mean value equal to $\boldsymbol{t_0}$ and standard deviation $\sigma_2$, the acceptance probability is $\alpha_2(\boldsymbol{t_0}, \boldsymbol{t''_0}) = \min\{1, A_2\}$ with $A_2$ equal to

$$A_2 = \frac{\{1 - \alpha_1(t''_0, t'_0)\}}{\{1 - \alpha_1(t_0, t'_0)\}} \times \frac{L(\boldsymbol{y}|k, \beta, t''_0, B)}{L(\boldsymbol{y}|k, \beta, t_0, B)} \tag{4.11}$$

#### 4.5.1.2 Updating the Amplitudes

To update the amplitudes, we use a gamma distribution denoted as $q_G$ with index equal to 1.5 and the current value of the Markov chain as follows:

$$q(\beta', \beta) = \prod_{j=1}^{k} q_G(\beta'_j|1.5, \beta_j) \tag{4.12}$$

and acceptance probability $\alpha_1(\beta, \beta') = \min\{1, A_1\}$ with $A_1$ equal to

$$A_1 = \frac{L(\boldsymbol{y}|k, \beta', t_0, B)}{L(\boldsymbol{y}|k, \beta, t_0, B)} \times \prod_{j=1}^{k} \frac{f_G(\beta'_j|a, b)}{f_G(\beta_j|a, b)} \frac{q_G(\beta_j|1.5, \beta'_j)}{q_G(\beta'_j|1.5, \beta_j)} \tag{4.13}$$

Since we are using a gamma distribution with scale parameter depending on the current value of the Markov chain, the rate of accepted values decreases as the current value increases. Indeed, if the current value is large, the gamma distribution will be more diffuse and vice versa. To solve this problem, we applied again delayed rejection using as a second proposal a normal distribution with mean value equal to $\beta$ and standard deviation $\sigma_3$. The second stage acceptance probability is $\alpha_1(\beta, \beta'') = \min\{1, A_2\}$ with $A_2$ equal to

$$A_2 = \frac{L(\boldsymbol{y}|k, \beta'', t_0, B)}{L(\boldsymbol{y}|k, \beta, t_0, B)} \times \frac{\{1 - \alpha_1(\beta'', \beta')\}}{\{1 - \alpha_1(\beta, \beta')\}} \times \prod_{j=1}^{k} \frac{f_G(\beta''_j|a, b)}{f_G(\beta_j|a, b)} \frac{q_G(\beta'_j|1.5, \beta''_j)}{q_G(\beta'_j|1.5, \beta_j)} \tag{4.14}$$

#### 4.5.1.3 Updating the Background

To perform this move, we propose a new estimate from a gamma distribution, $q_G$ with shape parameter equal to 1.5 and scale parameter the current value of the Markov chain.

The acceptance probability is $\alpha_1(B, B') = \min\{1, A\}$ with $A$ equal to

$$A = \frac{L(\boldsymbol{y}|k, \beta, t_0, B')}{L(\boldsymbol{y}|k, \beta, t_0, B)} \times \frac{f_G(B'|c, d)}{f_G(B|c, d)} \frac{q_G(B|1.5, B')}{q_G(B'|1.5, B)} \tag{4.15}$$

### 4.5.2 Variable Dimension Moves

In addition to learning about the parameters of a model, we might be interested in comparing different models. Such models could have equal dimension and we may want to find which one explains better the data, whether we have a mixture of distributions where the unknown is the number of elements, or perhaps a combination of both [152].

Our problem has a parameter space whose dimension is itself subject to inference, i.e. the number of returns is unknown. The goal is therefore to explore the different parameter subspaces in a single simulation run. Using the Bayesian framework, the problem is now formulated as a joint probability distribution of a model indicator, $k \in \{1, 2, 3, ..., K\}$, and a real stochastic vector of parameters of possibly varying dimension, $\phi$, given the data $y$, that is, $\pi(k, \phi|\boldsymbol{y})$. The parameter vector $\phi$ takes values in the set $\xi$ which defines the general space of solutions, $\xi = \cup_K \xi_k$. Given $K = k$, the parameter vector $\phi$ can only take values in $\xi_k$. In our problem, the set $\xi_k$ is taken to be $\xi_k \subseteq \Re^{n_k}$, where $n_k$ is the dimension of the space. Using the results of section 3.3.3 the general acceptance probability is given by

$$\alpha_m(k, \phi, k', \phi') = \min\left\{1, \frac{\pi(k', \phi'|\boldsymbol{y})}{\pi(k, \phi|\boldsymbol{y})} \frac{j_m(\phi')}{j_m(\phi)} \frac{g'_m(u')}{g_m(u)} \left|\frac{\delta(\phi', u')}{\delta(\phi, u)}\right|\right\} \tag{4.16}$$

for a move type $m$, $\left|\frac{\delta(\phi', u')}{\delta(\phi, u)}\right|$ is the Jacobian term which appears due to the deterministic transformation performed, $j_m(\cdot)$ the probability of performing a particular move and $g(\cdot)$ is the density from which the random values $u$ are drawn.

The birth/death(4) and split/merge(5) steps imply changes in the number of peaks (hence the number of parameters) of our distribution and therefore the reversible jump mechanism is needed. In these two specific cases the acceptance probability of equation. 3.34 can be reduced to

$$\alpha_m(k, \phi, k', \phi') = min\left\{\overbrace{\frac{\pi(k', \phi'|\boldsymbol{y})}{\pi(k, \phi|\boldsymbol{y})}}^{Target\ ratio} \times \overbrace{\frac{j_m(\phi')}{j_m(\phi)q(u)}}^{Proposal\ ratio} \times \overbrace{\left|\frac{\delta(\phi')}{\delta(\phi, u)}\right|}^{Jacobian}\right\} \tag{4.17}$$

Let $b_k$, $d_k$, $s_k$ and $m_k$ be the probabilities of a birth, death, split and merge respectively, where $k$ is the current number of peaks. Further, assume that when they are all allowed, they are equally probable. The following conditions apply:

- $b_k + d_k + s_k + m_k = 1$.
- If the current number of peaks, $k$, is equal to 0, that is, only background noise is present, only a birth step is allowed.
- If the current number of peaks, $k$, is equal to 1, only a birth, death or split step is allowed (with equal probability).
- If the current number of peaks, $k$, is equal to a predefined maximum number of peaks, $k_{\max}$, only a death or a merge step is allowed.

#### 4.5.2.1 Birth/Death Move

For the birth/death move we make a random choice between attempting to create a new peak or to delete an existing peak with probabilities $b_k$ and $d_k$. For a birth, the value of the amplitude $\beta_{k+1}$ and the position $t_{0_{k+1}}$ of the proposed new peak are drawn from:

$$\begin{aligned} \beta_{k+1} &\sim \Gamma(1, (\tfrac{1}{i_{\max}}) \textstyle\sum_{i=1}^{i_{\max}} y_i) \\ t_{0_{k+1}} &\sim \mathcal{U}(0, i_{\max}) \end{aligned} \tag{4.18}$$

The proposal ratio is therefore:

$$\frac{d_{k+1}}{b_k} \times \frac{1}{k+1} \times \{g_1(\beta_{k+1}) g_2(t_{0_{k+1}})\}^{-1} \tag{4.19}$$

where $g_1$ denotes a gamma PDF with shape parameter equal to 1 and scale parameter corresponding to the mean value of $\boldsymbol{x}$, and $g_2$ denotes a *uniform* PDF on the interval $[0, i_{max}]$.

From equation 4.17, 4.18 and 4.19 the expression for $A$ in the acceptance probability, $\alpha_m(k, \phi, k', \phi') = min\{1, A\}$, for the birth case can be expressed as:

$$\begin{aligned} A = \frac{L(\boldsymbol{y}|k+1, \beta_1, \ldots, \beta_{k+1}, t_0, \ldots, t_{0_{k+1}}, B)}{L(\boldsymbol{y}|k, \beta_1, \ldots, \beta_k, t_0, \ldots, t_{0_k}, B)} \times \frac{d_{k+1}}{b_k} \times \frac{1}{k+1} \\ \times \{g_1(\beta_{k+1}) g_2(t_{0_{k+1}})\}^{-1} \times f_G(\beta_{k+1}|a, b) \times \frac{1}{i_{max}} \end{aligned} \tag{4.20}$$

For the corresponding death move, we make a random choice between any existing peaks, the acceptance probability is $\alpha_m(k, \phi, k', \phi') = min\left\{1, A^{-1}\right\}$, with an appropriate change of index and $k' = k - 1$. For the birth and death step, the determinant of the Jacobian is equal to unity.

#### 4.5.2.2 Split/Merge Move

The split/merge move is more complicated. The choice of the deterministic function which links both moves is not straightforward since the "dimension-balancing" condition given by equation 3.29 has to be satisfied.

The split proposal begins by choosing a peak, $k_n$, at random to split into two new peaks, $k'_n$ and $k''_n$, with amplitude and position value equal to $(\beta'_n, t'_{0_n})$ and $(\beta''_n, t''_{0_n})$ respectively, conforming to equation 4.21

$$\begin{aligned} &u_1 \sim \mathcal{U}(0,1), && \Delta \sim \mathcal{U}(0, \Delta_1) \; where \; \Delta_1 = (t_{3_n} - t_{0_n}) \\ &\beta'_n = \beta_n \; u_1, && \beta''_n = \beta_n(1 - u_1) \\ &t'_{0_n} = t_{0_n} - u_1\Delta, && t''_{0_n} = t_{0_n} + u_1\Delta \end{aligned} \tag{4.21}$$

where $t_{0_n}$ is the time of the peak $n$ maximum and $t_{3_n}$ is the point at which the changeover between the third term and the fourth term of equation 4.1 occurs.

The proposal ratio of equation 4.17 becomes:

$$\frac{m_{k+1}}{s_k} \times \frac{1}{k+1} \times \{g_1(u_1)g_2(\Delta)\}^{-1} \tag{4.22}$$

where $g_1$ denotes the Uniform PDF, $\mathcal{U}(0,1)$, and $g_2$ denotes the Uniform PDF, $\mathcal{U}(0, \Delta_1)$.

The acceptance probability of this move is $\alpha(k, \phi, k', \phi') = min\{1, A\}$ where $A$ is defined as:

$$\begin{aligned} A = &\frac{L(\boldsymbol{y}|k+1, \beta', t'_0, B)}{L(\boldsymbol{y}|k, \beta, t_0, B)} \times \frac{m_{k+1}}{s_k} \times \frac{1}{k+1} \times \{g_1(u_1)g_2(\Delta)\}^{-1} \\ &\times \frac{f_G(\beta'_n|a,b) \times f_G(\beta''_n|a,b)}{f_G(\beta_n|a,b)} \times \frac{1}{i_{max}} \times 2\beta_n u_1 \end{aligned} \tag{4.23}$$

with $\beta' = \{\beta_1, \ldots, \beta'_n, \beta''_n, \ldots, \beta_k\}$, $t'_0 = \{t_{0_1}, \ldots, t'_{0_n}, t''_{0_n}, \ldots, t_{0_k}\}$, $\beta = \{\beta_1, \ldots, \beta_n, \ldots, \beta_k\}$, $t_0 = \{t_{0_1}, \ldots, t_{0_n}, \ldots, t_{0_k}\}$, $2\beta_n u_1$ is the Jacobian of the transformation from $(\beta_n, t_{0_n}, u_1, \Delta)$ to $(\beta'_n, t'_{0_n}, \beta''_n, i''_{0_n})$ and $f_G$ is the prior for $\beta$.

To perform the merge move we choose peaks that are adjacent in terms of $\Delta_1$. The acceptance probability is now $\alpha_m(k, \phi, k', \phi') = min\left\{1, A^{-1}\right\}$ with the corresponding change of index and $k' = k - 1$. The following parameter restructuring is also needed:

$$\beta_n \ = \beta'_n + \beta''_n, \quad t_{0_n} = \frac{t'_{0_n} + t''_{0_n}}{2} \tag{4.24}$$

## 4.6 Experimental and Simulated Results; Single Return.

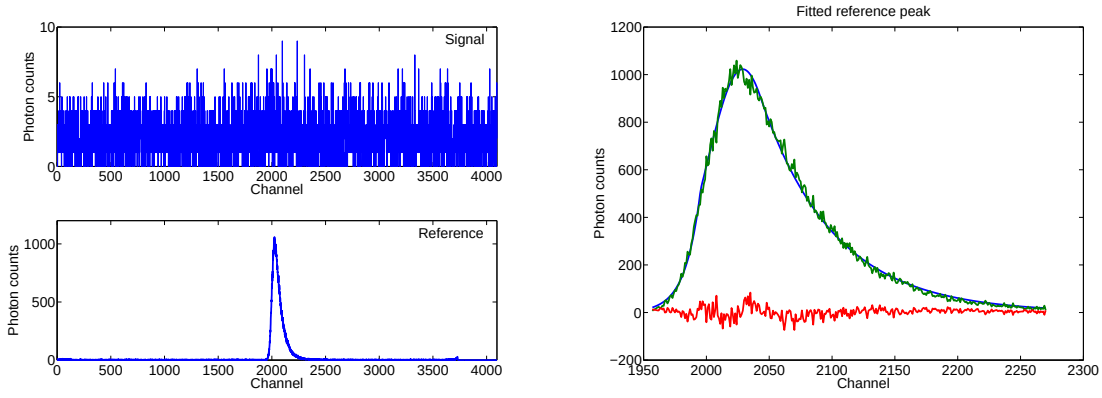
In this section, we present results on both simulated and real data having a single peak. To evaluate the MCMC we have performed some simulations in critical conditions, such as low signal-to-background ratio, comparing the estimates obtained with the semiparametric approach (simplified to one peak detection) explained in [156]. Moreover, we present some convergence results of the MCMC algorithm developed as well as some comments supporting the need for a delayed-rejection stage.

### 4.6.1 MCMC vs. BH-MLE

The purpose of this section is to compare and contrast MCMC techniques to the Bump Hunting Maximum Likelihood Estimation (BH-MLE) algorithm explained in Wallace *et al.* [156] by evaluating synthetic pulsed ladar data. The BH-MLE algorithm has two main stages: non-parametric bump hunting based on an analysis of the smoothed derivatives of the photon count histogram in scale space, and maximum likelihood estimation using Poisson statistics. Bump hunting is a non-parametric procedure [71] that provides an initial estimate of the number, amplitude and positions of the suspected returns. This uses Gaussian filters of progressively decreasing standard deviation to detect structure. Chaudhuri and Marron [30] have shown that there is a similarity between the statistical procedure of applying kernel density estimates and scale space filtering in computer vision. In their work, they derived SiZer (significant zero crossings of derivatives) maps, in which a zero crossing in a smoothed derivative at a certain scale was indicative of a mode, or peak, of a given width. However, in the approach of Wallace *et al.* [156], second derivative filters are also applied to the original data, and they detect peaks in the second, i.e. curvature, rather than zero crossings in the first derivative, i.e. slopes. This arises primarily from the desire to detect multiple returns when indicated only by multitangentiality and therefore,

they argued that this statistic gives a better measure of the presence or absence of a mode. Thus, the algorithm makes use of a threshold[1] to determine the presence or absence of a mode. Nevertheless, the curvature statistic presents some drawbacks since varying the curvature threshold will increase the chances of detecting the peak, but also will increase the probability of false detection in the scenario of multiple peaks. Once the initial estimate is obtained, they compute a maximum likelihood estimate (MLE) of the set of parameters of equation 4.4 that best explain the observed data.

To evaluate the MCMC algorithm and the BH-MLE algorithm we have performed simulations of peak detection for small signal-to-background ratios (SBR). Obviously for high SBR both methodologies yield very accurate results. Although the BH-MLE algorithm can be employed in situations where the number of peaks is unknown, in this section we restrict ourselves to the case where only one return is considered. The case in which multiple returns are considered is therefore postponed to later sections.



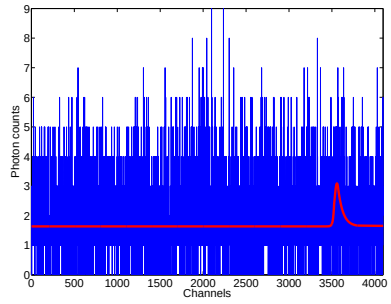

(a) (b)

**Figure 4.4:** (a) Simulated and reference signal used in the simulation (b) Reference signal (green), shape fitting of the reference signal (blue) and error incurred (red)

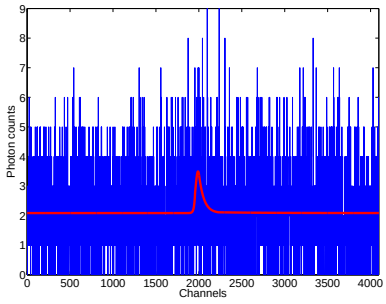
To simulate the return signal, we establish an initial fixed set of shape parameters obtained from a reference signal coming from a corner cube or a "good" external target (see figure 4.4(a)) to model the shape of the simulated response. The returned signal is then simulated using Poisson statistics as shown in figure 4.4(a) and then placed at

[1] In the following experiments the curvature threshold will be denoted simply as curvature to simplify notation

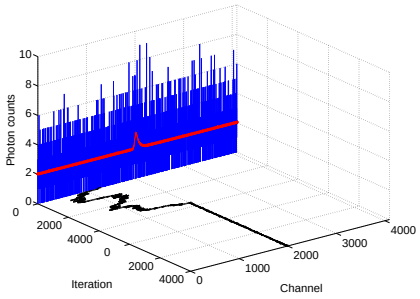
an arbitrary location, bin number 2000 and with an amplitude value of 1 photon count against a background of 2 photon counts. Figure 4.4(b) shows the instrumental signal (reference) obtained from a corner cube(green), the fit of this reference signal (in blue) using the four piecewise exponential model of equation 4.1 and the error incurred by using an 'operational' model (in red).



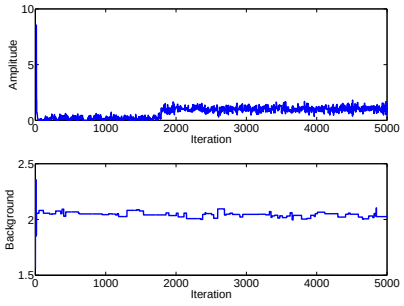

(a) Final estimate using BH-MLE (in red)




(b) Final estimate using MCMC (in red)




(c) Position trace plot




(d) Amplitude and Background trace plots

**Figure 4.5:** Results obtained using the BH-MLE and MCMC algorithms for the simulated data of figure 4.4

The synthetic data has been analysed with the model defined in section 4.3 for both methodologies. In the MCMC case, we set the parameter of the random walk proposal for the position parameter as $\sigma_1 = 25$. A priori, $t_0 \sim U[0, Number of channels$ showing that all the positions in the histogram are equally probable, $\beta \sim \Gamma(6, (\max(\boldsymbol{y})/2)/6)$, and $B \sim \Gamma(1.03, 10000)$. The parameters for the prior of the amplitude were chosen to allow small returns with similar height as the background. The parameters for the prior of the background were chosen to avoid background values very close to 0. Note that the experiments in this section do not use a delayed-rejection step in order to observe the behaviour of a standard MCMC algorithm. The MCMC algorithm has been run for

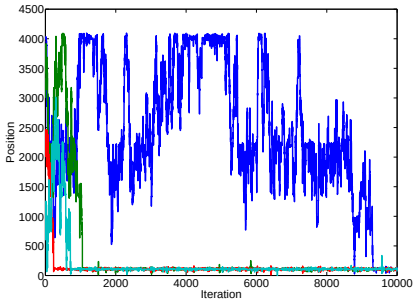
5000 iterations, with a burn-in period of 4000 iterations and random initial values. The BH-MLE algorithm has been used with different curvature thresholds from which two of them are key (those corresponding to threshold values of 3 and 4) since they imply the choice of the false return (see figure 4.5(a)) or the correct return (see figure 4.5(b)). The estimates obtained with the BH-MLE algorithm for both curvatures as well as those estimates obtained from MCMC marginal posterior means are shown in table 4.1. As it can be seen, for a curvature threshold value of 4 the BH-MLE algorithm detects a small false return. However, as we change the curvature threshold to 3 the correct return is detected although the estimates are not as good as those obtained with the MCMC algorithm. Figure 4.5(c) shows in red the final estimate of the MCMC algorithm and in blue the synthetic return. In this figure, the trace plot of the position parameter is also represented. The strong variations of the sequence of iterates generated correspond to the burn-in period of the chain. Figure 4.5(d) shows the trace plots for the amplitude and the background parameters. Note the poor mixing behaviour of the sequence of iterates of the background compared to that of the amplitude parameter. The burn-in period is also appreciable in the trace plot of the amplitude parameter.

**Table 4.1:** Illustrative results from simulations of figure 4.4

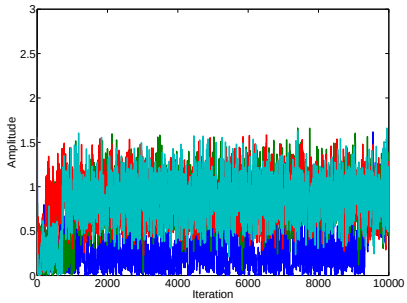
| | Position (Channel) | Amplitude (Photon counts) | Background (Photon Counts) |
|---|---|---|---|
| Actual Data | 2000 | 1 | 2 |
| BH-MLE (curvature = 4) | 3558.560 | 1.451 | 1.638 |
| BH-MLE Hunting (curvature = 3) | 1990.110 | 1.407 | 2.087 |
| MCMC (iterations = 5000) | 1996.920 | 1.042 | 2.043 |

Note that the BH-MLE algorithm is considered as a deterministic algorithm since for the same initial conditions we obtain the same results, however, the stochastic nature of the MCMC algorithm results in variations in the final estimate of the parameters. Repeatability and other performance measures are considered in more detail in the following sections.

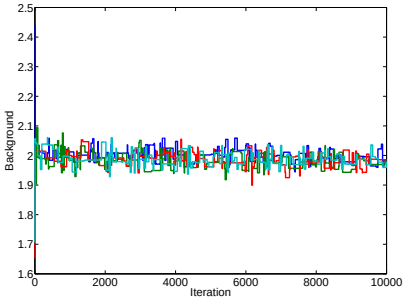
### 4.6.2 On the Convergence of the MCMC Algorithm



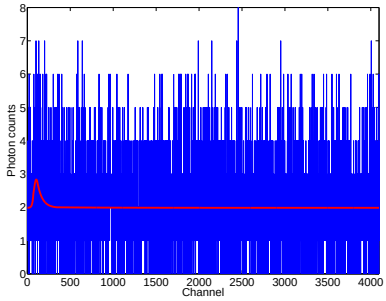

(a) Position trace plots




(b) Amplitude trace plots




(c) Background trace plots



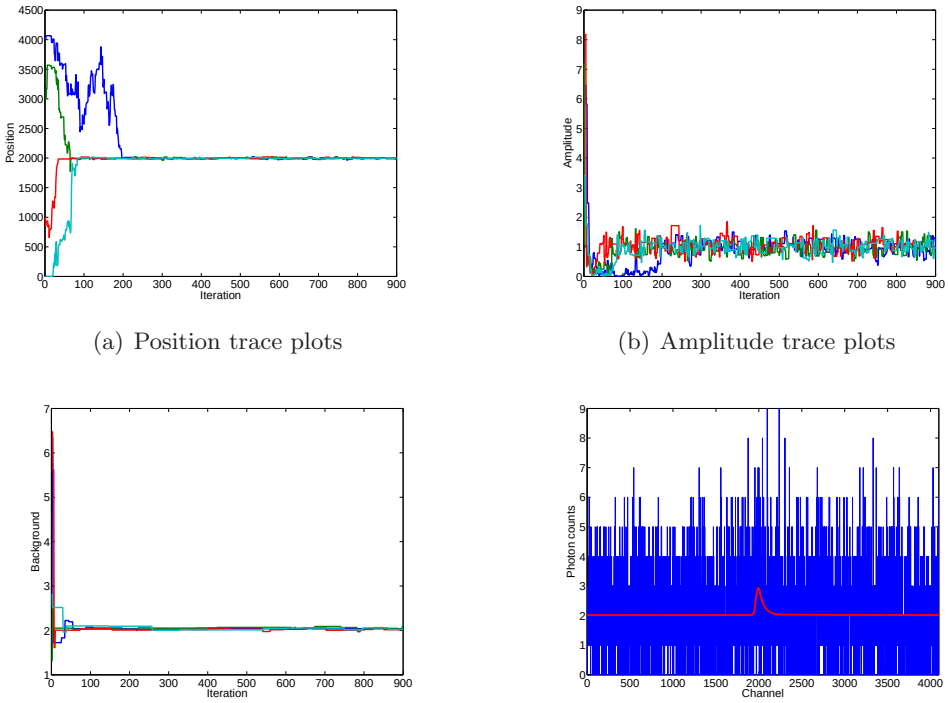

(d) Simulated return (in blue) and final estimate (red)

**Figure 4.6:** Results obtained using MCMC and Gelman and Rubin convergence diagnostic. For each parameter 4 chains were running in parallel.

Asymptotic convergence of the Markov chain generated is easy to prove in our case. Indeed, the invariance property of the chains holds since the detailed balance condition is satisfied. Aperiodicity and irreducibility are almost obvious for this situation since the likelihood never vanishes for any parameter values. The same results are valid for the multiple peak case. However, the assessment of convergence, i.e. to determine that the samples we are obtaining are representative of the underlying stationary distribution, is more difficult. The ergodic property of the Markov chain generated ensures that it will eventually cover all the target distribution, although to assess when this is going to happen is a hard issue. Slow convergence, even in a distribution that has a single mode, will produce poor estimates if we do not run the Markov chain for a sufficient time. Figure 4.6(a) shows an example of a trace plot (in blue) in which slow convergence can be observed. As explained in section 3.2.3.3, tools known as convergence diagnostics can be used to study

the behaviour of the Markov chain generated. In addition, to assess the convergence of the Markov chain we are interested in improving the mixing of the chain to allow better exploration of the posterior distribution and therefore a more efficient sample averaging through an improved convergence of the chain.

#### 4.6.2.1 Gelman and Rubin Convergence Diagnostic




(a) Position trace plots

(b) Amplitude trace plots

(c) Background trace plots

(d) Simulated return (in blue) and final estimate (red)

**Figure 4.7:** Results obtained using MCMC and Gelman and Rubin convergence diagnostic. In this experiment, 4 chains were running in parallel.

To monitor the convergence of the Markov chain, we use the method of Gelman and Rubin (see for instance [60]) based on the simulation of multiple chains in parallel and analysis of variances within the chains and between the chains in contrast to simple visual inspection. The different Markov chains are generated from different overdispersed starting points to ensure that all the major regions of the target distribution are explored in the simulations. The main idea is to form an overestimate and an underestimate of the variance of the target distribution, with the property that the estimates will be roughly equal at convergence but not before. The analysis of the variance is used to estimate the

factor, known as "estimated potential scale reduction", by which the scale parameter of the marginal posterior distribution of each variable might be reduced if the chain were run to infinity. Although best results are obtained for parameters whose marginal posterior densities are approximately normal, Gelman and Rubin's diagnostics works well in our synthetic data as the different time-series traces converge quickly.

Figures 4.6 and 4.7 display some trace plots for the different parameters, the simulated histograms and the final fit using MCMC. Figure 4.6(d) shows a simulated histogram (in blue) with a single return placed at an arbitrary location, bin number 100 and an amplitude value of 1 photon count against a background of 2 photon counts. The estimated return is represented in red. This data was analysed with the MCMC algorithm with a maximum number of 10000 iterations. The burn-in period was determined using the Gelman and Rubin algorithm. Figure 4.6(a) shows that convergence of one of the trace plots (in blue) is very slow suggesting that only after a burn-in of 9500 iterations the convergence to the stationary distribution is achieved. This value is consistent with the output given by the Gelman and Rubin algorithm. Similar conclusions are obtained looking at the trace plot associated with the amplitude value (figure 4.6(b)). On the other hand, figure 4.6(c) displays a very fast convergence of the background level.

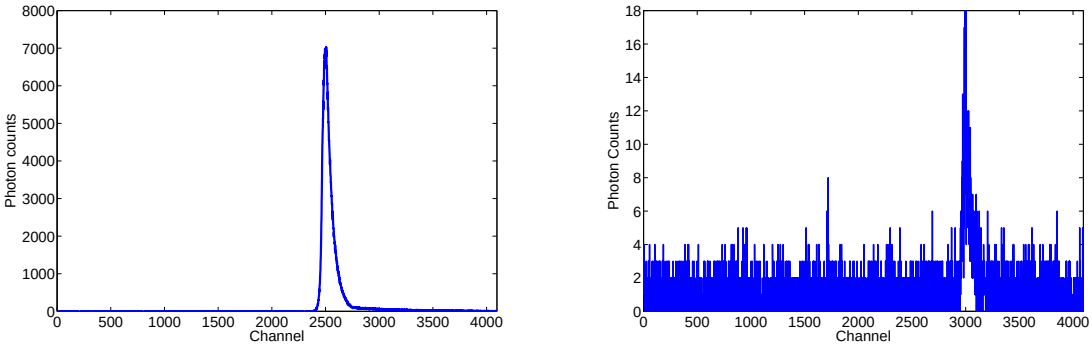
Figure 4.7 shows an example of simulated data in which the convergence is attained faster than in the previous case. Figure 4.7(d) shows the simulated histogram (in blue) with a single return placed at bin number 2000 with an amplitude value of 1 photon count against a background of 2 photon counts. In red is represented the final fit of the return using the estimates obtained from MCMC marginal posterior means. Figure 4.7(a) suggests that a burn-in of 200 iterations was sufficient to achieve convergence for the amplitude, position and background, and that samples $201 - 900$ from both chains may be assumed to arise from the marginal posterior distributions for each variable. Again the visual inspection is consistent with the results obtained with the Gelman and Rubin algorithm.

Note that other existing tools for assessing the convergence of a Markov chain such as that of Raftery and Lewis [35] could have been employed. However, methods based on a single long run can fail either because the simulations are moving very slowly or are stuck somewhere in the target distribution. As an example consider again the Markov

chain (in blue) in figure 4.6(a). This chain takes a long time to converge to the stationary distribution. If a single long run chain had been used, assessment of convergence for a single chain would have needed more iterations than a multiple-chains method. Due to the complexity of our model, each iteration is expensive in terms of computer time, and it is desirable to keep the simulation runs as short as possible.

#### 4.6.2.2 Delayed Rejection

In general, improved mixing can be achieved with a fine tuning of the proposal distribution. The rate of convergence will depend necessarily on the relationship between the proposal distribution and the target distribution. Note that even when the chain has converged, it may still mix slowly which implies that it has to be run longer to obtain accurate estimates from the sample path average. Slow mixing of the chain also implies high correlation between consecutive samples.




(a) (b)

**Figure 4.8:** Some data sets corresponding to real (a) and synthetic (b) data.

Ladar data are not exempt from this problem. Constructing a proposal which leads to rapid mixing is difficult since the domain of the parameters is large in the case of the position parameters, or even unbounded in the case of the amplitude and background parameters. Indeed, in the TCSPC data there is no "apparent" bound as the maximum number of photons that can be collected per temporal channel whereas the maximum level of intensity in the BIL case is limited by the CCD camera employed. These features suggest the use of proposal distributions with their domain in $\mathbb{R}^+$ that are capable of

proposing small and large values in a simulation run. Moreover, the number of temporal channels is usually large, *e.g.* of the order of 4096 for the TCSPC, and movement between regions separated by areas of low posterior support is difficult requiring proposals with high variance. However, if using a large spread proposal, once a region of higher probability is reached, exploring the surrounding area is difficult and unlikely to occur in a simulation run of limited length and a narrower proposal is therefore required.

To deal with these problems we propose to use a delayed rejection (DR) step which improves efficiency of the resulting MCMC estimators in a Metropolis-Hastings algorithm relative to Peskun's asymptotic variance ordering [110]. Two different stages have been proposed as explained in section 4.5. To analyse the performance of the DR algorithm we employ two sets of real (figure 4.8(a)) and synthetic data (figure 4.8(b)). Both data sets were analysed using a MCMC algorithm with a DR stage and without it.

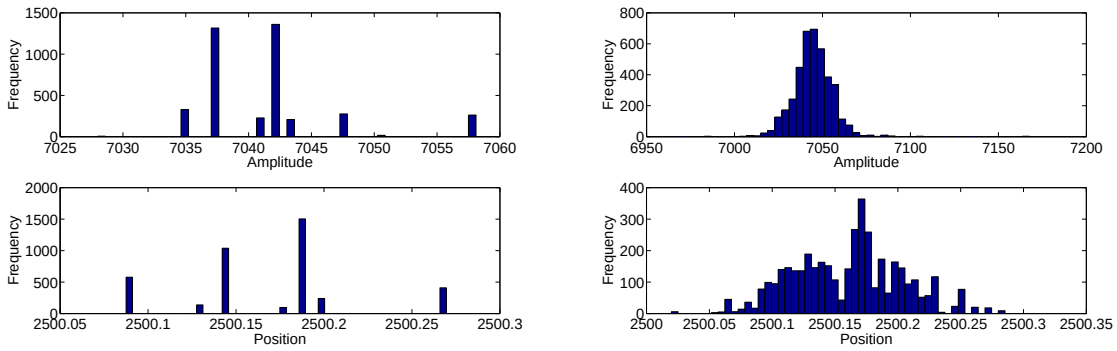
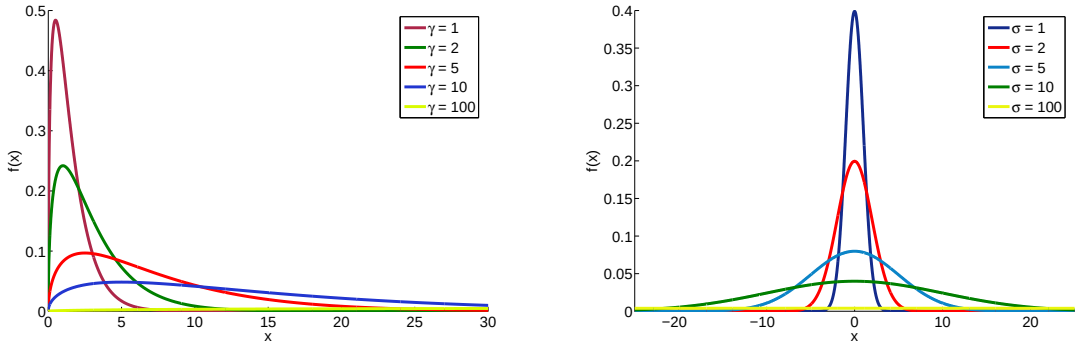



**Figure 4.9:** Histograms of the amplitude and position parameters with (b) and without (a) a delayed rejection step

The main difficulty of the first data set is to verify that the chain is not mixing slowly once it has achieved convergence . Figure 4.9(a) and figure 4.9(b) display the histograms of the results for the estimates of the position and the amplitude for a standard MCMC algorithm and one that includes DR respectively. The results correspond to 5000 iterations, following a burn-in period of 4000 iterations. The previously unspecified constants were set to $\sigma_1 = 1000$, $\sigma_2 = 10$, $\sigma_3 = 10$, $a = c = 1.03$, $b = 10000$ and $d = 1000$. It can be observed that for mixing, in the case using a standard MCMC algorithm, the Markov chain must perform

more iterations in order to calculate reliable statistics whereas the Markov chain generated using the DR algorithm has mixed satisfactorily. To explain the reason for this constrasting behaviour we have to observe carefully the proposal distributions used in the first stage which are common to both methodologies, albeit with different settings. Figure 4.10(a) shows several plots of the gamma distribution for a fixed shape parameter (value 1.5) and different values of the scale parameter (variable in our algorithm). Increasing the value of the scale parameter results in a more diffuse distribution. When this value is large enough, the gamma distribution has a behaviour similar to the uniform distribution giving similar probabilities to all locations. Proposing a value in a neighbourhood of the current value of the Markov chain becomes difficult and the chain must be run longer. On the other hand, the normal distribution used as proposal for the position parameter does not suffer from this problem as the scale parameter is fixed throughout the simulation. However, a bad mixing of the amplitude parameter and a 'large' value of the scale parameter will also harm the mixing of the Markov chain. Again, if the scale parameter of the Normal distribution is very large (see figure 4.10(b)), then exploring the neighbourhood of the current Markov chain estimate is more difficult as a result.




(a) Gamma distribution with shape parameter equal to 1.5

(b) Normal distribution with mean equal to 0

**Figure 4.10:** Examples of proposal distributions

Figure 4.8(b) shows a simulated histogram with a single return placed at an arbitrary location (bin number 3000) and an amplitude value of 10 photon counts against a background of 1 photon counts. To analyse these data we set the initial value of the Markov chain of the position parameter to location bin number 100. The objective of this experiment is to

show that a relatively small value of the scale parameter of the normal distribution will have repercussions on the convergence of the Markov chain. Figure 4.11 displays the trace plots for the different parameters when using a standard MCMC algorithm with scale parameter (standard deviation) equal to 10. As can be observed from the figure, the Markov chain has not converged to the stationary distribution since it needs longer runs to explore completely the state space. The trace plot corresponding to the background may indicate that the Markov chain has converged, however, it can be seen that its value is bigger than the true value. In some way, the background is trying to compensate the effects of not identifying the peak by increasing the likelihood. On the other hand, figure 4.12 displays the trace plots for the different parameters when using a MCMC algorithm with a delayed rejection step. The standard deviation of the normal distribution in the first stage was set to 1000 channels whereas the standard deviations of the second stage were set to 1 channel for the position proposal and 50 photon counts for the amplitude proposal. It can be observed that initially the first stage allows us to explore the state space very quickly till we find the peak. Once the peak is found, the values proposed in the first stage are persistently rejected and then the second stage is used proposing smaller values which allow us to explore the neighbourhood of the current Markov chain estimate. The good behaviour of the position parameter allows the Markov chain generated to converge quickly as well.

Finally, note that this methodology also fails in the case where no return is present and only background is represented in the histogram. The algorithm will proceed to fit very small false returns and will not converge to the actual stationary distribution. To deal with this problem, it is necessary to allow the algorithm to determine that no peak is present. To do that, it is necessary to use a RJMCMC methodology which allows an unknown number of peaks. However, the use of a delayed rejection algorithm is still important since the more generic proposals, *e.g.* random walk, are known to fail to explore large dimensions involving large sets of parameters.

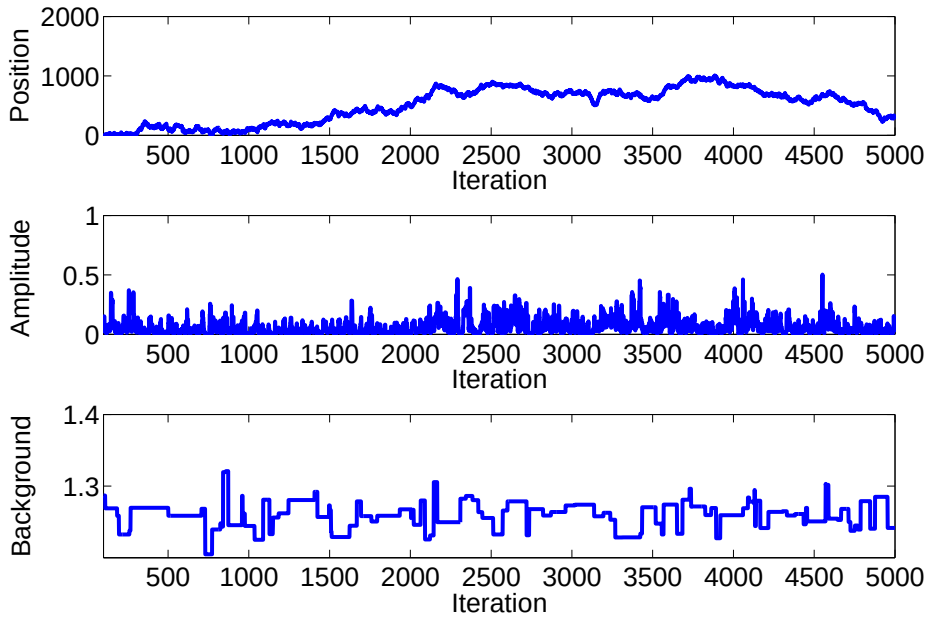



**Figure 4.11:** Trace plots for the different parameters when using a standard MCMC algorithm without a DR step

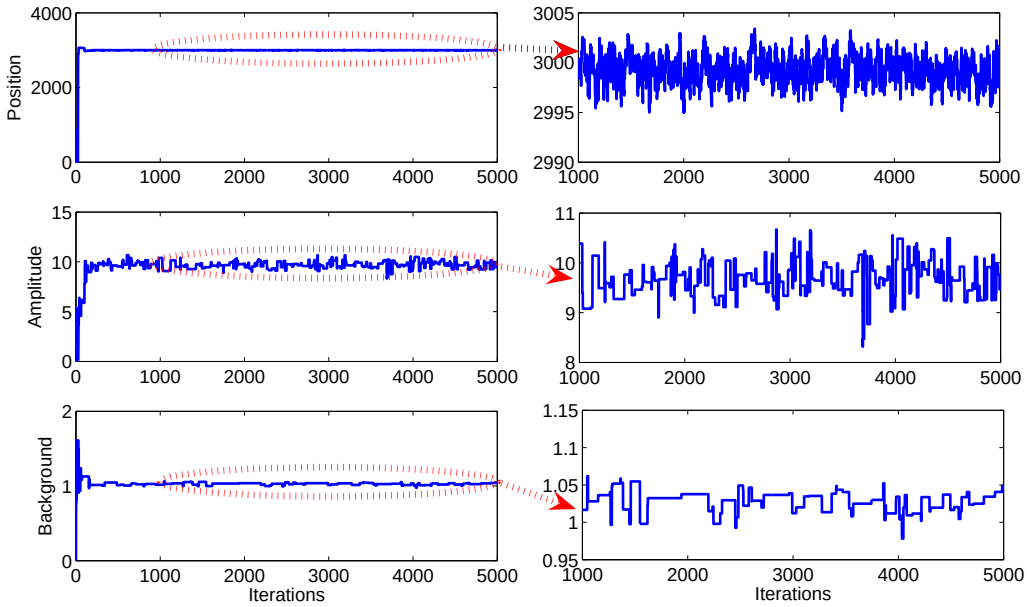


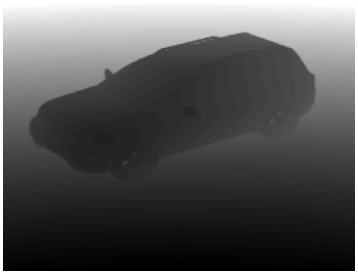

**Figure 4.12:** Trace plots for the different parameters when using a standard MCMC algorithm with a DR step. Expansions of the different graphs after burn-in period are also shown.

### 4.6.3 3D Ranging and Imaging

In this experiment, we considered the processing of synthetic and real TCSPC data as well as real range gated BIL data. The amplitude and position of the peaks of the synthetic data were simulated using intensity (figure 4.13(a)) and range (figure 4.13(b) information obtained with Cameosim$^{TM}$. Specifically, the object simulated was a Land Rover Freelander. The shape parameters were obtained from the fit of real data parameters and then Poisson noise was added to the signal. Figure 4.13(c) shows the intensity reconstruction of the Freelander whereas figure 4.13(d) shows a mesh reconstruction. Since this is a simulation, the ground truth is known and it is possible to assess the estimation error. Figures 4.13(e) and 4.13(f) display the error performed when estimating the intensity and position parameters respectively. The respective histograms for these errors are shown in figures 4.13(g) and 4.13(h). Note that the error in intensity is due to the strong Poisson noise used in the simulation, the biggest errors corresponding to areas with high-intensity values. Indeed, as the mean value parameter of the Poisson distribution increases, the distribution spreads out more and the fluctuations on the values are larger. On the other hand, the small errors made when estimating the position parameter are again due to the Poisson variations induced by the noise model.

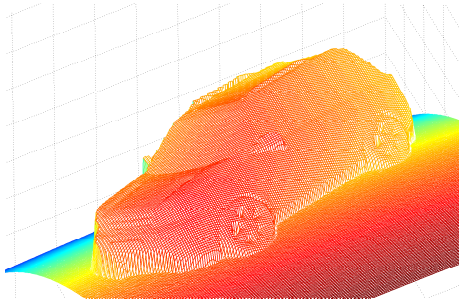
In the following experiment, we acquired real 3D images of a toy cat shown in figure 4.14(a) using the TCSPC system and a Land Rover shown in figure 4.14(e) using the BIL system and we have processed it using the MCMC algorithm considering only a single return with a delayed-rejection step. The results are obtained by marginal posterior averages of 1000 iterations, following a burn-in period of 900 iterations. The algorithm parameters used in the MCMC algorithm had the same value to those of section 4.6.2.2. Figure 4.14(b) shows a mesh reconstruction of the toy cat whereas figure 4.14(c) shows a depth profile of the previous image corresponding to row 50 where it can be seen that the algorithm captures the curvature of the object. Figure 4.14(d) shows the reflectance image obtained when estimating the amplitude parameters of equation 4.2. Figure 4.14(f) shows a zoom of the mesh reconstruction of the Land Rover.

(a) Simulated intensity

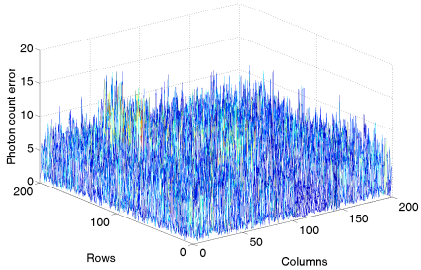


(b) Simulated depth

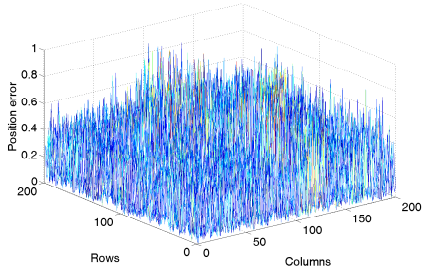
(c) Estimated intensity



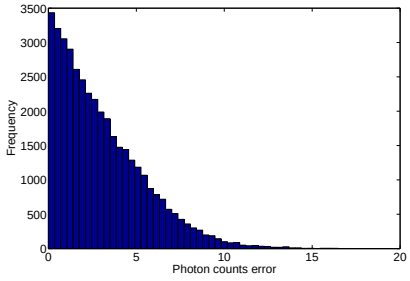
(d) Estimated depth



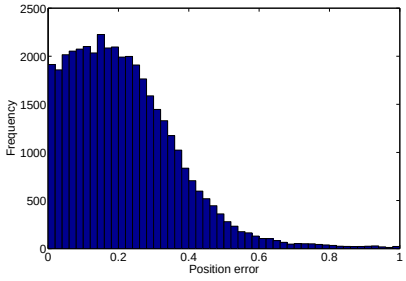

(e) Intensity error measured in photon counts




(f) Range error measured in bins ($1bin \simeq 6.1ps$)



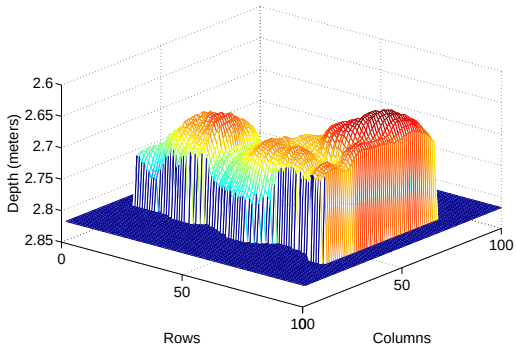

(g) Histogram of the intensity error (measured in photon counts)



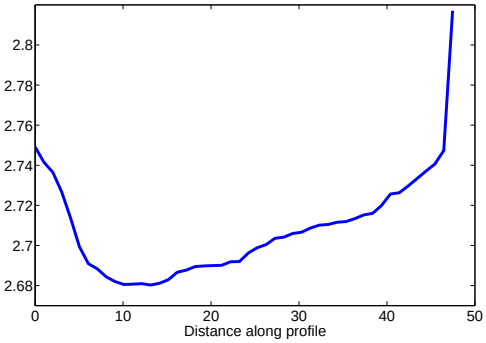

(h) Histogram of the range error measured in bins ($1bin \simeq 6.1ps$)

**Figure 4.13:** Simulated data of a Land Rover and estimates obtained using the MCMC algorithm

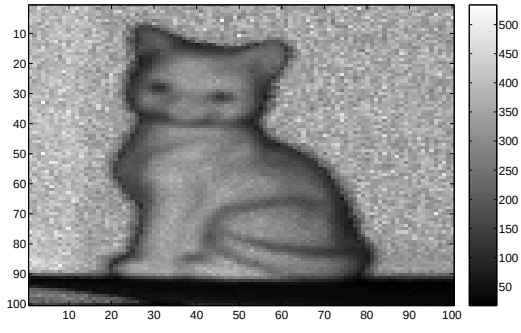
(a) Image of a toy cat




(b) Mesh of a toy cat

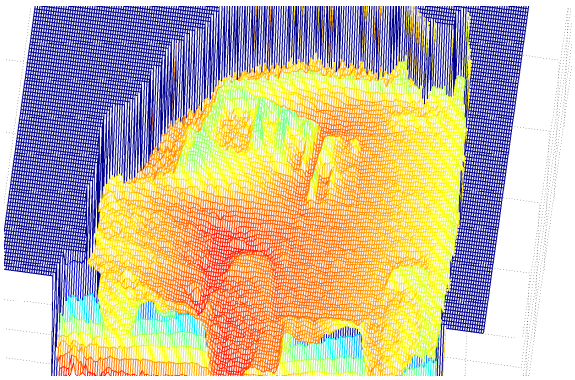



(c) Depth profile corresponding to row 50




(d) Reflectance image

(e) Image of a Land Rover



(f) Mesh of the Land Rover

**Figure 4.14:** Details from depth images acquired from a TCSPC system and BIL system

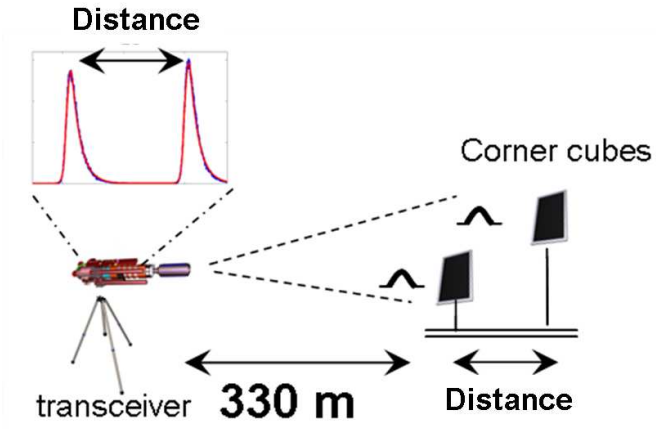



**Figure 4.15:** Scheme of the experiment of section 4.7.1

## 4.7 Experimental and Simulated Results; Multiple Returns.

The effectiveness of the RJMCMC algorithm in the processing of TOF signals is demonstrated for some synthetic and real data sets. In general, to process the data we run the algorithm a determined number of sweeps to explore the full posterior distribution and we discard all the samples corresponding to the burn-in. Then, we infer the number of peaks as that corresponding to the highest marginal posterior probability, $\hat{k} = argmax\ p(k|\boldsymbol{y})$. Once we determine the number of peaks, we extract estimates of the parameters from $p(\phi|k, \boldsymbol{y})$ by setting $p(\phi|k = \hat{k}, \boldsymbol{y})$. The values of the parameters are estimated as the mean values of the samples that correspond to such a parameter subspace. The data sets, along with the implementation parameters, are described for each experiment and a series of algorithm performance measurements are presented. Finally, the effectiveness of the RJMCMC algorithm for 3D ranging and imaging is demonstrated.

### 4.7.1 Algorithm Performance Measurements

The following experiment was performed to assess the ability of our algorithm to distinguish two surfaces separated by a finite distance. In other words, this experiment will allow us to investigate error due to short RJMCMC runs. For that reason, two retro-reflecting corner-cubes were placed at a distance of $330m$ from the TCSPC receiver described in Section 2.3 and the distance between these surfaces was varied from $1.7cm$ to $71.2cm$ as shown in figure 4.15. For a corner cube, all beams, independent of incident direction, are

reflected back in the original direction so the behavior is that of a perfect reflecting surface. The different data sets were collected and analysed with the RJMCMC algorithm. To assess the performance of the algorithm the following procedure was performed:

1. Run 100 independent Markov chains for a specific data set using RJMCMC.
2. For each Markov chain infer the number of peaks, positions, amplitudes and background.
3. Estimate the distance between the detected peaks.
4. Once the distances for each Markov chain are computed, calculate summary statistics (mean, median, root mean square error (RMSE), ...)

The support of the prior for the number of peaks is $[0, 1, 2]$. For each of the data sets, we report the results of 1000 sweeps, following a burn-in period of 4000 sweeps. Previous experiments have shown these values to be reliable. We set the previously unspecified constants as: $\sigma_1 = 1000$, $\sigma_2 = 10$, $\sigma_3 = 50$, $a = 6$, $b = (\max(\boldsymbol{y})/2)/6$, $c = 1.03$ and $d = 10000$. The values of $a$ and $b$ have been chosen to place low prior belief that peak amplitude will be small whereas the values of $c$ and $d$ have been chosen to represent vague prior knowledge of the background. For all the data set, the marginal posterior distribution of the number of peaks support strongly a number of peaks equal to 2. A scatter plot of the results obtained is shown in figure 4.16 and we have assessed the uncertainty of these measures by calculating the summary statistics which are illustrated in Table 4.2. We use the standard deviation to measure the range precision, the root mean square error (RMSE) as a measure of the error relative to the actual distance and the range of the samples to measure the dispersion of the observations. The quartiles $Q_1$ and $Q_3$ are also calculated in order to assess where 50% of the observations lie and therefore to show the presence of any possible outliers since some of the data sets present a distribution of the distance which is not symmetric around the mean value.

From figure 4.16 and table 4.2, it can be observed that the values with more dispersion are the data sets corresponding to distances of $1.7cm$ and $3.2cm$. These "large" values compared to those obtained with the other data sets are due to the time response of the

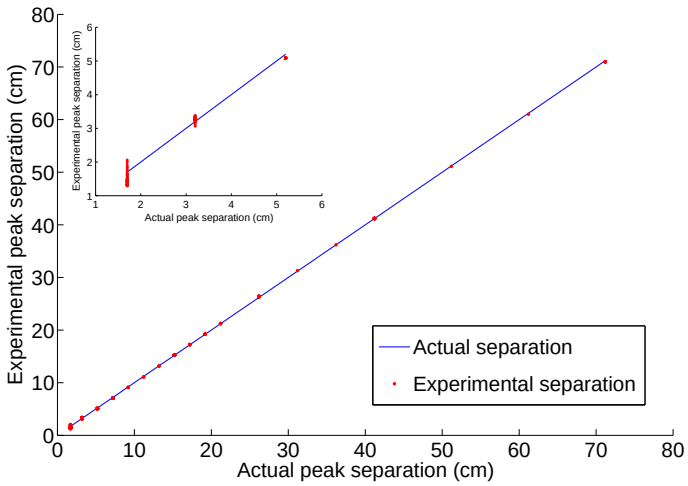



**Figure 4.16:** Calculated separation using RJMCMC vs. actual separation, the blue line shows the case of perfect agreement. The insert shows an expansion of the graph at the lower surface separations.

sensor which decreases with the range resolution of the system. Indeed, as the separation decreases, the peaks representing the position of the surfaces begin to merge and therefore it becomes increasingly difficult to calculate the separation of the peaks. Statistically, this difficulty is expressed as an increase in the range of the samples and the standard deviation in accordance to Table 4.2. From the 25% and 75% percentile values, it is possible to assess that the middle 50% of the values of the data set corresponding to distances of $1.7cm$ are allocated between the values 1.3376 and 1.4918. Clearly, these results underestimate the true distance of $1.7cm$. As we increase the distance between the surfaces, the results obtained are more consistent with the expected values. However, the results still display a bias with respect to the real value of the distance. This is due to two main causes. On the one hand the accuracy of the ruled gauge is limited to $1mm$. On the other hand the assumed peak shape is not tuned to the real peak shape. Nevertheless the statistics calculated are quite stable and consistent even using relatively short data runs.

To fully explore the advantages of our Bayesian approach we compare the RJMCMC algorithm developed with a basic centroid algorithm and the BH-MLE algorithm. These two algorithms are in essence deterministic in contrast to the stochastic nature of the RJMCMC algorithm. We use the mean value of the generated distances with RJMCMC as a bench-mark against which we compare the results obtained with centroid and BH-

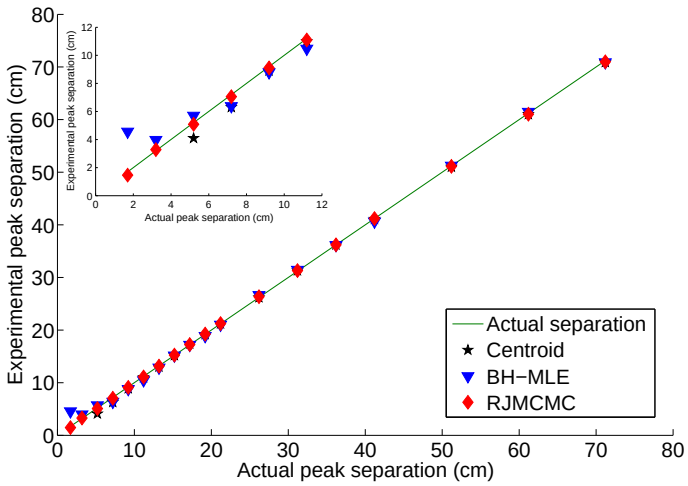
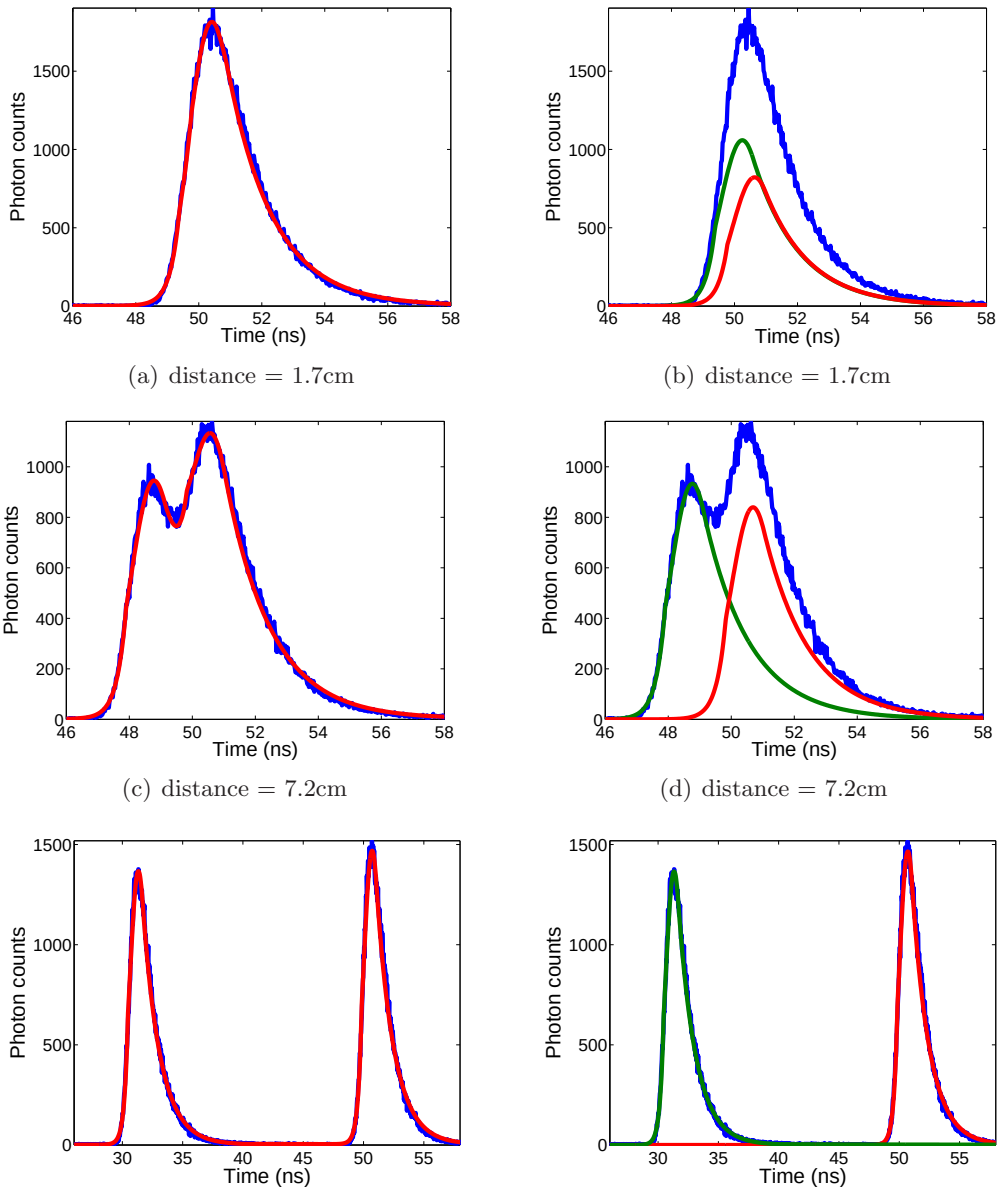



**Figure 4.17:** Calculated separation using different methods vs. actual separation, the green line shows the case of perfect agreement. The insert shows an expansion of the graph at the lower surface separations.

MLE. From figure 4.17 it can be observed that the centroid method is able to calculate with reasonable accuracy surface separations down to 5.2cm under the assumption that exactly two peaks are present. However, below this value this algorithm fails to detect two distinct surfaces and is not able to determine the number of peaks. On the other hand, the BH-MLE algorithm is able to calculate smaller separations between surfaces but fails to reasonably estimate the distance between peaks.

Figure 4.18 represents the number of detected photons (vertical axis) versus time (horizontal axis). Each recording on the vertical axis of these figures represents a single photon event, i.e. the first received photon from the target in response to a burst of laser energy in the form of a pulse. The number of photons collected depends on the laser repetition rate, the acquisition time and the properties of the optical system and target surface. Figure 4.18 (a), (c) and (d) show some examples of the final fit from RJMCMC estimation corresponding to distances of $1.7cm$, $7.2cm$ and $71.2cm$ whereas figures 4.18 (b), (d) and (e) show the different returns detected. Examples of return fitting using the BH-MLE algorithm can be found in [156].

RJMCMC algorithms are considered CPU-time consuming compared to simpler methods such as the centroid or BH-MLE. However, the complexity of these methods can be greatly reduced if we have a-priori information about the signal context. Further, the estimates




(a) distance = 1.7cm

(b) distance = 1.7cm

(c) distance = 7.2cm

(d) distance = 7.2cm

(e) distance = 71.2cm

(f) distance = 71.2cm

**Figure 4.18:** Some results of the experiment of section 4.7.1. Histogram of photon counts of the real data (blue) and final fit from RJMCMC estimation (red) in (a), (c) and (e). In figure (b), (d) and (f) the first return detected is represented in red whereas the green color represents the second return detected

**Table 4.2:** Summary statistics for the different data sets of the experiment of section 4.7.1

| Actual distance (cm) | Quartiles (cm) | | | | Mean (cm) | Standard deviation (cm) | RMSE (cm) | Range of samples (cm) |
|---|---|---|---|---|---|---|---|---|
| | 25% | 50% | 75% | 97.5% | | | | |
| 1.7 | 1.338 | 1.418 | 1.492 | 1.976 | 1.462 | 0.169 | 0.029 | 0.770 |
| 3.2 | 3.255 | 3.290 | 3.325 | 3.359 | 3.281 | 0.059 | 0.010 | 0.316 |
| 5.2 | 5.082 | 5.086 | 5.089 | 5.099 | 5.086 | 0.006 | 0.011 | 0.033 |
| 7.2 | 7.051 | 7.053 | 7.055 | 7.058 | 7.053 | 0.003 | 0.015 | 0.015 |
| 9.2 | 9.106 | 9.109 | 9.110 | 9.113 | 9.108 | 0.003 | 0.009 | 0.015 |
| 11.2 | 11.090 | 11.092 | 11.094 | 11.098 | 11.092 | 0.003 | 0.011 | 0.014 |
| 13.2 | 13.153 | 13.155 | 13.157 | 13.161 | 13.155 | 0.003 | 0.005 | 0.014 |
| 15.2 | 15.254 | 15.256 | 15.257 | 15.261 | 15.255 | 0.003 | 0.006 | 0.014 |
| 17.2 | 17.240 | 17.242 | 17.243 | 17.246 | 17.242 | 0.003 | 0.004 | 0.011 |
| 19.2 | 19.238 | 19.239 | 19.241 | 19.244 | 19.239 | 0.003 | 0.004 | 0.014 |
| 21.2 | 21.235 | 21.237 | 21.239 | 21.243 | 21.237 | 0.003 | 0.004 | 0.013 |
| 26.2 | 26.353 | 26.354 | 26.356 | 26.359 | 26.355 | 0.002 | 0.016 | 0.011 |
| 31.2 | 31.305 | 31.306 | 31.308 | 31.311 | 31.306 | 0.002 | 0.011 | 0.011 |
| 36.2 | 36.203 | 36.204 | 36.206 | 36.209 | 36.204 | 0.003 | 0.001 | 0.012 |
| 41.2 | 41.198 | 41.199 | 41.201 | 41.204 | 41.200 | 0.002 | 0.000 | 0.011 |
| 51.2 | 51.110 | 51.112 | 51.113 | 51.117 | 51.112 | 0.002 | 0.009 | 0.011 |
| 61.2 | 61.017 | 61.018 | 61.020 | 61.024 | 61.018 | 0.003 | 0.018 | 0.012 |
| 71.2 | 70.990 | 70.991 | 70.993 | 70.998 | 70.991 | 0.003 | 0.021 | 0.014 |

obtained with these algorithms are more accurate in both variable and fixed dimension as can be seen from the results presented here. Finally, RJMCMC methods naturally allow us to represent the uncertainties inherent in any inference from the data as shown in table 4.2.

### 4.7.2 Analysis of Multiple Returns

The purpose of this experiment is to assess the ability of our algorithm to detect multiple returns and much smaller return amplitudes in comparison with the background level. Again, the results obtained are compared to those obtained with the BH-MLE algorithm. Two different data sets were considered; one synthetic and the other real. The synthetic data provides ground truth to determine whether or not the algorithm is able to infer the exact number of peaks. On the other hand, the real data provides situations where the number of returns is unknown and the assumption of a unique return is not valid.

Figure 4.19 shows a simulation of 5 returns using a real target signature as a model placed at arbitrary locations and real data obtained from a complex target with multiple returns due to multiple and distributed or semi-transparent surfaces from an optical lens assembly. For the synthetic data, we report the results of 1000 sweeps, following a burn-in period of 4000 sweeps whereas for the real data the results correspond to 10000 sweeps, following a burn-in period of 9000 sweeps. We set the previously unspecified constants as: $\sigma_1 = 1000$, $\sigma_2 = 10$, $\sigma_3 = 50$, $a = 6$, $b = (\max(\boldsymbol{y})/2)/6$, $c = 1.03$ and $d = 10000$. The values of $a$, $b$, $c$ and $d$ have been chosen to make the prior on the amplitude and the background as vague as possible. The BH-MLE procedure was run for different values of the curvature statistic parameter even though this increases the probability of false detection of the algorithm. The maximum number of peaks was set as 10 for both methodologies in the case of the synthetic data and 20 for the real data case. For each data set, we compare the model with highest marginal posterior probability, $p(k|y)$, obtained with the RJMCMC algorithm to the best models chosen using AIC and BIC. Model selection using AIC or BIC works by calculating AIC/BIC scores for each candidate model and selecting the model with the lowest AIC/BIC as best. For us, candidate models are those models which follow the structure of equation 4.1 but that differ in the number of peaks (and possibly on the parameter values). The number of peaks is considered as the model indicator.

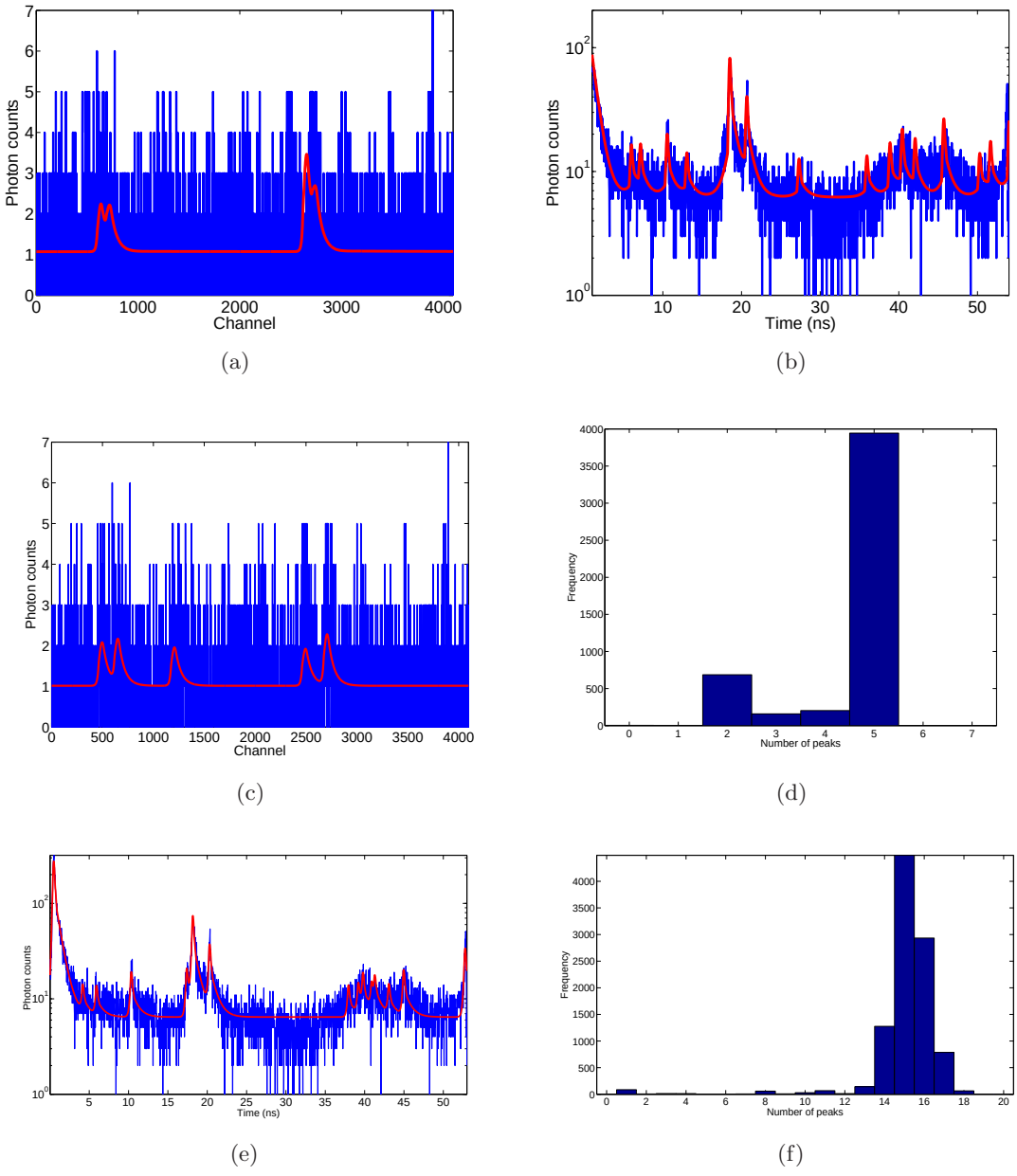
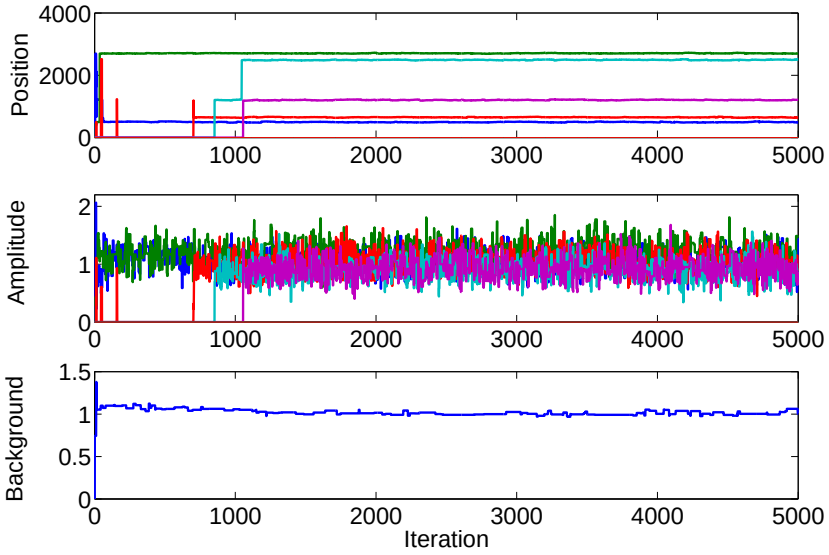


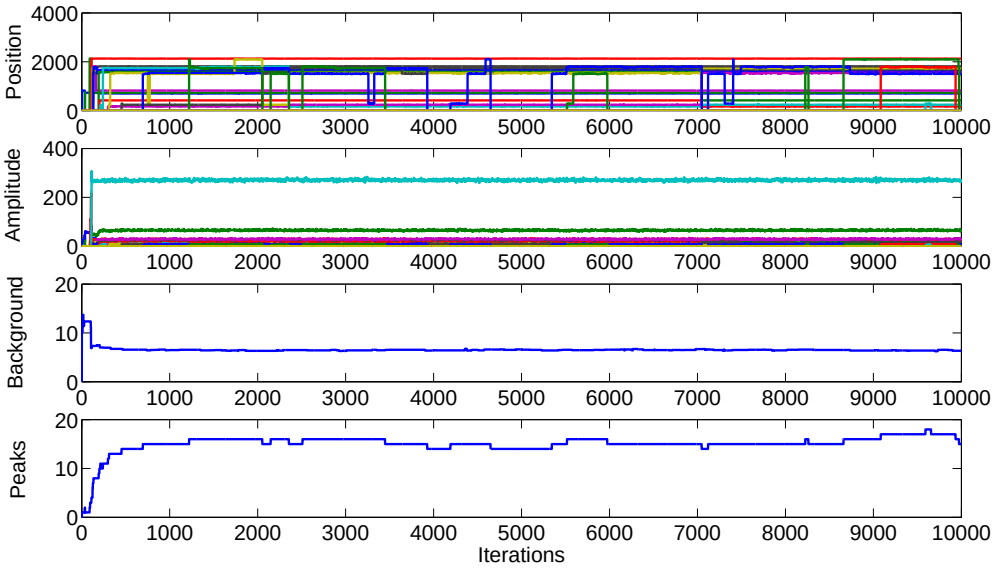

**Figure 4.19:** Analysis of TCSPC data. (a) Simulated histogram of photon counts (blue) and final fit from BH-MLE estimation (red) (b) Histogram of photon counts obtained with a TCSPC system (blue) and final fit from BH-MLE estimation (red) (c) Same histogram as that of (a) in blue and final fit from RJMCMC estimation (red) (d) Marginal posterior distribution of the number of peaks of the data of (c) obtained with RJMCMC (e) Same histogram as that of (b) in blue and final fit from RJMCMC estimation (red) (f) Marginal posterior distribution of the number of peaks of the data of (e) obtained with RJMCMC

The marginal posterior distributions of the number of peaks obtained with RJMCMC are shown in figure 4.19(d) and 4.19(f) for the synthetic and real data respectively. For the synthetic data set, the different methods agree that the best candidate model is that given by $k = 5$. Table 4.3 shows the final result of the BH-MLE for different values of the curvature statistic and the RJMCMC algorithm. As can be observed from the table, "large" values of the curvature statistic lead to under-estimation of the number of peaks present in the histogram whereas small values over-estimate the number of peaks. A curvature statistic of 3 correctly estimates the number peaks. However, most of them are false returns, i.e. do not correspond to actual peaks in the signal. Allowing a curvature statistic of 0 and incrementing the maximum number of returns to 20 will eventually find the 5 actual peaks but we shall find an extra 15 false returns. As a consequence, this methodology is quite sensitive to choice of the curvature statistics. This shows that even when we do not have any a-priori knowledge of the dimension of the parameter space, RJMCMC is able to infer the exact number of peaks, their amplitudes and their positions. The small variations are due to the existence of strong noise in the original, simulated photon count histogram.

Figures 4.20(a) and 4.20(b) show the different trace plots for the synthetic and real data respectively. In the synthetic data case, no label switching phenomenon is observable as opposed to the real data case. Indeed, a closer look at the graph of the generated trace plots displayed in figure 4.20 for the position parameter shows permutations of some of the trace plots. This causes severe problems when attempting to perform inference regarding the individual components of the mixture. In our case, this problem occurs when a change of dimension is performed and very small returns are processed. Taking mean values of the marginal posterior distributions to infer the parameter values may cause a bias on the genuine value of the parameter. This problem still remains even taking values after a long enough burn-in period. To solve this problem, we monitor the transitions of the Markov chain in which a change of dimension has actually occurred. Once the number of peaks is inferred using $\hat{k} = argmax\ p(k|y)$, we choose the largest sequence of consecutive iterates for which $\hat{k} = argmax\ p(k|y)$. Since no label-switching occurs for such a subspace of the space of solutions, the estimates obtained are not biased and represent accurate solutions.




(a) Synthetic data




(b) Real data

**Figure 4.20:** Trace plots for the different parameters using RJMCMC.

**Table 4.3:** Illustrative results from simulations of figure 4.19(c)

| | Actual values | | BH-MLE (Curvature = 5) | | BH-MLE (Curvature = 4) | | BH-MLE (Curvature = 3) | | BH-MLE (Curvature = 2) | | RJMCMC (5000 iterations) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Position | Amplitude | Position | Amplitude | Position | Amplitude | Position | Amplitude | Position | Amplitude | Position | Amplitude |
| 1st peak | 500 | 1 | 644.608 | 1.181 | 642.997 | 1.171 | 635.621 | 1.171 | 496.401 | 1.239 | 496.326 | 1.070 |
| 2nd peak | 650 | 1 | - | - | 2710.011 | 1.267 | 725.000 | 0.851 | 649.086 | 1.155 | 652.030 | 1.071 |
| 3rd peak | 1200 | 1 | - | - | - | - | 2656.000 | 1.267 | 808.000 | 0.703 | 1207.367 | 0.944 |
| 4th peak | 2500 | 1 | - | - | - | - | 2658.000 | 1.123 | 899.000 | 0.569 | 2494.939 | 0.909 |
| 5th peak | 2700 | 1 | - | - | - | - | 2748.000 | 1.015 | 1199.145 | 0.948 | 2707.642 | 1.239 |
| 6th peak | - | - | - | - | - | - | - | - | 1336.515 | 0.783 | - | - |
| 7th peak | - | - | - | - | - | - | - | - | 2064.338 | 0.678 | - | - |
| 8th peak | - | - | - | - | - | - | - | - | 2284.000 | 0.573 | - | - |
| 9th peak | - | - | - | - | - | - | - | - | 2493.186 | 1.085 | - | - |
| 10th peak | - | - | - | - | - | - | - | - | 2707.068 | 1.420 | - | - |
| B | 1 | | 1.056 | | 1.090 | | 1.072 | | 0.902 | | 1.014 | |

**Table 4.4:** Values of AIC, BIC and marginal posterior distribution, $p(k|\boldsymbol{y})$, for the different models

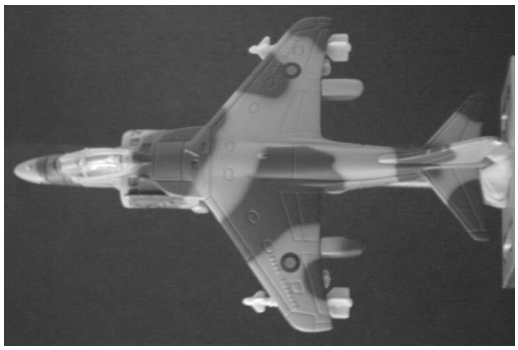
| Model | Number of parameters | AIC | BIC | $p(k\|\boldsymbol{y})$ |
|---|---|---|---|---|
| 0 | 1 | 21733.585 | 21750.558 | 0.0003 |
| 1 | 3 | 19202.808 | 19219.811 | 0.0086 |
| 2 | 5 | 18838.609 | 18866.947 | 0.0007 |
| 3 | 7 | 15424.315 | 15463.989 | 0.0014 |
| 4 | 9 | 13474.803 | 13525.812 | 0.0013 |
| 5 | 11 | 13307.255 | 13369.599 | 0.0001 |
| 6 | 13 | 13214.432 | 13288.111 | 0.0003 |
| 7 | 15 | 13318.816 | 13403.830 | 0.0007 |
| 8 | 17 | 13541.367 | 13637.717 | 0.0059 |
| 9 | 19 | 11805.894 | 11913.579 | 0.0008 |
| 10 | 21 | 11488.998 | 11608.018 | 0.0029 |
| 11 | 23 | 11412.920 | 11543.276 | 0.0068 |
| 12 | 25 | 11361.640 | 11503.331 | 0.0013 |
| 13 | 27 | 11286.296 | 11439.322 | 0.0142 |
| 14 | 29 | 11230.117 | 11394.478 | 0.1276 |
| 15 | 31 | 11284.399 | 11460.095 | 0.4488 |
| 16 | 33 | 11276.730 | 11463.762 | 0.2934 |
| 17 | 35 | 11303.961 | 11502.328 | 0.0786 |
| 18 | 37 | 11310.582 | 11520.284 | 0.0064 |

We have also calculated the AIC and BIC scores for the different candidate models using the RJMCMC estimates. Whereas in the synthetic data the estimated number of peaks is identical for the different methods, in the real data case, $\hat{k} = argmax\ p(k|y)$ occurs at $\hat{k} = 15$, the BIC criterion also supports this argument whilst the AIC criterion favours more complicated models in which $\hat{k} = 16$ as can be seen in table 4.3.

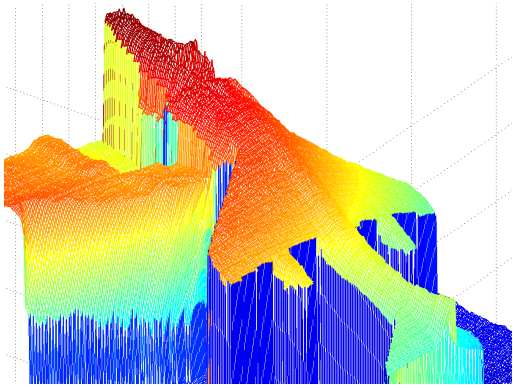
In other situations, the models proposed by AIC and BIC are not supported by the posterior distribution obtained by RJMCMC. This can be explained by Lindley's Paradox [92]. According to Lindley, the classical test can reject an hypothesis with an arbitrary small p-value (frequentist approach), while at the same time the evidence can convince us that the hypothesis is almost certainly true (Bayesian approach). In other words, it can happen that the log-likelihood increases faster than the number of parameters and therefore, the penalty against overfitting is not large enough.

### 4.7.3 3D Ranging and Imaging

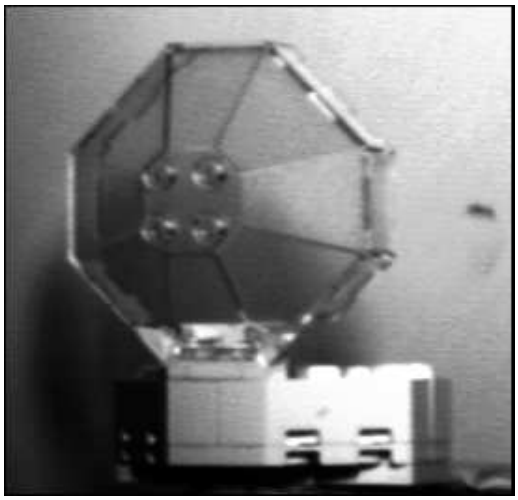
In this experiment, we acquired 3D images of a number of test objects of varying size at different distances and we applied the RJMCMC method to both TCSPC LaDAR data and range-gated BIL data. Figure 4.21 shows a series of examples from the TCSPC and



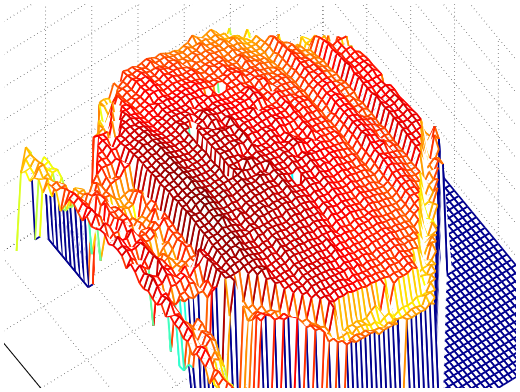
(a) Image of a toy fighter



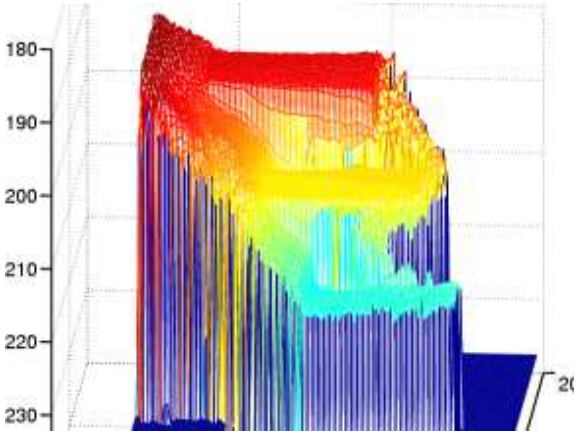
(b) Mesh of the toy fighter (distance$\sim 3.5m$)



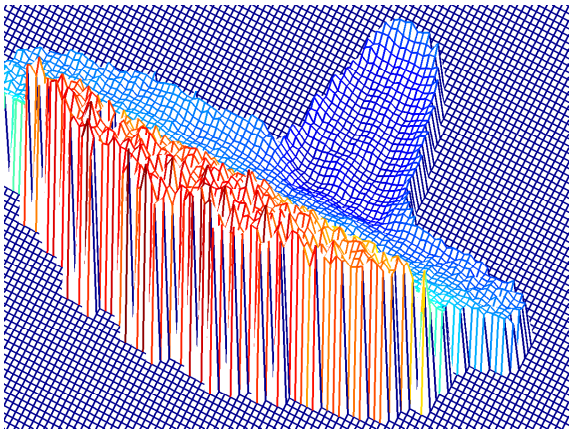
(c) Image of a Lego piece



(d) Mesh of the Lego piece (distance$\sim 2m$)




(e) Mesh of three planar targets (distance$\sim 200m$)



(f) Mesh of a trig point(distance$\sim 6.6km$)

**Figure 4.21:** Details from depth images acquired from a TCSPC system ($1bin \simeq 2.44ps$) ((b) and (d)) and BIL system ((e)and (f))

BIL data. Figure 4.21(a) and 4.21(b) show respectively a toy fighter (more specifically a Harrier) and its corresponding depth image. The toy fighter is mounted on a clamp that is also visible in the depth image. The data were obtained using a TCSPC scanned system. Figure 4.24(a) shows an example of a histogram return corresponding to one pixel of the image. Both, the return from the target and the reference used are shown in this figure.

Figures 4.21(c) and 4.21(a) show an example of a transparent object scanned using the TCSPC system. The object is a Lego piece with a transparent polygonal top, which is the surface of interest. The problem in this case is specularity more than transparency, since little of the reflected light may be directed back towards the detector. Figure 4.24(b) shows an example of a histogram return corresponding to one pixel of the image. The reference is centred approximately at channel 2700 whereas the return is about channel 1000. This return comes from one of the off-specular side faces, and has a maximum count of about ten photons.

Figure 4.21(e) shows a meshed reconstruction of the BIL image of the three planar targets that gives some idea of the depth variation across the targets. The peaks in the depth data adjacent to the planes are caused by ground reflections before or beyond the planar targets. The mean depths of the three planar targets were measured (by the laser) as $192.05m$, $204.00m$ and $215.5m$ respectively, in comparison with a ruled separation of approximately 12m between the targets. Figure 4.23 shows the temporal intensity response of a single pixel and an example of a single frame of a the BIL image acquired of the three planar targets. Figure 4.24(c) shows a multiple return example found in this data and its final fitting. This multiple return is possibly due to non-uniform reflectivity and/or ground reflections before or beyond the planar targets. The intensity values are plotted on a log scale for convenience, as the divergent beam is non-uniform, causing fall-off in intensity at the edges of the image and making a linear intensity image more difficult to interpret.

Figure 4.21(f) shows a meshed reconstruction from the BIL image sequence of a UK Ordinance Triangulation Point (trig point) shown in figure 4.22(a), located on top of a hill at several kilometres from the sensor. Reflections are also visible from the foreground below the target. Figure 4.22(b) shows a single frame BIL image of the trig point whereas Figure 4.24(d) shows the temporal intensity response of a single pixel when scanned through

(a) (b)

**Figure 4.22:** (a) UK Ordinance Triangulation Point (trig point) (b) Single frame BIL image of a trig point (long range)

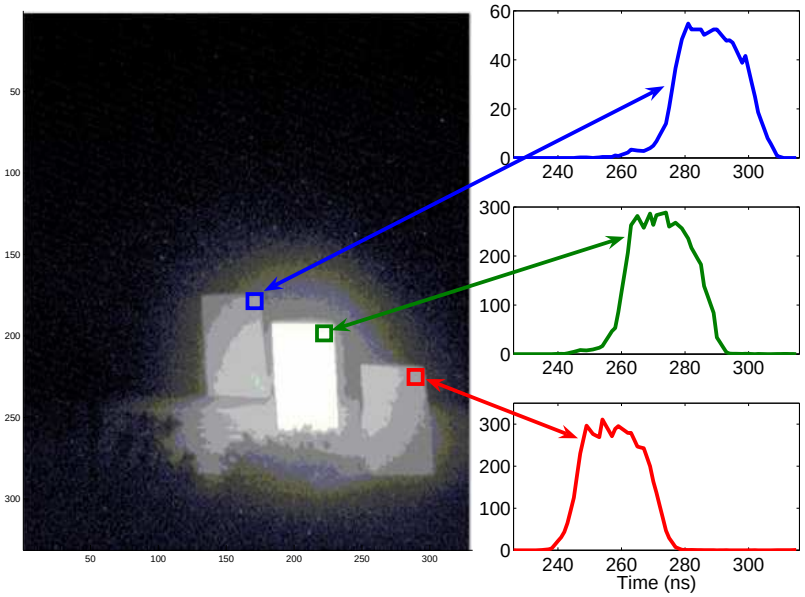
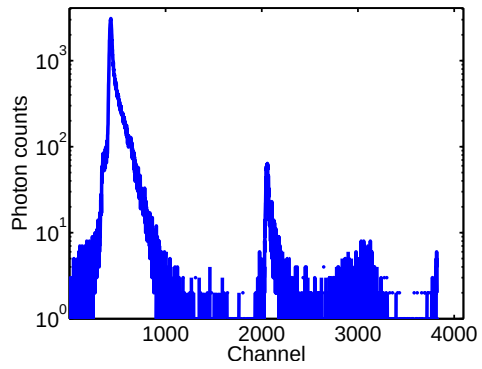
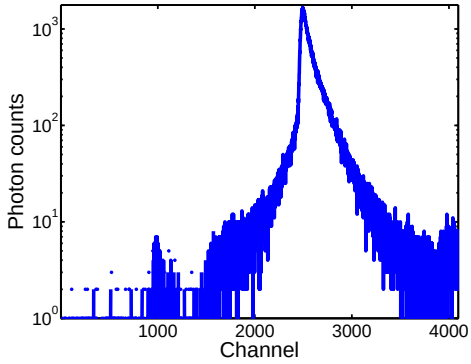


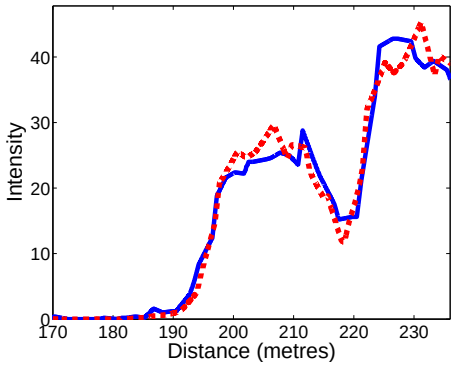

**Figure 4.23:** Log-intensity, single frame BIL image with stand-off distance centred on the middle of three uniform, matt planar targets, situated approximately 12m apart. The different intensity profiles show the distance variation along the different targets



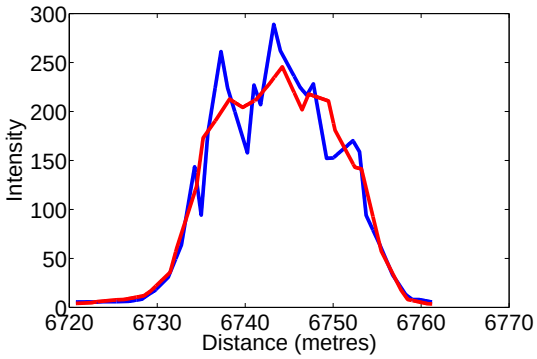

(a) Reference signal(left) and target return(right)




(b) Reference signal (right) and target return(left)




(c)




(d)

**Figure 4.24:** (a) Histogram of photon counts (vertical axis) against time (measured in 2.44*ps* channels, horizontal axis) corresponding to one pixel of the toy fighter data (b) Histogram of photon counts (vertical axis) against time (measured in 2.44*ps* channels, horizontal axis) corresponding to one pixel of the Lego data (c) Histogram of intensities (blue) of one pixel of three planar targets and final fit from the RJMCMC estimation (red) (d) Histogram of intensities (blue) of one pixel of the trig point and final fit from the RJMCMC estimation (red)

the trig point and its final fitting using RJMCMC. Given the distance involved, and the turbulence of the atmosphere between the sensor and the trig. point, the 3D mesh provides a very reasonable reconstruction of the 3D geometry of the stone and ground surfaces.

### 4.7.4 Influence of the Background Noise on the Number of Peaks

It is clear from equation 4.2 that the background noise will influence the number of peaks. To what extent is an important question. The choice of prior distributions for the number of peaks will also affect the marginal posterior distribution of the number of peaks. Throughout this section, we shall explain how the background noise and the prior distri-

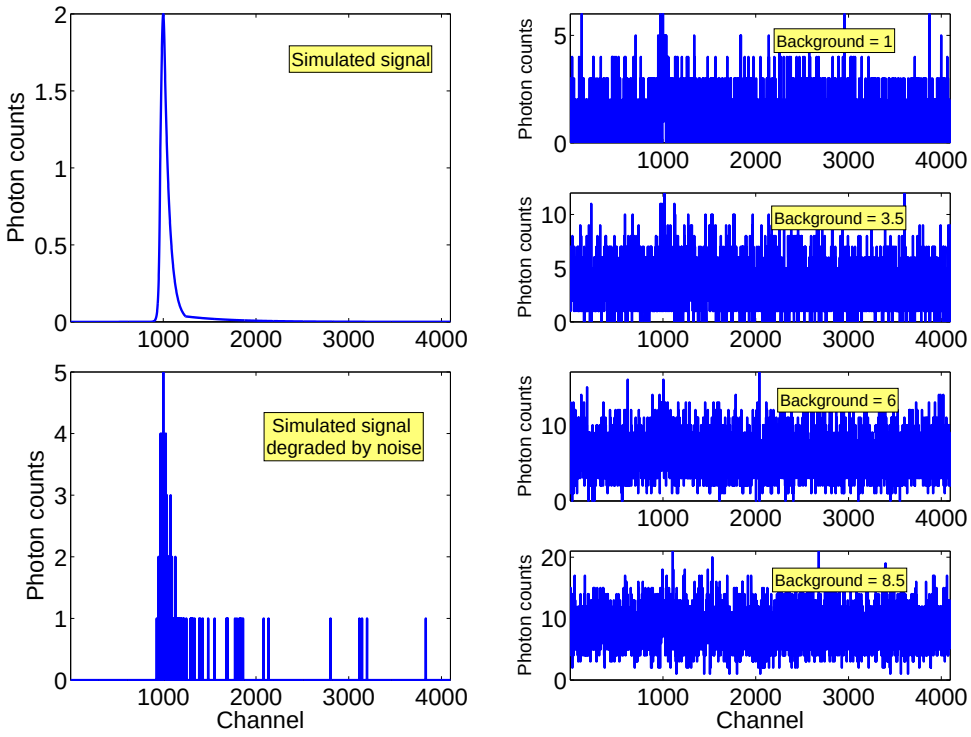



**Figure 4.25:** Details of a simulated signal before and after being degraded by noise and some examples in which the background noise is added

bution on the number of peaks affects the posterior distribution of the number of peaks. To quantify the effect of the background we simulated a return with one single peak as shown in figure 4.25. Then, this signal was degraded by Poisson noise. Subsequently, background noise was generated and added to the simulated signal. The expected value of this background noise was considered constant across all the channels taking values from 0 to 10 counts in steps of 0.5. Figure 4.25 shows some examples of the final simulated signal for different background levels.

In this experiment, the support of the prior on $k$ is $[0, 1]$. The parameters of the priors were set identical to those on section 4.7.1. We have run 100 independent Markov chains for each specific level of the background. Each Markov chain was run for 1000 iterations. Figure 4.26 shows the scatter plot of the estimated position value against the estimated amplitude value. These estimates were obtained as in previous sections using the mean value of the marginal posterior distributions, i.e. $p(\phi|k = \hat{k}, \boldsymbol{y})$ where $\hat{k} = argmax\ p(k|\boldsymbol{y})$.

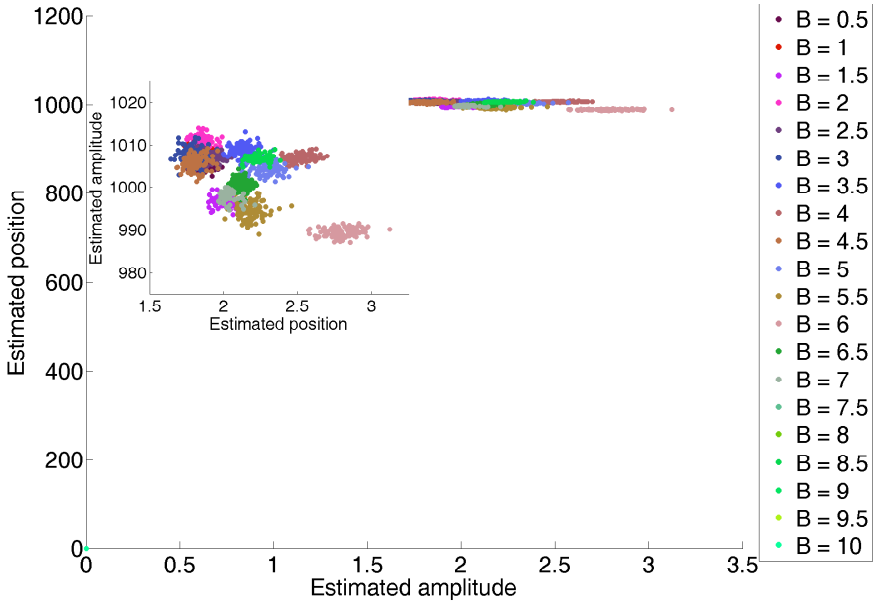



**Figure 4.26:** Estimated position value ($1bin \simeq 6.1ps$) vs. estimated amplitude value obtained with the RJMCMC algorithm. The insert shows an expansion of the graph not considering position/amplitude values equal to 0

From this scatter plot we can uncover relationships between these parameters. Two different data clusters are formed and are situated around channel position 1000 and channel position 0. As we shall see later, the values at channel 0 and amplitude 0 correspond to those cases where no peak was detected. As can be seen from the figure, small variations of the position parameter affect considerably the amplitude values. As the background level surpasses the background level 6.5, the amplitude and position values tend to 0.

Figures 4.27, 4.28 and 4.29 display the boxplots of the position, amplitude and background estimates respectively. These plots will allow us to detect and illustrate location and variation changes between datasets with different background level. Figure 4.27 shows that the median values of the estimates of the position are relatively similar when the background level is below 7. As the background level increases, the median value tends to 0 which is consistent with the fact that it gets more difficult to detect a peak. A similar situation occurs to the amplitude estimates. However, the variation of the amplitudes values is larger since they are more affected by the Poisson noise. The median values of the background estimates are consistent with the actual values of the background as shown in figure 4.29.

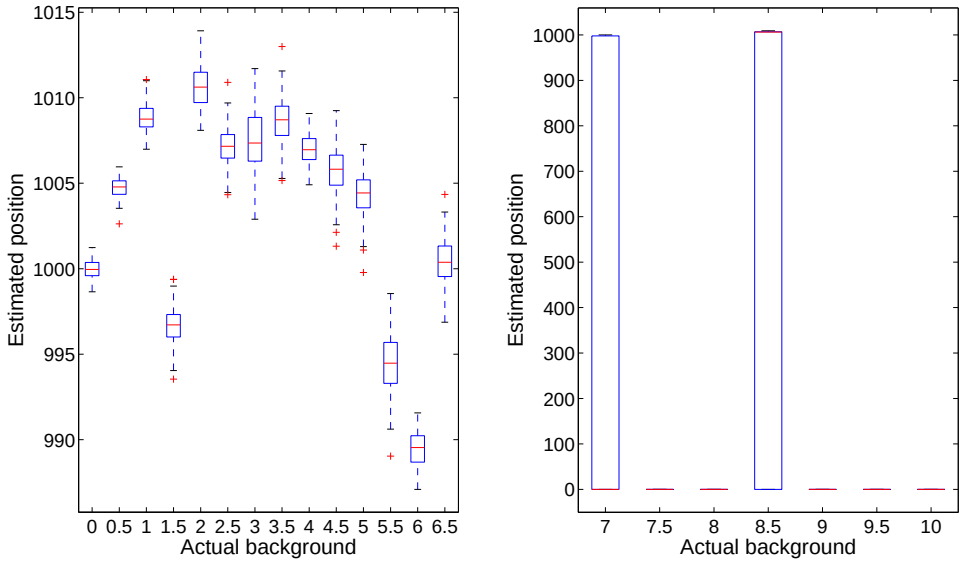



**Figure 4.27:** Boxplot of the estimated positions (100 values) at different simulated background levels ($1bin \simeq 6.1ps$)

The interquartile ranges for the position and amplitude estimates are reasonably similar for background levels below 7. Over this level, two datasets present large interquartiles ranges which suggests large data uncertainty. Figure 4.30 shows a histogram of the estimated number of peaks for different background levels. As the background level increases the probability of detecting a peak decreases. Particular cases are those corresponding to simulated background levels of 7 and 8.5 respectively. For these values, the probability of detecting or not a peak is similar. This is due in part to the prior chosen which gives low probability to smaller values of the amplitude. In the remaining cases, the algorithm suggests that no peak is present on the data as most of the estimated values are 0. In the background estimates case, the interquartiles are relative small suggesting accurate estimation of them. In general, the different datasets do not present marked skewness except in the cases where the background level is 7 and 8.5 where the distribution is right-skew with a large tail (see figures 4.27 and 4.28). This indicates that longer running times for the Markov chain generated are required to improve inference.

The above analysis show a very important feature. As the background level increases, the data provide very little information about the different parameters, the posterior is therefore dominated by the prior. On the contrary, as the background level decreases, the

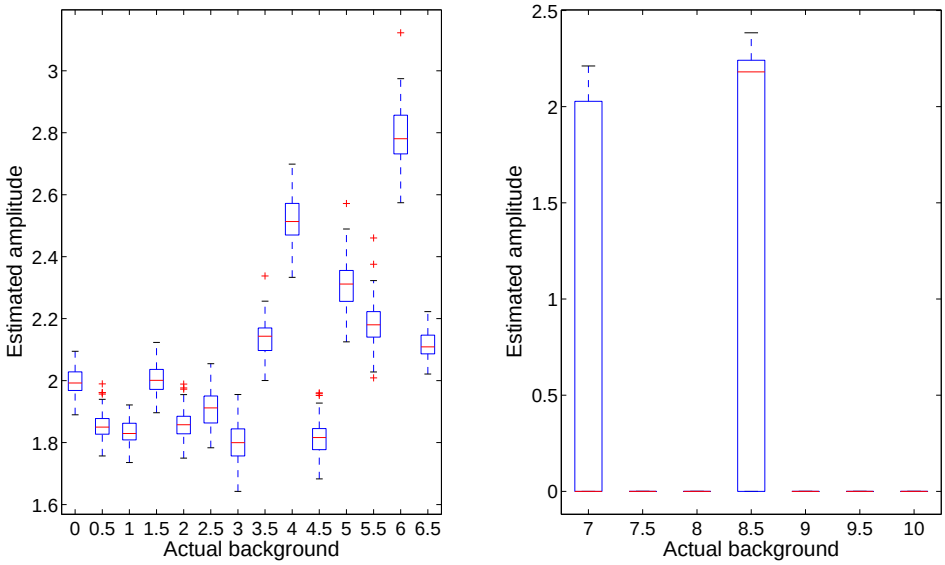



**Figure 4.28:** Boxplot of the estimated amplitudes (100 values) at different simulated background levels

posterior is dominated by the likelihood as the data becomes more informative. In the extreme case, i.e. background level equal to 0, the posterior distribution is unaffected by the prior distribution.

The prior on the number of peaks, $p(k)$ can be used to alleviate to some extent the pathology observed when the amplitude of the return is smaller than background. Indeed, this prior represents our subjective beliefs on the number of peaks in a particular situation. When the background level increases the behaviour of the posterior distribution is conditioned on the prior distribution chosen. Using a uniform distribution on the number of peaks to assess our ignorance on the exact number of peaks is preferred in this work since the aim is to detect all the unknown scatters present in a laser return. In any case, it is easy to convert results from an uniform prior to those corresponding to other priors as stated by Richardson and Green [119] using the following identity (general case)

$$p^*(k, \theta^k|\boldsymbol{y}) \propto p(k, \theta^k|\boldsymbol{y}) \frac{p^*(k)}{p(k)} \tag{4.25}$$

where $\theta^k$ is a set of parameters for dimension $k$ which in this particular section corresponds to $B$ and $p^*(\cdot|\boldsymbol{y})$ denotes the posterior for an alternative prior $p^*$. Note that to compensate pernicious effect of using an uniform prior in pathological cases, we use a more informative prior of the amplitudes to highlight the possible existence of peaks smaller than the

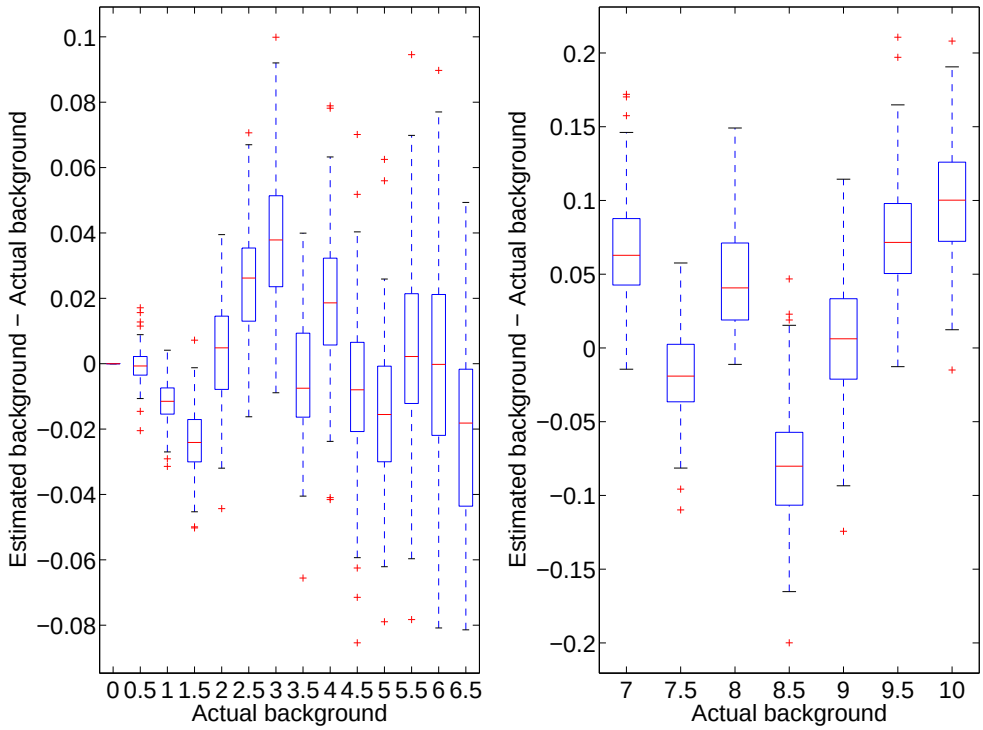



**Figure 4.29:** Boxplot of the estimated backgrounds minus actual values (100 values), to facilitate representation, at different simulated background levels

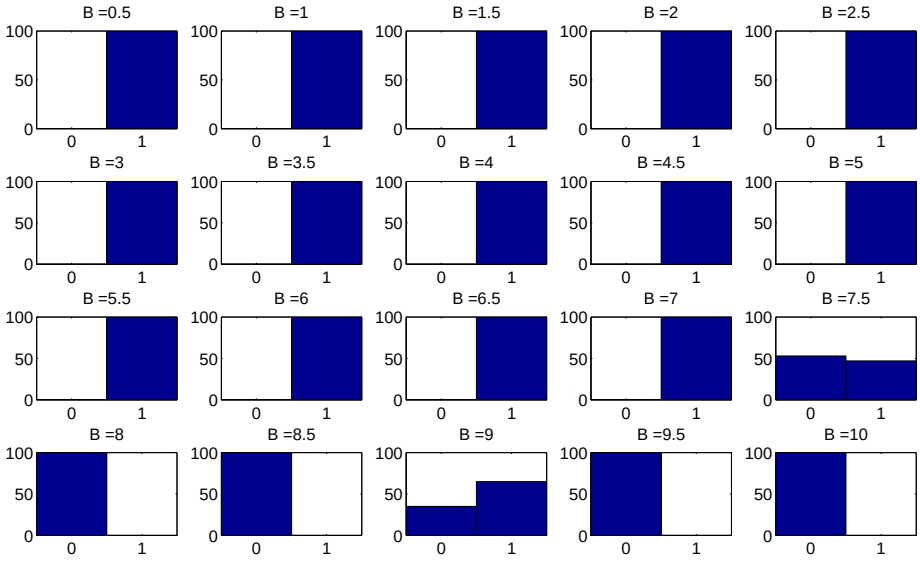



**Figure 4.30:** Histograms of the inferred number of peaks for different background levels

background noise.

## 4.8 Conclusions

We have described the development and application of RJMCMC techniques, incorporating delayed rejection, to process time-of-flight data using both single photon counting and variable temporal gating to extract range measurements. The modelling and processing strategy we have presented has been used to explore posterior distributions in different dimensional state spaces, providing a unified treatment for Lidar data acquired from both TCSPC and BIL sensors. One of the main advantages of this strategy is that it allows the user to incorporate uncertainty about the exact dimension of the problem. Indeed, the algorithm explores different dimensionality spaces and then infer the model, i.e. the number of peaks, which best fits the unknown reality. Moreover, the algorithm is able to infer the values of the parameters corresponding to the model selected. The experiments carried out have demonstrated the ability of this algorithm to determine the number, position and amplitude of the returns under different signal-to-background conditions and when the distance between scatterers is relatively small. It is possible to resolve multiple returns and hence characterise objects distributed in 3D space.

Chapter 5

# Bayesian Spatial Modelling of Lidar Signals

*Without the hard little bits of marble which are called "facts" or "data" one cannot compose a mosaic; what matters, however, are not so much the individual bits, but the successive patterns into which you arrange them, then break them up and rearrange them.*

*Arthur Koestler.*

## 5.1 Introduction

The goal of our work is a complete characterisation of the 3D surfaces viewed by a laser 3D imaging system. In general this includes measuring multiple returns viewed by a single pixel. However, most environments of interest contain spatial patterns that are better characterized by models that capture dependencies among pixels. Pixels contain spatial dependencies since some pixel configurations are more likely than others as multiple returns can be concentrated in certain regions and completely absent in others. Natural images exhibit strong contextual dependencies in the form of spatial interactions among components. There are many reasons why this occurs, for example, when mapping the pixels to landscape patterns, if a pixel is identified as water, it will be most likely surrounded by the same class of pixels. Moreover, different parts of an object are related through geometric constraints. If such a spatial interaction is well modeled, the classification accuracy can be improved [148].

In this chapter, we consider the complex scenario where spatial interaction is present. We incorporate spatial constraints in a multi-layered image (i.e. an image in which each pixel has several surface returns) for the first time. In each pixel there is a multivariate record so we use a mixture distribution for each pixel instead of a "global mixture" that includes all

the possible returns. Moreover, the spatial dependencies are introduced through the prior distribution of the number of components of the mixture. We incorporate two new moves within the RJMCMC algorithm, "spatial mode jumping" and a spatial birth/death process that incorporates spatial information in what is now a mixture of proposal distributions. We also show a delayed-rejection step [144] for a variable dimension setting in a RJMCMC algorithm.

In the next section, we discuss why spatial contextual information is important in the TCSPC and BIL systems and related work. In Section 5.3 we consider a model for TCSPC and BIL data which incorporates spatial constraints. In Section 5.4 we describe our research methodology using spatial constraints in a Bayesian framework. In Section 5.5 we show how the two proposed moves mentioned previously can be incorporated in a RJMCMC analysis of the posterior distribution obtained using a Bayesian approach. In section 5.6 we present results obtained with this methodology. Finally, in Section 5.7 some conclusions are drawn.

## 5.2 Background and Related Work

One of the basic goals of computer vision is to produce a 3-dimensional description of any scene from the surrounding world, given its intensity/reflectance or range image(s). Such a description could then be used in a variety of ways in tasks, *e.g.* terrain mapping, object recognition, medical image analysis, and robot navigation. 3D reconstruction techniques are typically divided into two families: range-based techniques and intensity image-based techniques (see chapter 2). Each one of these techniques has advantages and limitations.

TCSPC and BIL systems are examples of 3D Ladar/Lidar techniques that can acquire both depth and reflectance images of objects. Roughly speaking, each pixel of these systems records a multivariate measurement which can be considered as an observed photon or intensity histogram which in turn is considered as a sample of a non-normalized statistical mixture distribution. In the histograms created, the returns present come not only from the first surface encountered by the projected laser signal, but also from subsequent surfaces in its path as explained in section 4.2.

Computer vision is therefore confronted with the inverse problem of recovering surfaces from images. Unfortunately, this is an ill-posed problem in a Hadamard sense [112] and providing a satisfactory solution is a hard issue. Many deterministic and stochastic approaches have been proposed to deal with this problem when each pixel consists of a univariate measurement. The common point of these approaches is the consideration of spatial dependencies among the different pixels. Standard regularization methods [69, 96] are deterministic approaches that restrict the number of possible solutions by imposing the smoothness constraint, which prevents the local behaviour of the reconstruction from becoming too disjoint. Although standard regularization methods may lead to satisfactory solutions, they cannot deal with a few general problems. Indeed, since these algorithms require that the search space has only one local minimum, they cannot deal with discontinuities, and stochastic approaches to the problem are needed. Further, fusion of information from multiple modules, *e.g.* intensity and range detectors, is not possible [95].

Spatial statistics provides a framework to cope with the problems associated with standard regularization methods and to quantify generic and a priori knowledge. Spatial statistics attempts to develop the best representation of the data set from which the maximum statistically robust information can be extracted. Within the framework of spatial statistics, the Bayesian approach has played predominant role in image analysis and reconstruction, pioneered by Grenander [67], Geman and Geman [59] and Besag [14]. In the Bayesian paradigm, a model of the supplier response of the device, the likelihood, is combined with the prior distribution to obtain the posterior distribution of the quantity of interest. This prior distribution contains a priori knowledge of the phenomenon under observation and information about the local regularities that are believed to exist. Markov random field (MRF) (see Besag [13]) theory provides a convenient and consistent way for modeling context dependent entities such as image pixels. Although the posterior probability completely describes the state of certainty about any possible image, it is often necessary to select a single image as the result or reconstruction. One basic use of Bayesian image analysis is therefore to produce point estimates of pixel images. The most obvious Bayesian point estimate is that which maximizes the a-posteriori probability (Maximum a posteriori (MAP)), however, if multimodality in the space of solutions is suspected, finding the global maximum is extremely difficult. Other ways to obtain estimates include Markov

chain Monte Carlo (MCMC) methods, such as Gibbs sampling or simulated annealing [59], marginal posterior modes (MPM) [15] or deterministic methods, such as iterated conditional modes (ICM) [14] or modified expectation maximization (MEM) [62]. Among all these methods, MCMC is usually the only approach for assessing the uncertainty in the estimates. Examples of surface reconstruction using the Bayesian approach are found in many applications such as single photon emission computerised tomography (SPECT) [62] and ranging [162].

Many spatial problems, however, are inherently multivariate, in that more than one variable is typically measured at each spatial location. Multivariate data analysis allows users to display many different spatial data layers. Mardia [93] introduced a multivariate MRF model for image processing although, initially, this model received little attention due to computational difficulties. In recent years, interest has grown in the analysis of multivariate lattice data with special emphasis on remotely sensed data and more concretely in multispectral image processing [118] in applications of target detection and scene classification and segmentation. Multispectral imaging systems use several spectral bands to acquire data. Assuming we can treat this data as a collection of vector-valued observations on a two-dimensional lattice, a multivariate MRF (more concretely a multivariate Gaussian MRF) can be used to extract relevant information as proposed by Hazel [74]. However, as the number of bands increase, the data volumes are substantially increased and this makes difficult to employ multivariate MRF.

In many situations the observed multivariate data may arise from a time series of measurements of some determined phenomenon as in the case of TCSPC and BIL data. A simple approach to the problem would be to treat the time series of observations at a given image pixel as independent and identically distributed to that for all other pixels, and to analyse them with an appropriate model to understand the underlying theory of the data points. To model such data one can use a parametric approach as in chapter 4 making use of appropriate mixture distributions and then use the techniques described therein to obtain estimates of the different parameters. However, such an approach may give suboptimal results because it neglects the correlations between the parameters in neighboring pixels. For instance, returns coming from the same surface are expected to have similar

amplitude, position and number of peaks. In a frequentist framework, applying EM algorithms to learning spatial models is not straightforward, as there may be context-specific constraints that are often difficult to incorporate. Some examples of these algorithms can be found in the literature [42, 159], however, either they do not consider multivariate data or problems with an unknown number of dimensions.

Recently, there have been various approaches in introducing spatial dependencies in a mixture distribution to take into account spatial heterogeneity using a Bayesian approach [52, 66]. However, all of them consider that the different observations come from a unique mixture distribution in which the number of components and some of the parameters (maybe all) are variable. In these approaches, the spatial dependencies were introduced either through the weights of the mixture distribution [52] or through the different allocation variables [66], and posterior inference was performed using RJMCMC algorithms. Initially, one may consider applying a multivariate MRF to model the spatial dependencies among the different parameters of the different mixture distributions. However, it is not clear how to use these structures in a variable dimension setting where the different peaks are continuously created or deleted to better explore the space of solutions.

## 5.3 Model

To interpret the TCSPC and BIL data we follow the same approach as in Chapter 4. However, we have changed the notation to include spatial contextual information. Consider a rectangular grid of pixels, labeled $m = 1, 2, \ldots, N$ at each of which a multivariate measurement $\boldsymbol{y}^m = \{y_1^m, y_2^m, \ldots, y_t^m\}$ is available. Each multivariate measurement is a histogram of photon counts or intensities in which $y_i^m$ is the value recorded in channel $i$, $i = \{1, \ldots, t\}$.

We use indistinctly the notation $f_{op}$ when referring to the TCSPC operating model or the BIL instrumental response and make the same assumptions as in the previous chapter. We summarise them here. In the same histogram, several signals can be present and that these signals will be observed against a finite background level, $B_m$, whose expected value is considered as a constant across all the channels, the observed histogram in pixel $m$,

$\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m)$, can be considered as a sample of a statistical mixture distribution with density

$$\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m) = \sum_{j=1}^{k_m} \beta_{mj} \cdot f_{op}(i, t_{0_{mj}}) + B_m \tag{5.1}$$

where $k_m$ is the number of peaks, $\boldsymbol{\beta_m} = (\beta_{m1}, \beta_{m2}, \ldots, \beta_{mk_m})$ is a vector of amplitude factors and $\boldsymbol{t_{0_m}} = (t_{0_{m1}}, t_{0_{m2}}, \ldots, t_{0_{mk_m}})$ is a vector of the time of the peak maxima.

If the time resolution is sufficiently fine, the value $y_i^m$ recorded can be considered to be a random sample of a Poisson distribution with intensity equal to $\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m)$ which depends on the different model parameters.

$$y_i^m \sim Poisson(\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m)) \tag{5.2}$$

In order to build up a likelihood function, we make two further assumptions. First, the observations recorded in each channel $i$ of the histogram are conditionally independent given the value of the parameters. Second, the spatial dependencies are included in Bayes' equation through parameters representing spatial interactions so that the different $\boldsymbol{Y} = (\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^N)$, are therefore spatially independent given the parameters. The likelihood for the total array is expressed by the following equation

$$L(\boldsymbol{Y}|\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}) = \prod_{m=1}^{N} \prod_{i=1}^{i_{\max}} e^{-\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m)} \frac{\mathcal{F}_m(i, k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, B_m)^{y_i^m}}{y_i^m!} \tag{5.3}$$

where $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$, $\boldsymbol{\Omega} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta_N})$, $\boldsymbol{T} = (\boldsymbol{t_{0_1}}, \boldsymbol{t_{0_2}}, \ldots, \boldsymbol{t_{0_N}})$ and $\boldsymbol{B} = (B_1, B_2, \ldots, B_N)$.

## 5.4 Bayesian Inference with Spatial Constraints

The objective is inference about the parameters of equation 5.3 in order to obtain accurate estimates of the number of peaks, position, amplitude and background of the returned signals. If we do not have any spatial contextual information then these unknowns are regarded as drawn from appropriate independent prior distributions. As in the previous chapter, the position parameter, $t_{0_{mj}}$, is drawn from a flat uniform distribution with support in the interval $[1, i_{max}]$, to reflect in some way our complete prior ignorance of

the true position of the peaks. For the prior distribution of scale parameters $\beta_{mj}$ and $B_m$ we assume a gamma distribution since the support of our parameters is on $\Re^+$. The support of the prior for the number of peaks, $k_m$, is $[0, ..., k_{max}]$ with $k_{max}$ chosen to be suitably large. The full joint prior distribution when no spatial constraints are considered is therefore

$$\pi(k_m, \boldsymbol{\beta}, \boldsymbol{t_0}, \boldsymbol{B}) = \left(\frac{1}{k_{max}}\right)\left(\frac{1}{i_{max}}\right)^k f_G(B_m|c,d)\prod_{j=1}^{k} f_G(\beta_{mj}|a,b) \tag{5.4}$$

where $f_G$ is the probability density function (pdf) of a gamma distribution with mean $a \cdot b$ ( $c \cdot d$ respectively) and variance equal to $a \cdot b^2$ ($c \cdot d^2$ respectively).

In this chapter, we consider that observations that correspond to nearby locations are more likely to have similar numbers of peaks than observations from locations that are far apart. Therefore, we generalise the prior distribution given by equation 5.4 including a penalty function, the Potts model, that discorages adjacent pixels to have different numbers of peaks. The Potts model is a generalisation of the Ising model introduced in section 3.5.1.2 in which the random variable is allowed to have more than two different values. The Potts model has been used previously in image processing applications and in disease mapping applications to model allocation variables. We follow the formulation used in [66] in which $\boldsymbol{k}$ is modeled jointly

$$p_1(\boldsymbol{k}|\psi) = \frac{e^{\psi \cdot U(\boldsymbol{k})}}{\sum_{k \in \{1,2,\cdots,k_{max}\}^N} e^{\psi \cdot U(\boldsymbol{k})}} \tag{5.5}$$

where $U(k) = \sum_{m \sim m'} I[k_m = k_{m'}]$ are the number of like-labeled neighbouring pairs in the configuration $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$. The parameter $\psi$ is nonnegative and controls the amount of "smoothing", that is, $\psi = 0$ corresponds to a priori spatial independence of the number of peaks and as $\psi \to \infty$ we favour patterns where the number of peaks in neighbouring pixels tend to be similar. In what follows, $\psi$ is assumed known and will have a fixed value.

The full joint prior distribution is now modeled as

$$\boldsymbol{f}(\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}) = p_1(\boldsymbol{k}|\psi) \times \prod_{m=1}^{N} \pi(k_m, \boldsymbol{\beta_m}, \boldsymbol{t_{0_m}}, \boldsymbol{B_m}) \tag{5.6}$$

Using equations 5.3, 5.6 and Bayes' theorem the target distribution can be expressed as

$$\pi(\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}|\boldsymbol{Y}) \propto L(\boldsymbol{Y}|\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}) \boldsymbol{f}(\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}) \tag{5.7}$$

since we just have to know the posterior distribution up to a normalizing constant.

As simulating from the joint probability distribution is difficult we propose to update the different parameters using their respective full conditional distributions. We define a Markov random field and use a second-order neighbourhood. The Hammersley-Clifford theorem [13] ensures that we can use the full conditional distribution of equation 5.5.

$$p_m(k_m|\ldots) \propto \exp(\psi \sum_{m' \in \ \delta_m} I[k_m = k_{m'}]) \tag{5.8}$$

where $\delta_m$ denote the neighbours of $m$. In this way, a rather complex multivariate probability distribution of a MRF can be obtained by successive simulations from the full conditional distributions.

## 5.5 RJMCMC Methodology

The Bayesian models that are proposed in this chapter are too complex to be amenable to analytical calculations. We use reversible RJMCMC techniques [66] to infer the number, positions and amplitudes of the returned signals from target surfaces. This allows moves between state spaces with different dimensionality, i.e. different numbers of viewed surfaces, and moves when the dimension is fixed to improve the estimation of the parameters of such subspace, i.e. position (depth) and amplitude (reflectance) of surfaces. RJMCMC provides a natural method for updating beliefs in response to new information and incorporates prior knowledge into the analysis. Indeed, RJMCMC techniques allow us to explore the full posterior distribution of the parameters of the mixture distribution of equation 5.1, given the data values $Y$ supplied by the different histograms. The parameters we seek are estimated from the values of a Markov chain whose limiting distribution is a target distribution $\pi$. In our case, $\pi$ arises from the posterior distribution defined by equation 5.7, nevertheless, we use simpler versions of it involving only full conditional distributions. The Markov chain constructed involves moves of various types. These are:

1. Updates to parameters $\boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}$,
2. Random birth of a peak,
3. Death of a peak,
4. Random splitting of a peak into two peaks
5. Merging of two peaks into a single peak

To explore the different state spaces we also incorporate a delayed-rejection step in the RJMCMC algorithm as proposed by Green and Mira [65]. As in the standard RJMCMC algorithms, our problem starts with the formulation of an intractable posterior distribution $\pi(\cdot)$ from which we want to sample and the construction of a Markov chain whose limiting distribution is this posterior distribution. In a Metropolis-Hastings update, such a Markov chain is generated by proposing new values for some or all variables drawn from a proposal distribution and then an accept/reject decision is made, depending on the generated values and on the form of the proposal and target distributions. In the delayed-rejection strategy, if a candidate move is rejected, we make another attempt to move using a second proposal instead of turning to the next transition. There is no restriction on the number of stages to be used although we shall use only two stages in this work, to reduce the computational load.

To perform a RJMCMC with a delayed-rejection stage, three different acceptance probabilities have to be calculated. Using Green and Mira's notation [65] for general state spaces, to move from a state $x$ to a state $xx$, we propose to draw a random number $u_1$ from a known density $g_1$. The new stage $xx$ is calculated as $xx = h_1^+(x, u_1)$. The reversed move, from $xx$ to $x$, is performed by drawing $u_1'$ from $g_1'$ and calculating $x = h_1^-(xx, u_1')$ where $h_1^+$ and $h_1^-$ are deterministic mappings. The acceptance probability of this first stage is given by

$$\alpha_1(x, xx) = min\left\{1, \frac{\pi(xx)g_1'(u_1')}{\pi(x)g_1(u_1)}\left|\frac{\delta(xx, u_1')}{\delta(x, u_1)}\right|\right\} \tag{5.9}$$

If this candidate move is rejected, a new move to state $z$ is proposed in a similar way to the previous candidate move, this move is accepted or rejected with a probability given

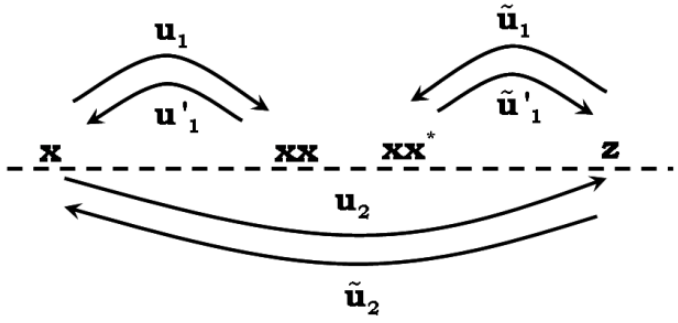



**Figure 5.1:** Schematic adapted from [65] representing the variables in play

by

$$\alpha_2(x,z) = min\left\{1, \frac{\pi(z)\tilde{g}_1(\tilde{u}_1)\tilde{g}_2(\tilde{u}_2)\{1-\alpha_1(z,xx^*)\}}{\pi(x)g_1(u_1)g_2(u_2)\{1-\alpha_1(x,xx)\}}\left|\frac{\delta(z,\tilde{u}_1,\tilde{u}_2)}{\delta(x,u_1,u_2)}\right|\right\} \tag{5.10}$$

where $z = h_2^+(x,u_1,u_2)$ and $x = h_2^-(z,\tilde{u}_1,\tilde{u}_2)$. The values $\tilde{u}_1, \tilde{u}_2$ and $u_2$ are drawn from $\tilde{g}_1, \tilde{g}_2$ and $g_2$ respectively.

To obtain the numerator of the second acceptance probability we need to calculate a third acceptance probability, $\alpha_1(z,xx^*)$, corresponding to a fictional stationary Markov chain started in $z$ which proposes a move to $xx$ at the first stage, rejects it and accepts a second stage move to $x$. Although this chain is not really implemented, it is needed to ensure the reversibility condition and therefore preserve the stationary distribution. In the original work of Tierney and Mira [144], it was required that $xx = xx^*$. However, this assumption limited the application of the delayed-rejection algorithm to variable dimension settings. Fortunately, this is not a required condition to preserve the stationary distribution and alternative values of $xx^*$ can be devised. A schematic diagram of the moves adapted from [65] is presented in figure 5.1. The acceptance probability of this virtual stage is given by

$$\alpha_1(z,xx^*) = min\left\{1, \frac{\pi(xx^*)\tilde{g}_1'(\tilde{u}_1')}{\pi(z)\tilde{g}_1(\tilde{u}_1)}\left|\frac{\delta(xx^*,\tilde{u}_1')}{\delta(z,\tilde{u}_1)}\right|\right\} \tag{5.11}$$

where $\tilde{u}_1'$ is such that $z = h_1^+(xx^*,\tilde{u}_1')$. Note that these different stages do not destroy the Markovian property of the sampler and therefore the asymptotic Markov chain theory used for standard MCMC methods holds true. The advantage of this strategy is apparent if the optimal variance of the proposal is not constant throughout the state space. Furthermore, this stage is allowed to depend on the current state of the chain and also on that just

proposed and rejected. Therefore we learn from the information obtained during the run of the chain.

### 5.5.1 Fixed-dimension Moves

Assume we are at pixel $m$ and the current state of the parameter variable set is $\phi$, with $\phi = (\boldsymbol{k}, \boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B})$, we propose a move $n$ to a state $\phi' = (\boldsymbol{k}', \boldsymbol{\Omega}', \boldsymbol{T}', \boldsymbol{B}')$ drawn from the proposal probability density, $q_n(\phi, \phi')$. The proposed move is accepted with a probability given by equation 5.9. If the proposed move is accepted, the new value of the Markov chain becomes $\phi'$ otherwise we keep the current value $\phi$. Note, that if the number of peaks does not change, i.e. a fixed dimension of the space of solutions, the acceptance probability of equation 5.9 reduces to the Metropolis-Hastings acceptance probability and we perform a parameter updating using the full conditional distributions as a target distribution. In this work, the different parameters $\boldsymbol{\Omega}, \boldsymbol{T}, \boldsymbol{B}$ are updated separately and as a consequence each of them having a different acceptance probability. In our case, we draw the proposal of the amplitude, $\beta'_{mj}$, and the background, $B'_m$, from the gamma distributions $\Gamma(1.5, \beta_{mj})$ and $\Gamma(1.5, B_m)$ respectively as in section 4.5.1 as well as the delayed-rejection move associated with the update of the amplitude parameters. The proposal for the position is more complicated since it involves spatial information from neighbouring pixels.

#### 5.5.1.1 Spatial Mode Jumping

In practice, the objects imaged by the system that have the same attributes tend to cluster. Hence, group of pixels with same number of peaks and similar parameter values are expected to occur together. Spatial interactions for positions are also expected to occur. However, only spatial interactions for the number of peaks are represented in the prior. Furthermore, the Potts model constrains the number of peaks without considering the relative positions and amplitudes of the current pixel with respect to its neighbourhing pixels. Since such contextual information is not available in the prior, it may happen we encounter situations in which the state space is not explored properly. Therefore, we suggest a proposal distribution for position moves which can exploit that we expect data to show correlated positions. To this end, we incorporate a new move, called "spatial mode jumping" (inspired by the work of Tjelmeland and Hegstad [147]), within the RJMCMC

algorithm described in chapter 4. The "spatial mode jumping" move proposes an update to the position of a peak in pixel $m$ so that its position corresponds to a random perturbation of the position of a peak in pixel $m'$ with some probability. In this way, we incorporate spatial information on the proposal distribution.

There are two different ways of incorporating this new move. On the one hand, we may consider "spatial mode jumping" as a distinct move of the "position update" move. On the other hand, we may merge the "spatial mode jumping" move and the "position update" move into a single move. Both methodologies have their own advantages and disadvantages. An independent move is easier to implement, however, it is expected that a large number of rejections in the RJMCMC algorithm will be attained. If both moves are merged, a delayed-rejection step can be used every time we reject a "spatial mode jumping" move. This delayed-rejection step will propose an update of the current position parameter instead of a spatial mode jumping. The main disadvantage of this move is the increase of complexity of the acceptance probability. In this work, we shall use the second approach to present an uniform theory of spatial and non-spatial position update. The new step is therefore based on proposing moves of the current peak in pixel $m$ that depend on its current position and on the current positions of its pixel neighbours, $\delta_m$. One could consider the whole array of pixels as the neighbourhood. However, in our implementation the size of this proposal is limited by the size of the second-order neighbourhood of the Potts model. We propose to use a mixture of proposals distributions which take account of the current values of both the peaks of $m$ and $\delta_m$ given by the following expression

$$Q(\theta, \theta') = w_m \cdot q_m(\theta, \theta') + \sum_{l \in \delta_m} w_l \left( \sum_{t=1}^{k_{max}} \nu_{lt} \cdot q_t(\theta, \theta') \right) \tag{5.12}$$

where $\theta$ is the current estimate of the parameters, $\theta'$ the proposed values, $q$ are proposal distributions, $k_{max}$ is is a suitable large number of peaks and $w$ and $\nu$ are weights that satisfy $\sum_{l=1}^{9} w_l = 1$ and $\sum_{t=1}^{k_{max}} \nu_{lt} = 1$ respectively. These weights define positive probabilities in such a way that at each step one of proposals is selected according to these probabilities.

#### 5.5.1.2 Computational Implementation

In this section, we describe how the "spatial mode jumping" move described in section 5.5.1.1 has been implemented using a RJMCMC algorithm with delayed rejection. The key element of this move is the mixture of proposals given by equation 5.12. To calculate the acceptance probability we need first to define the different probabilities involved in the mixture of proposals. Indeed, each component density of the mixture has to be sampled from some probability distribution to generate a proposed value for the Markov chain. The weights $w$ will determine from which pixel we are going to "borrow" some information to propose a new value and depends on the current number of peaks. Assuming we are at pixel $m$, the following conditions apply

1. The probability of choosing pixel $i'$ is $w_{i'} = \frac{I[k^{i'} \neq 0]}{\sum_{s=1}^{9} I[k^s \neq 0]}$, where $k^s$ is the number of peaks at pixel $s$ and $I[\cdot]$ is the indicator function,
2. If no peaks are available on the neighbourhood to extract some spatial information, $w_m = 1$, $w_{i' \neq m} = 0$ and we propose updates in the manner of previous chapter.

On the other hand, the weights $\nu_l$ of a pixel $l$ are chosen proportionately to the number of peaks on such pixel. In this work, uniformity on the choice of the peak is chosen for simplicity. More complex weighting schemes based on Euclidean distance could be used. However, these will require to use some kind of labeling of the different variables which would complicate the analysis of the data.

Suppose we are updating the values of peak $i$ at pixel $m$ and we choose pixel $i'$ with some probability determined by $w_{i'}$. If $i' \neq i$, we choose peak $j$ conditioned on $i'$, with probability $\nu_{i'j}$. The proposed position for peak $i$, $xx^i$, is drawn from the following normal distribution $\mathcal{N}(x^j, \sigma_j)$ where $x^j$ is the current position of the peak $j$ at pixel $i'$. If no peak from neighbouring pixel is selected, we update the current estimate of the Markov chain using a standard random walk, $xx^i \sim \mathcal{N}(x^i, \sigma_i)$ where $x^i$ is the current position of the peak $i$ at pixel $m$. Initially, $\sigma_j$ can be as small as desired and different for each pixel. To simplify, we set $\sigma_j = \sigma_1$, $\forall j \neq i$ as we do not have any knowledge of the solution space. In general, we shall make $\sigma_i \gg \sigma_j$. Allowing $\sigma$ large in the first stage of the algorithm will

allow better exploration of the state space. The acceptance probability of the first stage is given by

$$\alpha_1(x^i, xx^i) = min\left\{1, \frac{\pi(xx^i)\cdot\left(w_m\cdot q'_m(xx^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot q'_t(xx^i)\right)\right)}{\pi(x^i)\cdot\left(w_m\cdot q_m(x^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot q_t(x^i)\right)\right)}\right\} \tag{5.13}$$

where $q_m(\cdot) = q'_m(\cdot) = \mathcal{N}(x^i, \sigma_i)$ and $q_{lt}(\cdot) = q'_{lt}(\cdot) = \mathcal{N}(x^t, \sigma_t)$. This algorithm is valid as it is just a particular case of the Metropolis-Hastings algorithm.

For the delayed-rejection step, we propose to update the current position not considering spatial dependencies. It is also possible to perform a delayed-rejection step which learns from the previous rejection and proposes another "spatial mode jumping" step. The procedure for the second stage is similar to that of the first stage, although using equation 5.10 and not considering the adjacent pixels. Indeed, the proposed value $z^i$ is drawn from the following normal distribution $\mathcal{N}(x^i, \sigma_2)$. In other words, to move from $x$ to $z$ (and respectively from $z$ to $x$) we perform a standard random walk with standard deviation equal to $\sigma_2$. The value of $\sigma_2$ is in general small to allow better local exploration.

As stated before, to calculate the acceptance probability of the second stage it is necessary to calculate the acceptance probability of the virtual step given by equation 5.11. The procedure followed is identical to that of the first stage but starting from state $z$ instead of state $xx^*$. To simplify (and because we do not perform a change-of-dimension move) we assume that the virtual stage $xx^*$ is equal to the previous rejected step $xx$. The acceptance probability is given by the following equation

$$\alpha_1(z^i, xx^i) = min\left\{1, \frac{\pi(xx^i)\cdot\left(w_m\cdot \tilde{q}'_m(xx^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot \tilde{q}'_t(xx^i)\right)\right)}{\pi(z^i)\cdot\left(w_m\cdot \tilde{q}_m(z^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot \tilde{q}_t(z^i)\right)\right)}\right\} \tag{5.14}$$

where $\tilde{q}'_{1l} = \tilde{q}_{1l} = q'_{lt}$ and $\tilde{q}'_m = \tilde{q}_m = q'_m$. Thus, the acceptance probability of the second stage is $\alpha_2(x, z) = min\{1, A\}$, where

$$A = \frac{\pi(z)\cdot\tilde{q}_2(\tilde{u}_2)\cdot\left(w_m\cdot\tilde{q}_m(z^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot\tilde{q}_t(z^i)\right)\right)\cdot\{1-\alpha_1(z, xx)\}}{\pi(x)\cdot q_2(u_2)\cdot\left(w_m\cdot q_m(x^i) + \sum_{l\in\delta_m} w_l\left(\sum_{t=1}^{k_{max}}\nu_{lt}\cdot q_t(x^i)\right)\right)\cdot\{1-\alpha_1(x, xx)\}} \tag{5.15}$$

and where $\tilde{q}_2 = q_2 = \mathcal{N}(x^i, \sigma_2)$. Since the dimension-balancing holds in all the equations, this algorithm is a-priori valid.

### 5.5.2 Variable-dimension Moves

The birth, death, split and merge moves imply changes in the parameter-space dimensionality ($k$ changes). In the case of a birth or split move, the acceptance probability of equation 5.9 can be reduced to:

$$\alpha_m^1(\phi, \phi') = min \left\{ 1, \overbrace{\frac{\pi(\phi'|y)}{\pi(\phi|y)}}^{Target\ ratio} \times \overbrace{\frac{r_n(\phi')}{r_n(\phi)q(u)}}^{Proposal\ ratio} \times \overbrace{\left| \frac{\delta(\phi')}{\delta(\phi, u)} \right|}^{Jacobian} \right\} \tag{5.16}$$

where $r_n(\cdot)$ is the probability of choosing move type $n$ when in state $\phi$, $u$ is a vector of continuous random variables which ensures the reversibility of the deterministic function $\phi' = h(\phi, u)$, which allows the move to a higher-dimensional space, and $q(\cdot)$ is the pdf of $u$. The Jacobian term joint to this deterministic function ensures the "dimension-balancing" condition which allows reversible jumps between different state spaces. The Jacobian arises from the change of variable from $(\phi, u)$ to $\phi'$. Note that the death and merge moves are obtained by inverting equation 5.16 with an appropriate change of index.

The split and merge moves are identical to those explained in section 4.5.2. For a split, a peak is chosen at random and divided into two peaks. The split is performed so that the two new peaks equate to the total mass of the peak split. For a merge, the initial peaks to be amalgamated must be within a temporal separation below a certain threshold, and mass balancing is again considered, that is, the total mass of the new peak must be equal to the addition of the total mass of the initial peaks. However the birth and death process are different in the sense that they also include spatial information and a delayed-rejection step to improve mixing of the chain. For computational reasons, when using a delayed-rejection step, as will be explained in section 5.5.2.2, we have split the birth move as well as the death move into two different kernels respectively. As a consequence, we consider a birth move, called $0 - birth\ step$[1], to propose the birth of a peak in situations where there is only background noise. The $0 - birth\ step$ does not incorporate a delayed-rejection step. In the remaining cases, the birth move uses a delayed-rejection step so that is the reason we distinguish the different birth moves. A similar situation occurs for the death move. In this case, we consider a death move, called $0 - death\ step$, to propose the death

[1]The 0 in $0 - birth\ step$ stands for the current number of peaks whereas in the $0 - death\ step$ case, it stands for the updated number of peaks is the move is accepted

of a peak when only a single peak is present. Again, this move does not make use of a delayed-rejection step, whereas this step is used in the remaining cases. Since the different new kernels fulfill the detailed balance condition as well as aperiodicity and irreducibility, the new mixture of kernels satisfies these conditions and the asymptotic theory of standard Markov chains holds.

Let $b_k^0$, $d_k^0$, $b_k$, $d_k$, $s_k$ and $m_k$ be the probabilities of a $0-birth$, $0-death$, birth, death, split and merge respectively, where $k$ is the current number of peaks. Further, assume all allowable moves are equiprobable. The following conditions apply:

- $b_k^0 + d_k^0 + b_k + d_k + s_k + m_k = 1$.
- If the current number of peaks, $k$, is equal to 0, that is, only background noise is present, only a $0-birth$ step is allowed.
- If the current number of peaks, $k$, is equal to 1, only a $0-death$, birth or split step is allowed.
- If the current number of peaks, $k$, is equal to 1 and smaller than a predefined maximum number of peaks, $k_{\max}$, only a birth, death, merge or split step is allowed.
- If the current number of peaks, $k$, is equal to $k_{\max}$ only a death or a merge step is allowed.

#### 5.5.2.1 Spatial Birth/Death Process

A way to penalise the existence of a small or large number of peaks in a histogram is by using a prior on the number of peaks. In general, this requires a very good knowledge of the problem and the use of very subjective priors. In applications with strong spatial dependencies among the different parameters, one includes this fact in the prior distribution, in our case the Potts model. However, the information contained in the Potts model is not enough, the reason being the lack of information given by this model on the peak to be removed. To deal with this problem we incorporate a spatial birth/death move into the previously described RJMCMC algorithm which penalises in some way the creation or removal of a peak given the information of the neighbourhood of the pixel. Generally, for

a birth we propose values for the parameters of the new peak. In practice, in the case with non spatial information, we propose values drawn from a gamma and a uniform distribution for the amplitude and the position of the peak respectively. For a death, one of the current peaks is chosen at random to be removed. However, in the spatial birth process the proposed values for the amplitude and the position of the new peak are drawn from two mixture distributions whose elements incorporate spatial information from neighbouring pixels in a similar way to that of the "spatial mode jumping". Furthermore, the birth and death moves incorporate a delayed rejection step which learns from the previous rejected value and therefore allows some kind of learning.

Assume we are at pixel $m$, the general expression for the mixture of proposals is similar to that of the "spatial mode jumping" move and is given by the following equation

$$Q(\cdot) = w_m \cdot q_m^1(\cdot) + \sum_{l \in \delta_m} w_l \left( \sum_{t=1}^{k_{max}} \nu_{lt} \cdot q_t^2(\cdot) \right) \tag{5.17}$$

where $q^1$ and $q^2$ are proposal distributions, $k_{max}$ is the maximum number of peaks and $w$ and $\nu$ are weights that satisfy $\sum_{l=1}^{9} w_l = 1$ and $\sum_{t=1}^{k_{max}} \nu_{lt} = 1$ respectively. The first term of equation 5.17 ensures that a birth move is going to be proposed when no peaks are available in the neighbourhing pixels whereas the second term includes spatial information in the proposal distribution. The term $q^1$ represents a gamma distribution when proposing the amplitude of the new peak whereas it represents a uniform distribution when proposing the position of such new peak. The term $q^2$ represents a normal distribution and includes spatial information in a similar spirit to that of the "spatial mode jumping" move. That is, the proposed position/amplitude is drawn from a normal distribution whose mean value is the current position/amplitude of a peak of a neighbouring pixel. The different parameters of these distributions will be explained in the next section.

#### 5.5.2.2 Computational Implementation

In this section, we describe how the "spatial mode jumping" move described in section 5.5.2.1 has been implemented using a RJMCMC algorithm with delayed rejection. Key elements of this move are the mixtures of proposals given in a general form by equation 5.17. As in the "spatial mode jumping", each component density of the mixture has to be sampled according to some probability to generate a proposed value for the Markov

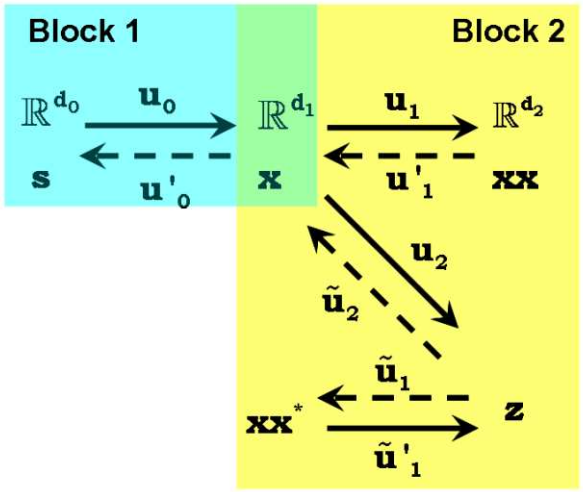



**Figure 5.2:** Relationships among variables for delayed rejection in variable dimensions adapted from [65]. Different colors represent different blocks. Block 1 represents the $0-birth$ and $0-death$ stages. Block 2 represents the birth and death stages with delayed rejection. All these moves include spatial contextual information. The area where the two blocks are superimposed represent common variables for the moves.

chain. This probability is given by the different weights employed in the mixtures of proposals. The procedure to choose a component of the mixture is identical to that explained in section 5.5.1.2. As commented in previous sections, we have split the standard birth and death moves into four different moves $0-birth$, $0-death$, birth and death. This is because there are situations in which the delayed rejection step is not worth implementing, as in the case when we propose a death of a peak when only one peak is present. Indeed, in a standard death move a peak is chosen to be deleted uniformly among the peaks contained in such pixel. If only a single peak is available, a delayed-rejection stage would imply persistent rejection of this move. Therefore, this move is computationally non-efficient. As the birth and death processes are needed in tandem, if one of them does not make use of a delayed rejection step, consequently the other one does not make use of it either. Figure 5.2 shows a schematic of the different stages and variables involved when the number of peaks is 0, 1 and 2. Block 2 corresponds to the standard birth and death moves with delayed rejection. The stages in block 2 can be replicated for larger dimensions. On the other hand, block 1 corresponds to the $0-birth$ and $0-death$ moves and is only used in transitions from 0 to 1 peak and 1 to 0 peaks.

Suppose we are updating the values of peak $i$ at pixel $m$ and we choose pixel $i'$ with some probability determined by $w_{i'}$. If $i' \neq i$, we choose peak $j$ conditioned on $i'$, with probability $\nu_{i't}$. We propose to have a birth in pixel $i$ at position $u_{1p}$ and with amplitude $u_{1a}$ where $u_{1p} \sim \mathcal{N}(x_p^{i'j}, \sigma_b)$, $u_{1a} \sim \mathcal{N}(x_a^{i'j}, \sigma_b)$ and $x_p^{i'j}$ and $x_a^{i'j}$ are the position and amplitude of peak $j$ in pixel $i'$ respectively. An exception occurs when choosing the current pixel $i$, the proposed values for the amplitude and the position are drawn from the $\Gamma(1, (\frac{1}{i_{\max}}) \sum_{i=1}^{i_{\max}} y_i^m)$ and $\mathcal{U}(0, i_{\max})$ distributions respectively which do not incorporate spatial dependencies as in section 4.5.2.1. This allows us to propose birth and death moves within a pixel in cases where no peaks have been detected in the neighbouring pixels. The acceptance probability of the first stage is given by the following expression

$$\alpha_1^i(k^i, x^i, k^i + 1, xx^i) = min\left\{1, \frac{\pi(k^i + 1, xx^i) \cdot d(k^i + 1) \cdot (1/(k^i + 1))}{\pi(k^i, x^i) \cdot b(k^i) \cdot Q_p(u_{1p}) \cdot Q_a(u_{1a})}\right\} \tag{5.18}$$

where $k^i$ is the current number of peaks in pixel $i$, $Q_a(\cdot)$ is the mixture of proposals of the amplitude, $Q_p(\cdot)$ is the mixture of proposals of the position, $x^i$ and $xx^i$ are the parameter vectors of the current and proposed state respectively and the Jacobian term is 1. The death move acceptance probability is the reciprocal of the acceptance probability of the birth move with an appropriate change of index. The expression of this acceptance probability is valid for both block 1 and block 2 with an appropriate change of variables.

With reference to figure 5.2, the second stage corresponding to block 2 is calculated. Again we need to calculate the acceptance probability of the virtual stage given by equation 5.11 which would correspond to a hypothetical death move if we were in state $z$. Thus, we propose to remove a peak in pixel $i$ with probability $1/(k^i+1)$, in which case the acceptance probability is given by

$$\alpha_1(k^i + 1, z^i, k^i, xx^{*i}) = min\left\{1, \frac{\pi(k^i, xx^{*i}) \cdot b(k^i) \cdot Q_p(\tilde{u}_{1p}) \cdot Q_a(\tilde{u}_{1a})}{\pi(k^i + 1, z^i) \cdot d(k^i + 1) \cdot (1/(k^i + 1))}\right\} \tag{5.19}$$

where $\tilde{u}_{1a}$ and $\tilde{u}_{1p}$ are respectively the amplitude and the position of the peak to be removed, $z^i$ and $xx^{*i}$ are the parameters vectors of the proposed state in the second stage and proposed virtual state respectively and the Jacobian term is 1.

It only remains to calculate the acceptance probability corresponding to the second stage given by equation 5.10. To propose a new birth, we use a rudimentary form of learning in which we avoid proposing peaks of adjacent pixels previously rejected in the first stage. In

a similar way, if a peak has been proposed to be removed in a death move and this move has been rejected, in the second stage we propose to remove one of the remaining peaks. The complete expression for the second-stage acceptance probability for a birth move is $\alpha_1(k^i, x^i, k^i + 1, z^i) = min\{1, A\}$ where $A$ is

$$A = \frac{\pi(k^i + 1, z^i) \cdot d(k^i + 1) \cdot (1/(k^i + 1)) \cdot (1/(k^i))}{\pi(k^i, x^i) \cdot b(k^i) \cdot Q_a(u_{1a}) \cdot Q_p(u_{1p}) \cdot Q_a(u_{2a}|u_{2a} \neq u_{1a}) \cdot Q_p(u_{2p}|u_{2p} \neq u_{1p})} \times \frac{\{1 - \alpha_1(z, xx^*)\}}{\{1 - \alpha_1(x, xx)\}} \tag{5.20}$$

and $u_{2a}$ and $u_{2p}$ are respectively the amplitude and the position of the peak. Note that the Jacobian term is again equal to 1.

## 5.6 Experimental Results

In this section, we shall illustrate the performance of the Bayesian model with spatial constraints developed on both real and synthetic data set. All the results displayed correspond to runs of 200 sweeps of the algorithm after a burn in of 800 sweeps. The common constants to the previous approach without spatial constraints are set with the values previously specified whereas the new unspecified constants are set as: $\sigma_i = 5$, $\sigma_1 = 0.01$ and $\sigma_b = 0.1$. As it was commented in previous sections, the parameters $\sigma_1$ and $\sigma_b$ represent a small random perturbation of the parameters of a peak of a neighbouring pixel. For that reason, small values of these parameters are envisaged. The value of the parameter $\sigma_i$ has been chosen taking into account preliminary experiments. Different values are considered for the smoothing parameter $\psi$ to observe the sensitivity of the results to variations of this parameter. The maximum number of peaks is set to be equal to 5 for the experiments envisaged. The number of peaks is inferred as in section 4.7, i.e. that corresponding to the highest marginal posterior probability, $p(k|\boldsymbol{y})$. Once we estimate the number of peaks, we extract estimates of the parameters from $p(\phi|k, y)$ by setting $k$ equal to the estimated number of peaks. The values of the parameters are estimated as the mean values of the samples that correspond to such a parameter subspace. The different pixels are accessed in a random fashion instead of using a raster scan of image array in order to avoid a bias of results.

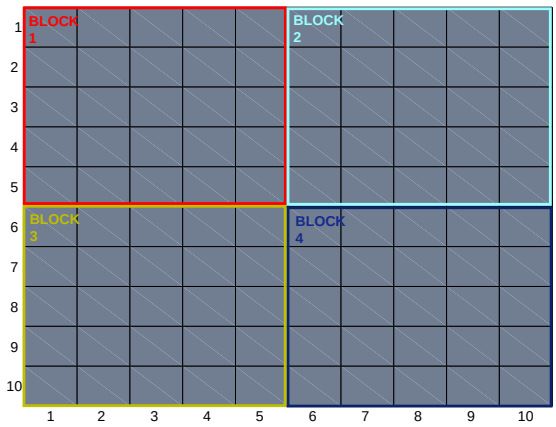



**Figure 5.3:** Decomposition in blocks of the simulated array for illustrative purposes

### 5.6.1 Synthetic Data

In this section we report some simulation results that compare some aspects of the performance of the spatial Bayesian model developed in this chapter to that of chapter 4. We use several criteria for comparison of this model as in [66]. For simulated data, we know the ground truth of the different parameters of the mixture distribution used and it is denoted as $F_{real}$. For each pixel $i$, we calculate the posterior mean squared error (MSE), $MSE_i = E((F^i_{real} - F^i_{estimated})^2|\boldsymbol{y})$ where $F_{estimated}$ is the estimate of the mixture distribution we have obtained. To summarise the performance over the whole array, we compute the root averaged MSE (RAMSE) defined as $RAMSE = (\sum_i MSE_i/N)^{1/2}$. Further, we have computed the deviance information criterion (DIC) described in [132] to compare both models using an approximate decision-theoretic justification. The DIC is composed of two terms: the posterior expected deviance, $E(D|y)$, and a penalty term $p_D$. $p_D$ can be interpreted as the effective number of parameters in the model. In our case, $D$ is the Poisson deviance and is given by $D = 2\sum_i(\boldsymbol{y}^i log(y/F_{real}) - (\boldsymbol{y}^i + F_{real}))$ as explained in [98]. $p_D$ is the difference between the posterior expectation of the deviance and the deviance at the posterior mean of the parameters and represents the model complexity. $E(D|y)$ represents the goodness of fit. Further information on this criterion can be found in the references previously mentioned.

Figure 5.4 displays an array of simulated histograms corresponding to block 1 of figure 5.3.

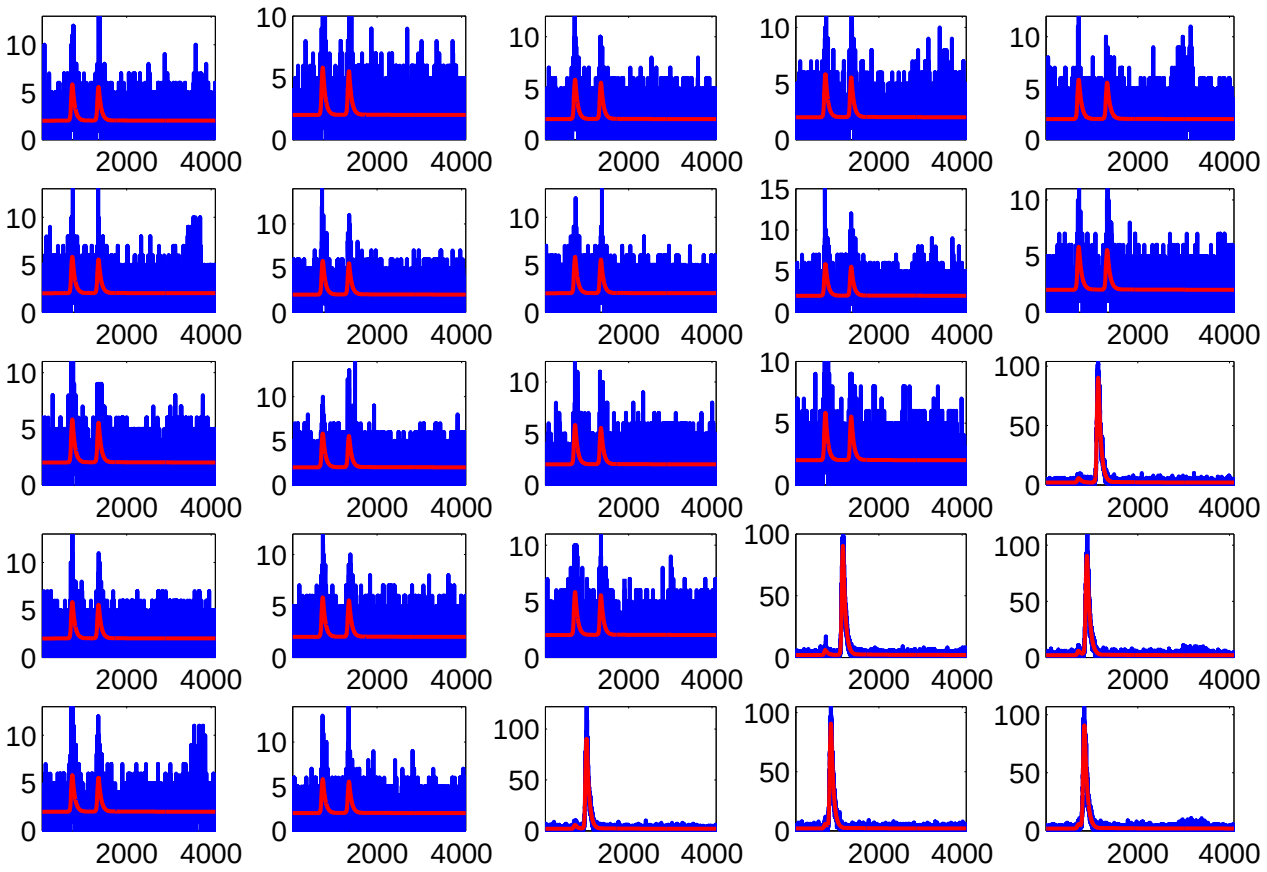
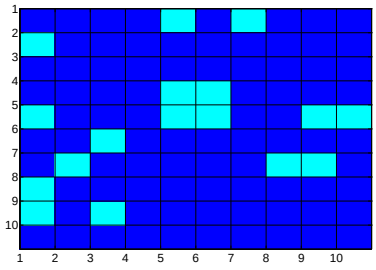


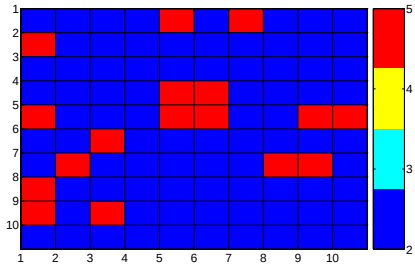

**Figure 5.4:** Actual values of the mixture distribution (red) and noisy histogram (blue) with clutter corresponding image pixels with coordinates $rows \in \{1, 2, 3, 4, 5\}$ and $columns \in \{1, 2, 3, 4, 5\}$

The remaining blocks are shown in appendix A. The array of histograms is divided into different block for illustrative purposes only. In this array, each pixel has 2 different returns whose parameters are known. We have also added clutter to some pixels chosen at random, the location of this clutter is also random. For instance, this clutter can be observed in the pixel located at row 1 and column 5. Figure 5.5 shows a colour-coded representation of the estimated number of peaks when no spatial constraints or different values of the parameter $\psi$ are used. Figure 5.5(a) indicates in which pixels we have added the clutter (light blue). It can be observed, when no spatial constraints are used, the algorithm determines that the clutter is either a return or multiple returns together. As the value of $\psi$ increases, there is an effective "smoothing" on the number of peaks and all the pixels tend to have only two peaks. However, inferring the current number of peaks does not mean that we have detected the current peaks. In figure 5.6 we display a scatter plot of the channel position against the intensity value for the actual simulated value, and the values obtained with $\psi = 1$ and $\psi = 10$. Besides of variations in the estimated values, it can be observed that some of the points accumulate between channel positions 2500 and 3000, that is, the area where the clutter was generated. As can be observed, one of the points corresponds to the case where $\psi = 20$ which means that one clutter return was preferred against the true simulated peak. The fitting of this outlier can be seen in the pixel of the right bottom corner in figure A.5. In this case, we have a border effect in which the pixels in the boundaries of the image have less information from the neighbouring pixels which increases the uncertainty associated with estimates in these pixels(see figure A.1). The remaining block structures of the pixels and the estimated values are shown in appendix A.

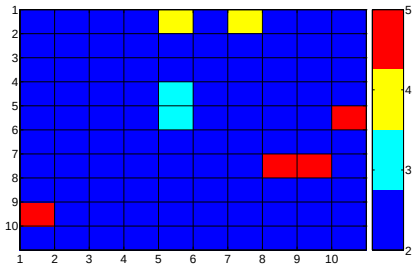
The models with and without spatial constraints are compared on the basis of RAMSE, DIC, $E(D|y)$ and $p_D$ for different values of the parameter $\psi$. The different results for these criteria are shown in table 5.1. The table shows that the spatial model with $\psi = 10$ gives more faithful estimation of the underlying peaks in terms of RAMSE. However, the variation of RAMSE between models with spatial constraints is so small that it seems that all are perfectly plausible. It seems that averaging over the whole array corrects the estimated error calculated by RAMSE since only 17 pixels present clutter against the remaining 83 which the algorithm (in both versions with and without spatial constraints)



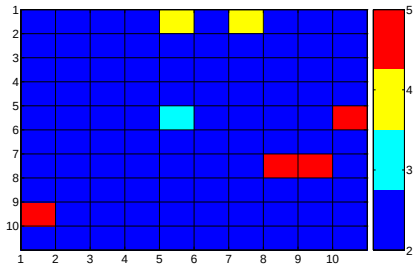
(a) True position of the clutter (ground truth)



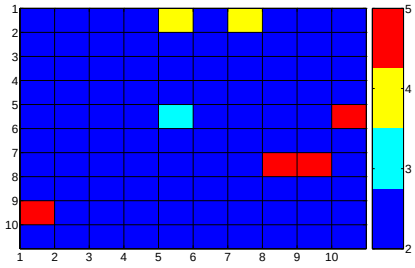
(b) No spatial constraints



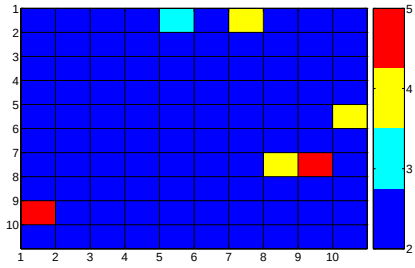
(c) $\psi = 0.25$



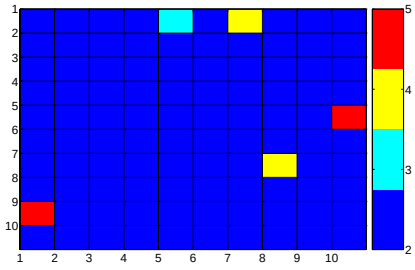
(d) $\psi = 0.5$



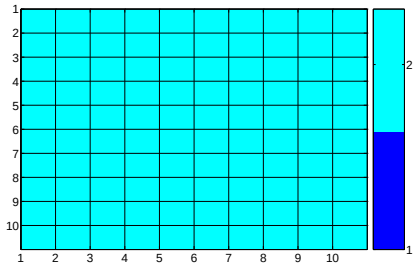
(e) $\psi = 0.75$



(f) $\psi = 1$



(g) $\psi = 10$



(h) $\psi = 20$

**Figure 5.5:** Estimated number of peaks with and without spatial constraints from simulated data (whole array)

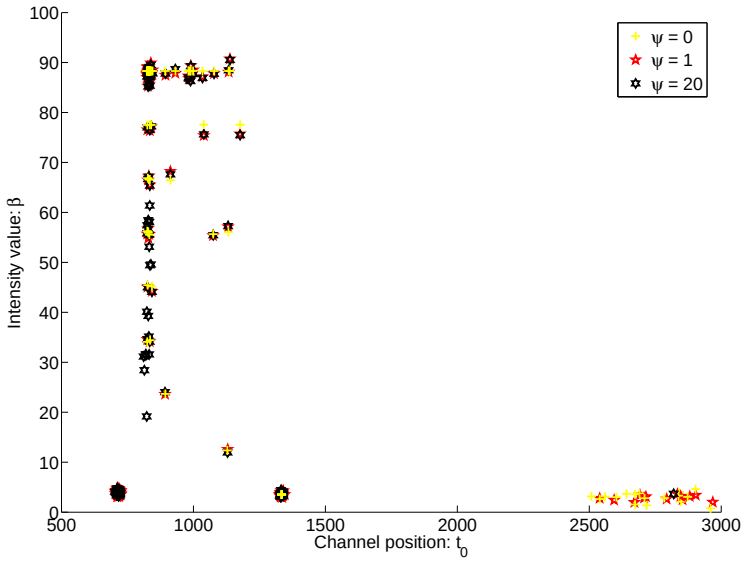



**Figure 5.6:** Scatter plot representing channel position against intensity value. The actual simulated values and those obtained for the parameters $\psi = 1$ and $\psi = 20$ are represented

estimates accurately. If one accepts the DIC principle as the most adequate model measure since it is applicable to real data and no ground truth is needed, then the model without spatial constraints is the one which best explains the data. A similar result is obtained for $E(D|y)$ whereas $p_D$ indicates that the best model is that with $\psi = 1$. These results are consistent since information criteria rely heavily on the data and do not make any distinction on whether or not the return detected is false. They favour the model which best describes the data.

**Table 5.1:** Simulation results comparing RJMCMC with and without spatial constraints.

| Smoothing parameter | RAMSE | DIC | $E(D\|y)$ | $p_D$ |
|---|---|---|---|---|
| No constraints | 0.316 | 463976.054 | 462727.755 | 1248.299 |
| 0.25 | 0.21 | 469304.319 | 468897.615 | 406.704 |
| 0.5 | 0.205 | 469647.117 | 469154.308 | 492.809 |
| 0.75 | 0.21 | 469760.579 | 469228.735 | 531.844 |
| 1 | 0.202 | 469361.138 | 469290.94 | 70.195 |
| 10 | 0.201 | 470330.748 | 469912.312 | 418.436 |
| 20 | 0.222 | 474119.728 | 473731.597 | 388.131 |

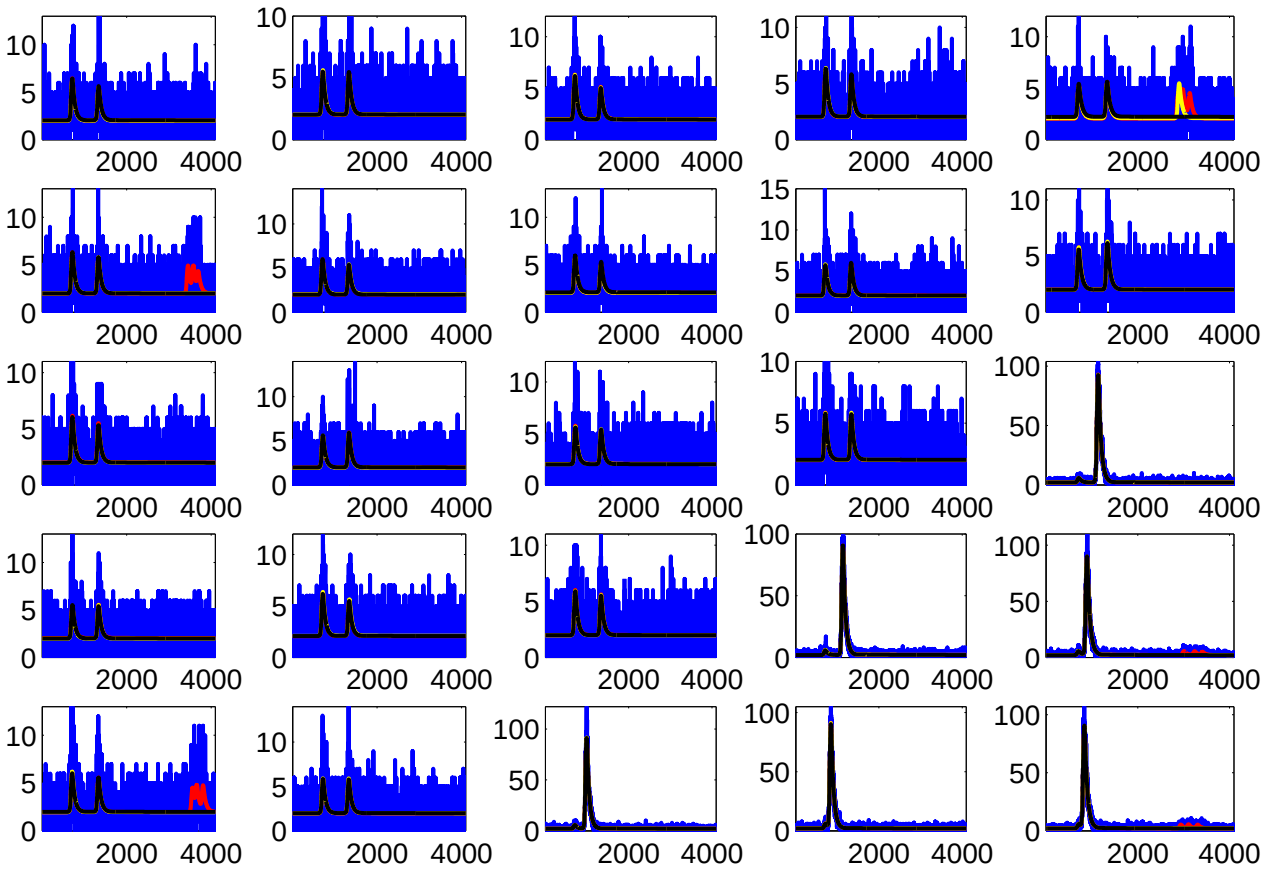



**Figure 5.7:** Array of photon count histograms (in blue) corresponding to block 1 of figure 5.3 and final fit without spatial constraints(red), with $\psi = 1$ (yellow) and $\psi = 20$ (black).

### 5.6.2 TCSPC and BIL Data

In this experiment, we analyse TCSPC and BIL data that correspond to a pixel array of 50 by 50 histograms of photon counts and reflectance/intensity values respectively. The TCSPC histograms come from the structure shown in figure 5.8(a). The area analysed corresponds to that shown in the reflectance image of figure 5.8(b). These data have been analysed with the following values for the parameter $\psi = \{0.25, 0.5, 0.75, 1, 2, 5, 10\}$ of the Potts model defined by equation 5.5. The maximum number of peaks in this experiment was set to 10 and the initial number of peaks was selected randomly. Figures 5.9 and 5.10 display the estimated number of peaks for each value of the smoothing parameter using a colour-coded image. As can be seen from the different figures, as the smoothing parameter increases its value the algorithm tends to form clusters, that is, areas with the same number of estimated peaks are connected by paths from neighbour to neighbour. Having seen the scene, initially, one may think that the number of peaks should be 2, one coming from the semitransparent surface of the toy house and the other coming either from the background or the toy cat. However, this is not the case as shown in the different figures. The "extra" number of peaks is possible due to the footprint of the beam impinging on a target with surfaces distributed in depth. Some examples of multilayered images are shown in figure 5.11 where all the returns are represented. The different surfaces are not perceptible with this representation due to difficulties with Matlab. When the smoothing parameter is equal to 10 (see figure 5.10(d)), the spatial interactions are so strong that the algorithm is not able to "kill" some of the unwanted returns. In this case, the posterior distribution is dominated by the prior information so the death move does not help to eliminate these peaks.

Figure 5.8(c) shows the toy cat and its depth variation using a false color image whereas figure 5.8(d) shows a depth profile of the previous image corresponding to row 30 where we can see that the algorithm captures the curvature of the object. This profile is less smooth than that of figure 4.14(c). However, the reason for this is the different size of the array (50 by 50 against 100 by 100) and the complexity of the experiment which requires the laser by focused on to the scene in a sub-optimal way. Figures 5.8(e) and 5.8(f) show a mesh reconstruction of the toy cat ($\psi = 5$) and a three layer representation of the complete

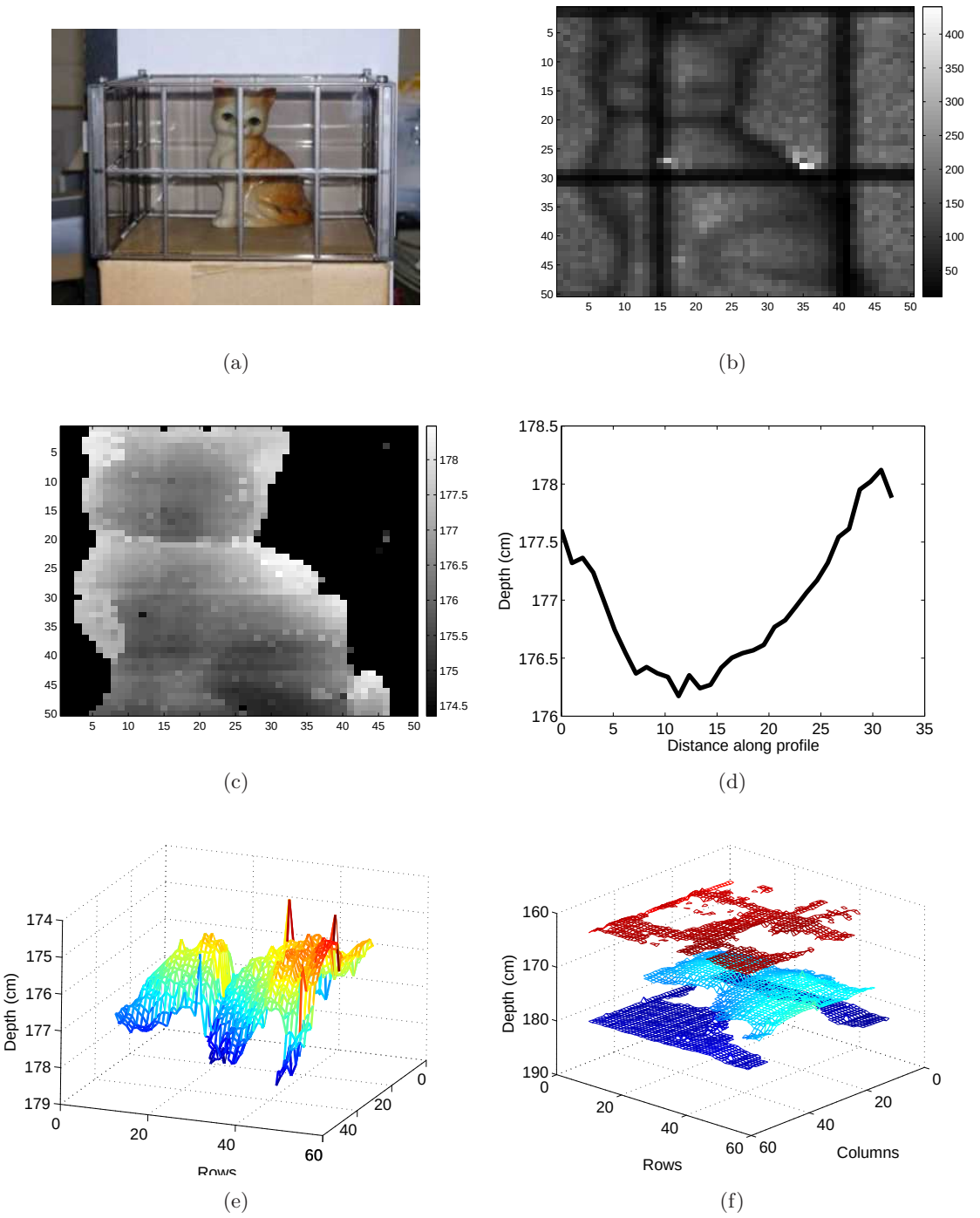
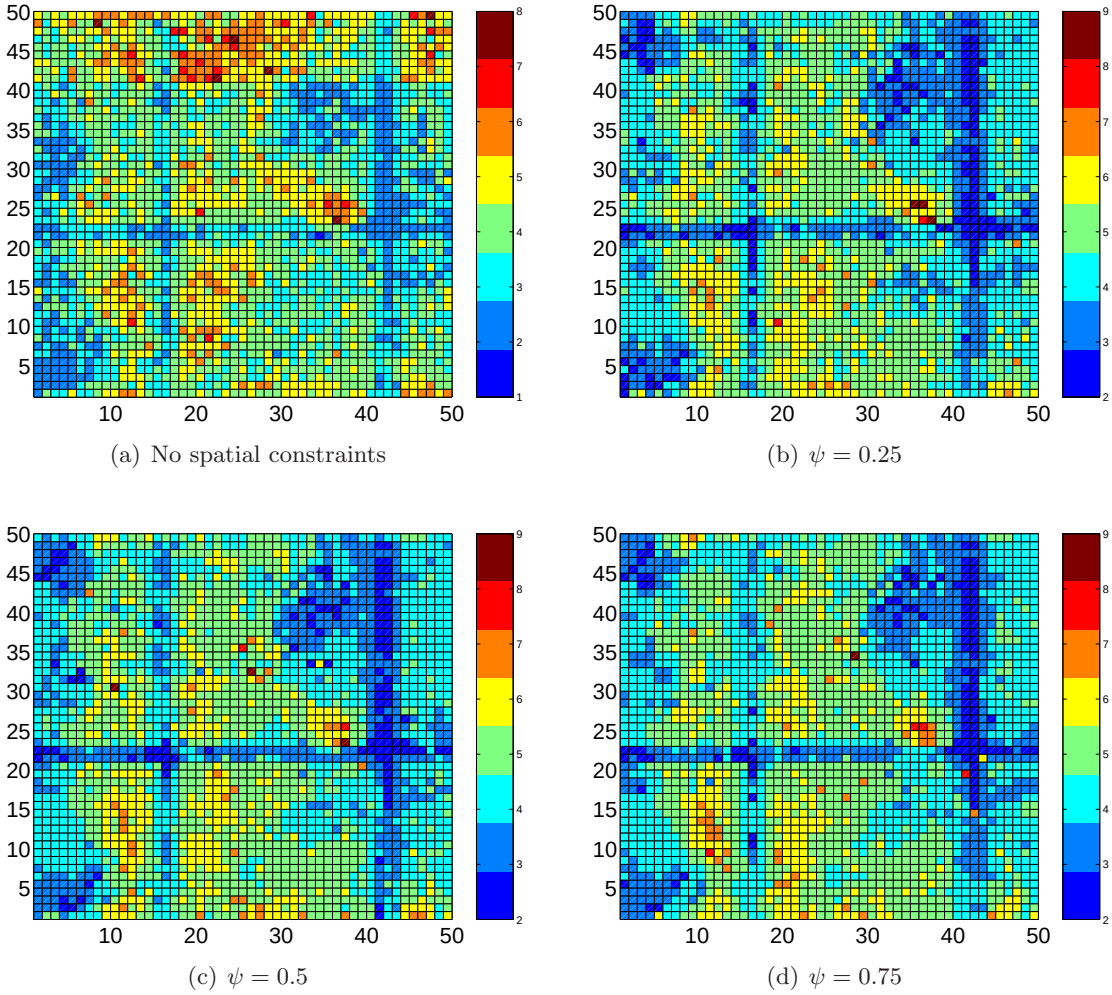



**Figure 5.8:** (a) Image of complex multiple return structure composed of a toy cat and a semi-transparent toy greenhouse (b) Reflectance image obtained with TCSPC system, the unit is photon counts (c) False color depth image of the toy cat (d) Depth profile of the toy cat corresponding to row 30 (e) Mesh of the toy cat (f) Multiple layer structure of the target

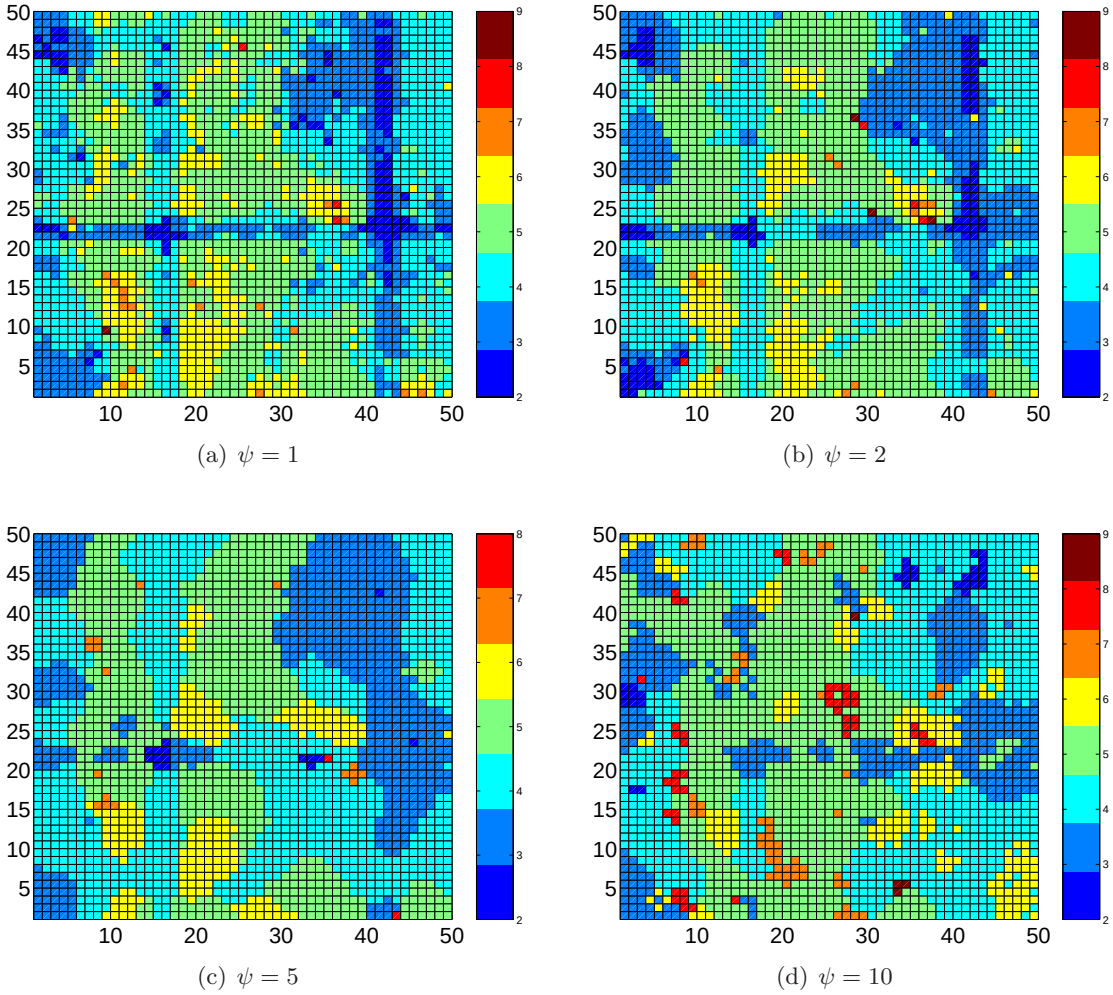


(a) No spatial constraints (b) $\psi = 0.25$

(c) $\psi = 0.5$ (d) $\psi = 0.75$

**Figure 5.9:** Estimated number of peaks of the TCSPC data using RJMCMC with and without spatial constraints

target where only 3 returns have been used to facilitate the display. The first layer (closer to the detector) represents the first glass of the toy greenhouse. Note, that all the values are not available. This is due to three main factors: First the specularity of the surface is very poor (semitransparent surface). Second, the returns are well below the background noise. The acquisition time considered was very short and the incidence of the laser was not normal. Third, the prior on $k$ is "oversmoothing" the number of peaks. The second layer corresponds to the toy cat whereas the third layer corresponds to the second glass of the toy greenhouse. The distance between the pair of glass surfaces is estimated to be about $18.0cm$ which is consistent with the dimensions of the toy greenhouse.

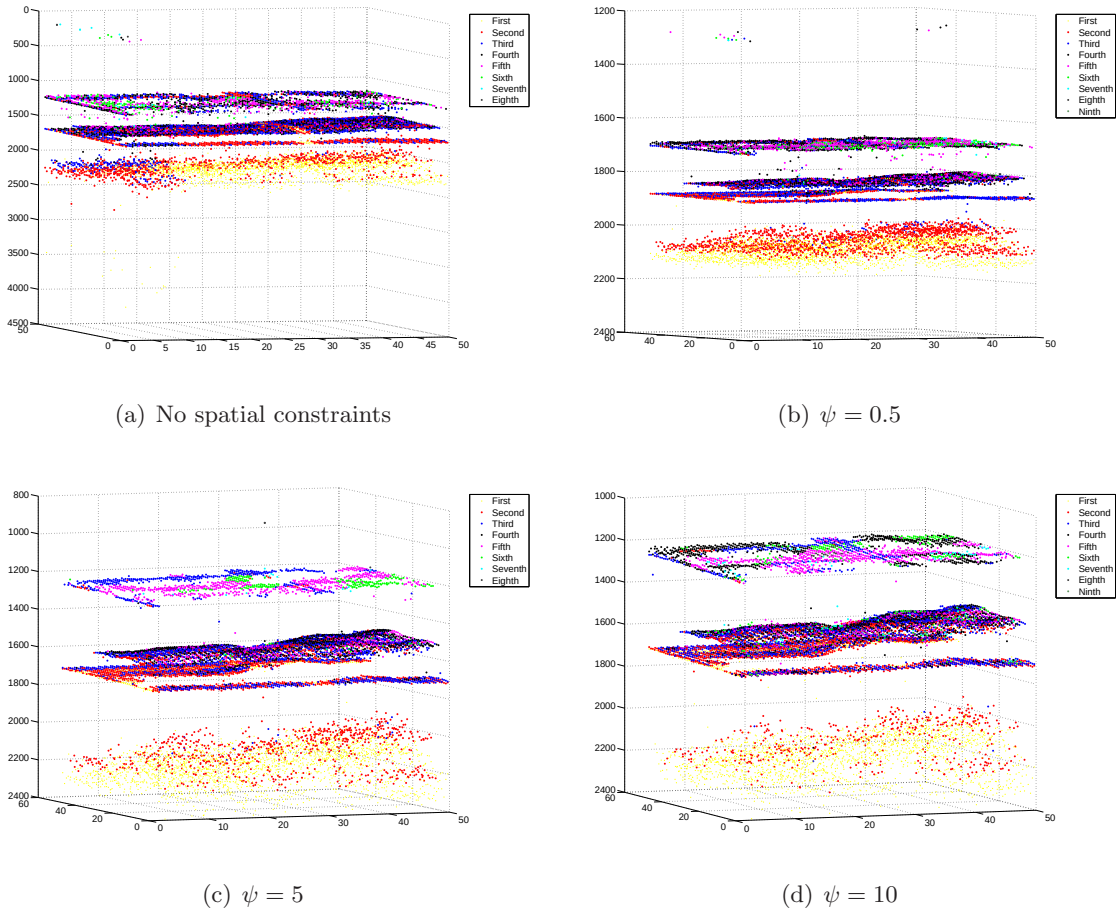


(a) $\psi = 1$ (b) $\psi = 2$

(c) $\psi = 5$ (d) $\psi = 10$

**Figure 5.10:** Estimated number of peaks of the TCSPC data using RJMCMC with spatial constraints

The performance of the model is also illustrated on BIL data. The data set chosen is that corresponding to the trig point shown in figure 4.22. This data presents only one unique return as can be seen in figure 5.13. Thus, we have tested this data under two different assumptions to better understand the behaviour of the method. On one hand, only one peak or background was considered. The spatial information was incorporated through the proposal distributions. On the other hand, we allowed up to 5 possible returns in our calculations. Figures 5.14(a) and 5.14(b) display the estimated number of peaks when only background or 1 peak is allowed with a smoothing parameter of 0 and 0.5 respectively. Although both figures look similar, it can be appreciated that figure 5.14(b) helps to "smooth" the number of peaks in some places where previously it was considered




(a) No spatial constraints

(b) $\psi = 0.5$

(c) $\psi = 5$

(d) $\psi = 10$

**Figure 5.11:** Multiple-layer structure of the experiment of figure 5.8(a) with and without spatial constraints

there was only background. If the smoothing parameter is increased, the results obtained are similar to those obtained with $\psi = 0.5$. Figure 5.15(a) shows a depth reconstruction of the trig. point whereas figure 5.15(b) shows the reflectance image obtained (amplitude values of the mixture). Figure 5.14(c) and 5.14(d) displays the estimated number of peaks when considering up to 5 peaks, and $\psi = 10$ and $\psi = 30$ respectively. In the first case, it can be observed that 3 different clusters have been created, one corresponding to only background, another to 1 return and the last one to 2 peaks. Since we are using a look-up table as the instrumental function to which we compare the returns, this does not take into account different effects such as the atmospheric turbulence and speckle. One peak is not enough and the algorithm needs at least two peaks to explain the data as reflected in the posteriors. In addition to this problem, when considering $\psi = 30$, a small portion of the image corresponding to the foreground is considered as background. As in the TCSPC

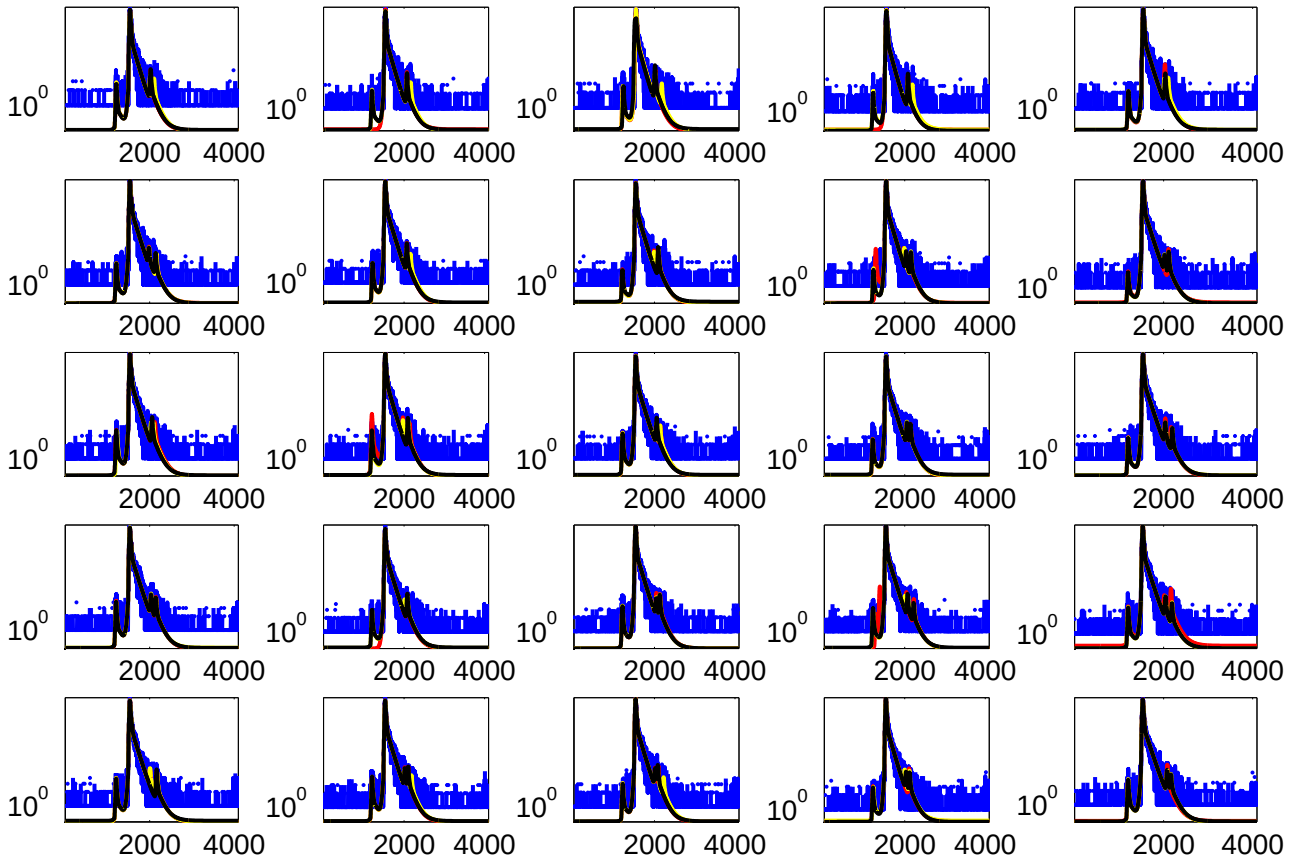



**Figure 5.12:** Array of photon count histograms (in blue) of the structure of figure 5.8(a) and final fit without spatial constraints(red), with $\psi = 1$ (yellow) and $\psi = 5$ (black). The axis are represented in log scale for convenience

case, the penalty associated prior information is "stronger" than the likelihood function and the peaks cannot be added or removed.

## 5.7 Conclusions

We have described the development and application of RJMCMC techniques incorporating spatial contextual information to process time-of-flight data using both single photon counting and variable temporal gating to extract range measurements. This spatial information was incorporated on the prior distribution through a Markov Random Field on the number of peaks and on the proposal distribution through the development of two new moves: the spatial mode jumping and the spatial birth/death processes. Our results show that it is possible to resolve multiple returns and hence characterise objects distributed in 3D space and to extract the underlying spatial structure in the data processed. This spatial information can be used to reduce the influence of clutter of unknown origin by eliminating possible false returns. Furthermore, we have described a delayed-rejection stage which allows us to perform rudimentary learning and therefore improve the performance of this stage.

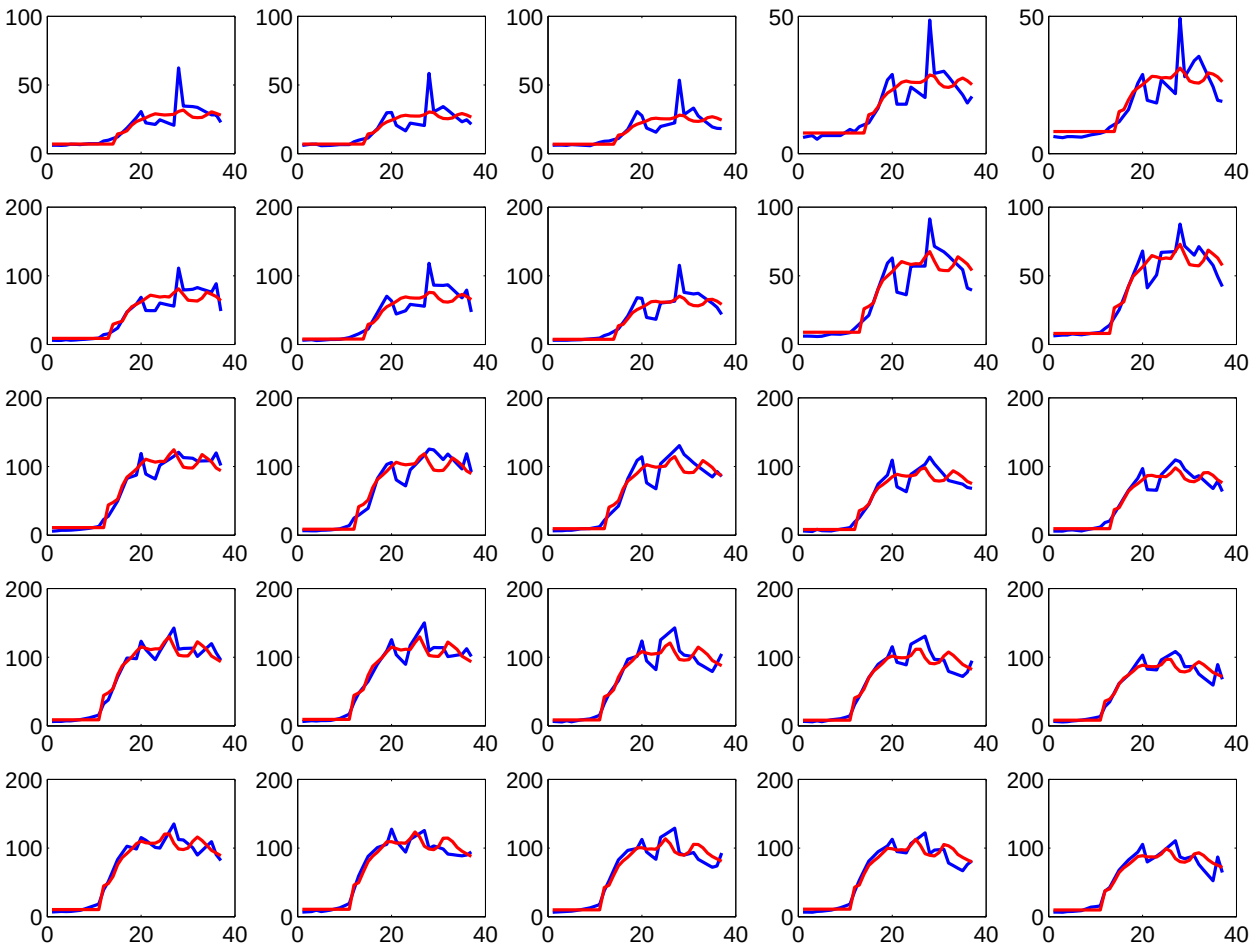



**Figure 5.13:** Array of intensity/reflectance histograms (in blue) of the trig point and final fit using the proposed model. Only 25 pixels from an image of 2500 are represented. The vertical axis represent intensity values whereas the horizontal axis represents non-normalised temporal channels

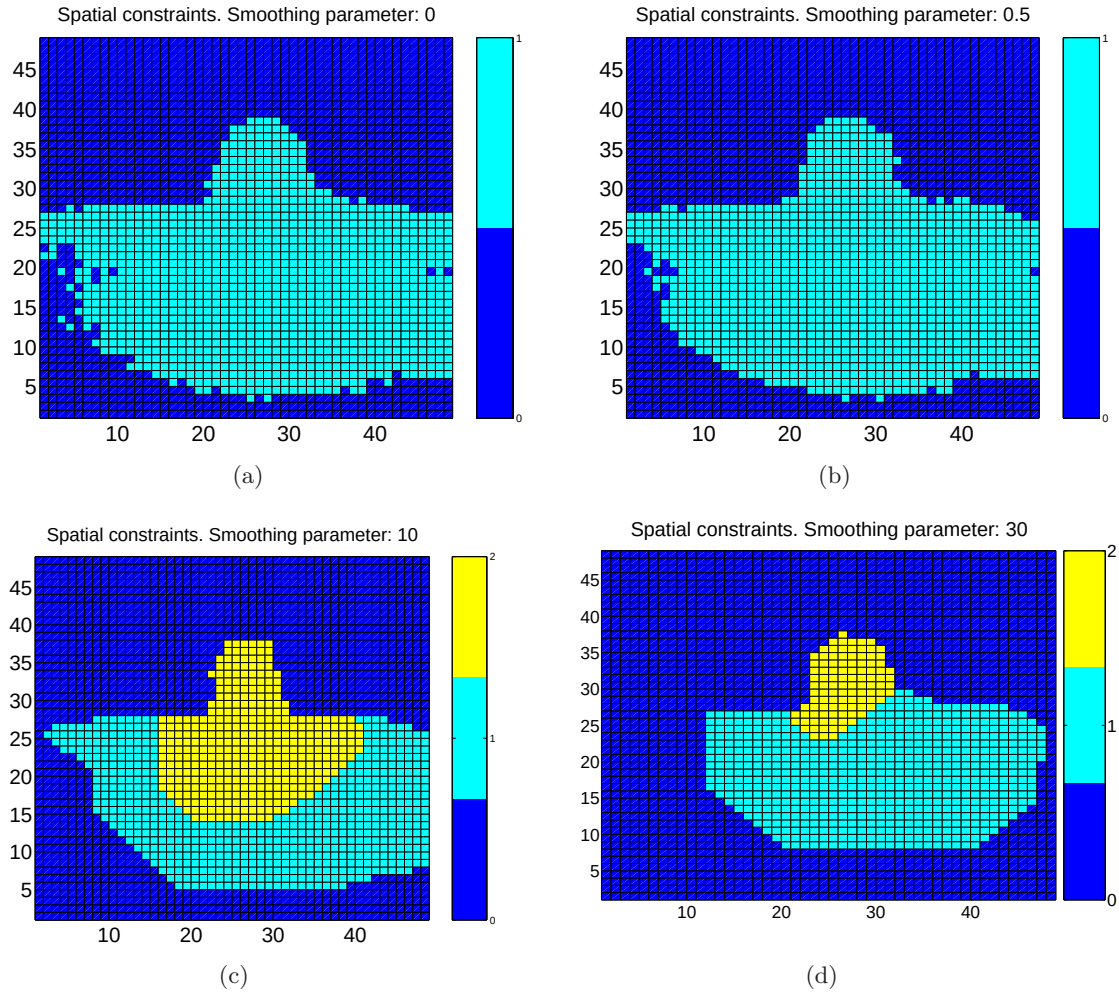
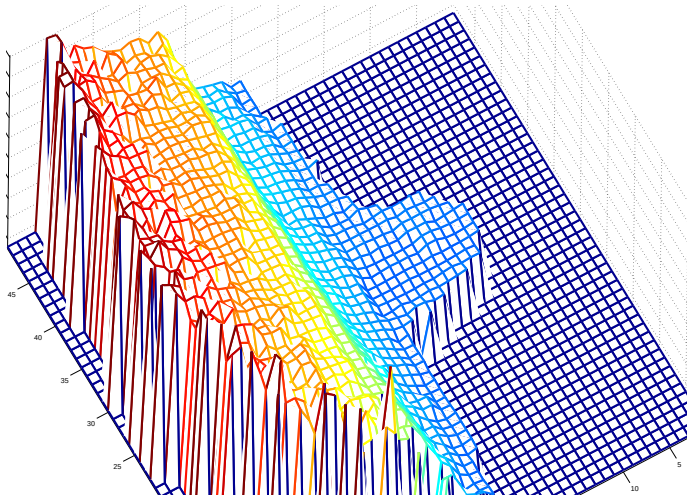
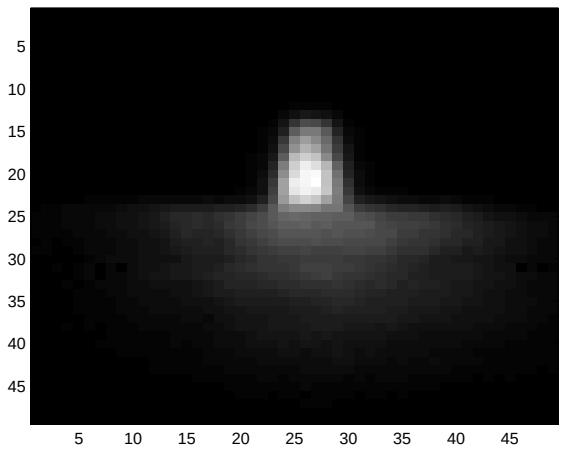



**Figure 5.14:** Estimated number of peaks of the BIL data of the trig point of figure 4.22(b).




(a) Mesh of a trig point




(b) Reflectance image of the trig point

**Figure 5.15:** Details from depth and reflectance images of a trig point (distance $\sim 6.6km$) obtained with a BIL system

Chapter 6

# Conclusions and Further Work

*Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.*

*Winston Churchill.*

## 6.1 Summary of Main Findings

The work described in this thesis has been concerned with the development of a stochastic Bayesian framework for the analysis of lidar/ladar data obtained from two different pulsed laser systems (TCSPC and BIL). Particular emphasis has been placed on cases where the returned signal is weak or contains multiple returns, due to the spread of the laser beam or due to reflection from multiple, axial surfaces. The goal of this thesis was to improve the accuracy of measurement of such systems and to provide a complete classification as follows: 1) the number of scatterers present; 2) the amplitude of each of the signals; 3) the positions of each of the signals; and 4) the background level; using a maximally automatic technique.

The Bayesian approach employed allows incorporation of any prior knowledge on the number of scatterers, their relative size and positions in the analysis. Moreover, the posterior density gives an indication of the certainty with which these can be estimated after observing the data. Finally, by using the posterior as a prior for analysis of subsequent data it is possible to amalgamate observations in a coherent way. With that aim, several prior models have been proposed that allow us to detect very weak returns hidden in the background noise. The choice of prior distributions in our approach is difficult since some parameter ranges are unbounded, which makes the results very sensitive to this choice. Two physical models which represent the likelihood of our system have been presented: a

parametric model in the TCSPC case and a look-up table in the BIL case. In both cases, the posterior distributions have been properly explored by using reversible jump Markov chain Monte Carlo techniques to determine the parameters of the model and to assess the unknown number of scatterers.

Throughout this thesis, we have considered a range of scenarios to determine the advantages of the Bayesian approach. Initially, we considered each pixel or scanned element in isolation, i.e. we did not examine the spatial dependencies among adjacent pixels. Then, we assumed the existence of only one scatterer in the analysis of the return signal which allowed us to assess the convergence and mixing behaviour of the Markov chains generated and to choose appropriate proposal distributions. Subsequently, we allowed the possibility of an unknown number of scatterers. The method was shown to be effective and robust in detecting multiple returns from noisy data, as shown by some performance measurements to determine the accuracy of the algorithm. We also performed comparisons of our method with other state-of-the-art algorithms through the use of both synthetic and real data. The algorithm proposed here showed superior accuracy in determination of the number of scatterers, their amplitudes and times of arrival. Moreover, the Bayesian approach allowed us to quantify the uncertainty in these parameters. We also acquired 3D images of a number of test objects of varying size at different distances to show the capability of the algorithm to create 3D reconstructions. Finally, we extended the algorithm to cases in which we considered spatial correlations between neighbouring pixels. To that end, we used a MRF, concretely a Potts model, as a prior distribution of the number of peaks. This prior favours patterns in which neighbouring pixels have the same number of peaks. Further, we proposed two new moves, "spatial mode jumping" and a "spatial birth/death process", which included spatial information from neighbouring pixels. We also made use of a delayed rejection step in fixed and variable dimension settings to improve the mixing of the chain. In the spatial case, the delayed rejection step used a rudimentary form of learning which allowed us to discard proposed values previously rejected in the first stage.

## 6.2 Further Work

There is much further analysis that could be done on data of the type presented in this thesis. Several assumptions have been made that can be modified to improve the results obtained. For instance, one assumption is that the "underlying shape" of peaks in the photon return signal is independent of the peak amplitude. Despite the good results shown in previous chapters, the process of fitting the reference signal presents non-linearities. As the amplitude of the peak decreases, the shape of the returns changes slightly and therefore the accuracy of the measurements is affected. On the other hand, the shape parameters from the reference signal have been obtained using a constrained MLE process. An extension of the algorithm should include the estimation of this parameters using a MCMC algorithm. Some initial work has been carried out regarding to this issue as shown in appendix B, but more has to be done to produce an effective algorithm. Indeed, although the formulation of the problem in a Bayesian framework is not difficult once the constraints are given in a form of appropriate prior distributions, the algorithm seems to fail to find an appropriate shape for the return. This is due to the long third exponential tail given by equation 4.1 which contributes more to the log-likelihood term than the Gaussian shape of the return. One possible solution to this problem might be to specify prior distributions for each term of the four-piecewise exponential in such a way to weight them according to their "importance" with respect to the other terms. The results obtained from this modification may be used to develop a new fuller Bayesian approach in which all the parameters of the model are estimated. This would allow a more accurate description and interpretation of the model and to consider the effects on the return signal as it propagates through the atmosphere. A further analysis should also consider the speckle and turbulence effects which can influence lidar measurements from a diffuse hard target.

In the case of BIL data, a further improvement is the identification of a suitable parametric model for the expected return which will improve the accuracy of the estimates performed by the current algorithm. Further, the parametric model might include atmosphere and speckle effects as in the TCSPC case. If the parametric model designed is close to the real behaviour of the system, a BIL data simulator may be constructed. This simulator should allow the generation of ground truth data which can be used to compare the different

algorithms used to estimate depth in the BIL framework. Some recent work has been done on BIL data simulators (see Andersson's work [5]). However the physics model used is very simple as the noise is assumed to be normally distributed, not considering sensor and data-dependent noise.

We might also consider continuous estimation of range data using standard Kalman filter or sequential Monte Carlo methods such as particle filters as mentioned in chapter 2 and compare its performance to the RJMCMC algorithm developed in this thesis. In the extreme case, the first received photon provides an initial estimate of the target position to an accuracy of one bin width. As each received photon is recorded the estimate of both position and error is updated until a suitable convergence is found. This should speed up the process of classification adaptively, for example, data from good reflectors requires less time to process than poor data. We anticipate that these algorithms will require substantial modification to deal with the problem of an unknown number of components and to assess uncertainty since sequential Monte Carlo methods only give single estimates.

As explained in chapter 2, lidar systems normally use a single wavelength, coherent source; however, if several laser sources at different wavelengths are used [153] this allows us to combine distance measurements with an assessment of the surface colour signature, which augments the limited reflectance information that can be obtained by integrating a single laser wavelength return and may be more indicative of the nature of the target. For instance, the analysis of histograms where several wavelengths are present may help us to distinguish whether two returns in the observed histograms are due to two distinct opaque surfaces at different locations in a divergent beam, or to two co-axial responses from semi-transparent reflectors. To process this data, ideas can be borrowed from the analysis of multispectral data for scene classification. As the number of spectral bands used will remain fixed when processing the data, one may use a multivariate Gaussian MRF to establish spectral dependencies between the parameters of the different bands employed.

Further improvements to the analysis of spatial data can be performed. Although the RJMCMC moves used in this thesis make no reference to the normalizing constant, its computation is needed in cases where a fully Bayesian analysis of the data is required.

In general, the hyperparameters of a given model are selected according to some a-priori knowledge or estimated by the use of some training sets. In many cases, the values of the hyperparameters have to be re-calculated when using different data sets. The "smoothing parameter" of the Potts model introduced in chapter 4 is an example of this issue where the calculation of the normalising constant, i.e. the partition function, is very difficult. In the last few years, several attempts have been made to perform a truly Bayesian analysis, accounting for uncertainty in the hyperparameter values (see for instance [56, 81]). A similar approach might be performed to lidar data to implement an adaptive smoothing parameter. The main drawback of these approaches is the high computational cost involved.

The use of a multivariate MRFs for variable dimension settings is another approach which deserves future research. In chapter 5 we commented that a multivariate MRF could not be used in a variable dimension setting since we had a data association problem. Indeed, as the number of peaks in a pixel varies randomly to better explore the parameter space it is difficult to link the parameters of one pixel with its equivalent parameter in a neighbouring pixel. A way to solve this problem is to introduce a missing data formulation of the model, in which there is a bounding parameter which associates new peaks with existing peaks or disassociates peaks in the case of a death event. This will create a new layer in the hierarchical Bayes approach but it will allow us to create a more flexible multivariate MRF which can cope with variable dimension settings.

To date, all the analyses we have done are on images collected with fixation on each pixel, that is the sensor stares along a given axis in space, the data is collected, and then the sensor moves to the next position. With new scanning and data acquisition hardware, it should be possible to scan continuously through *(x,y)* or $z$-space and record returns from each single pulse in a space-time continuum. This would allow much faster acquisition and processing of range data, and conceivably could lead to adaptive scanning and processing. However, this would introduce many variations to the analyses presented here to cope with the continuous processing.

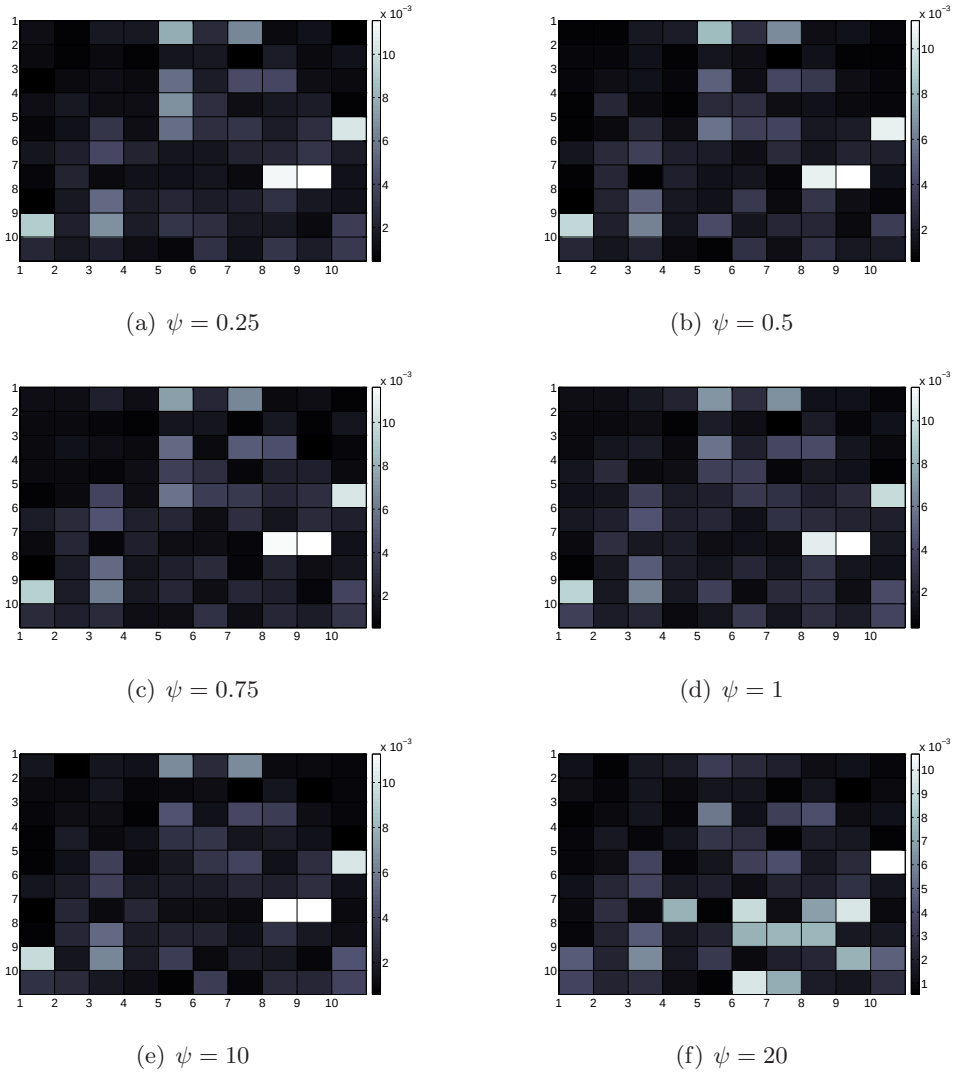
Finally, the posterior distributions obtained could be used as prior information, in a more complex Bayesian framework, for simultaneous detection, tracking and recognition of 3D

objects by including ideas of sequential Monte Carlo methods.

Appendix A

# Further Results

In this appendix, we show some figures corresponding to the analysis of simulated data of section 5.6.1. Figure A.1 displays for each pixel the root MSE between the simulated and estimated distributions of peaks. We also show the remaining blocks of 5 by 5 pixels used in the simulation of 5.6 as well as the results obtained when using our Bayesian approach with and without spatial constraints. Figure A.2 corresponds to block 2, figure A.3 to block 3 and figure A.4 to block 4.



(a) $\psi = 0.25$ (b) $\psi = 0.5$

(c) $\psi = 0.75$ (d) $\psi = 1$

(e) $\psi = 10$ (f) $\psi = 20$

**Figure A.1:** Root Mean Squared Errors for the simulated data

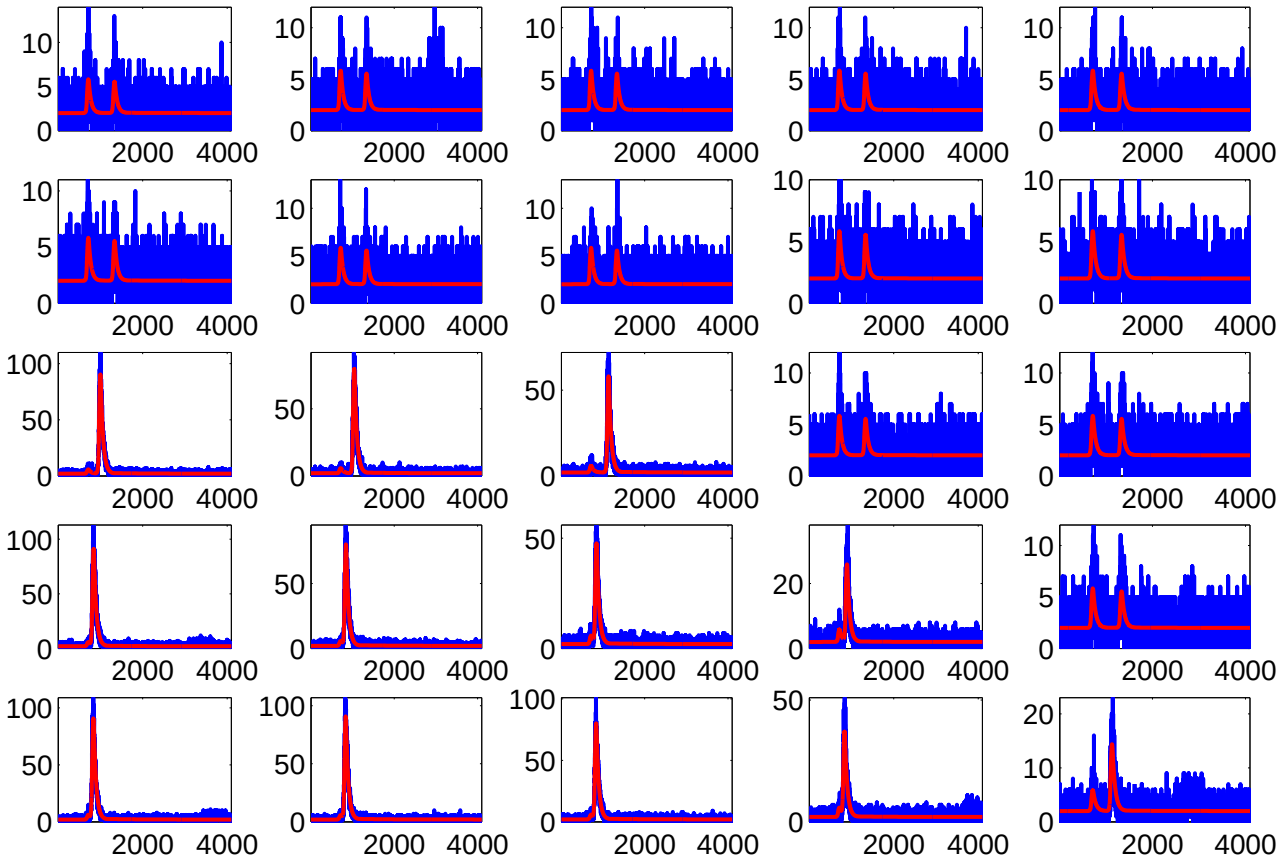



**Figure A.2:** Actual value of the mixture distribution (red) and noisy histogram (blue) with clutter corresponding image pixels with coordinates $rows \in \{1, 2, 3, 4, 5\}$ and $columns \in \{6, 7, 8, 9, 10\}$

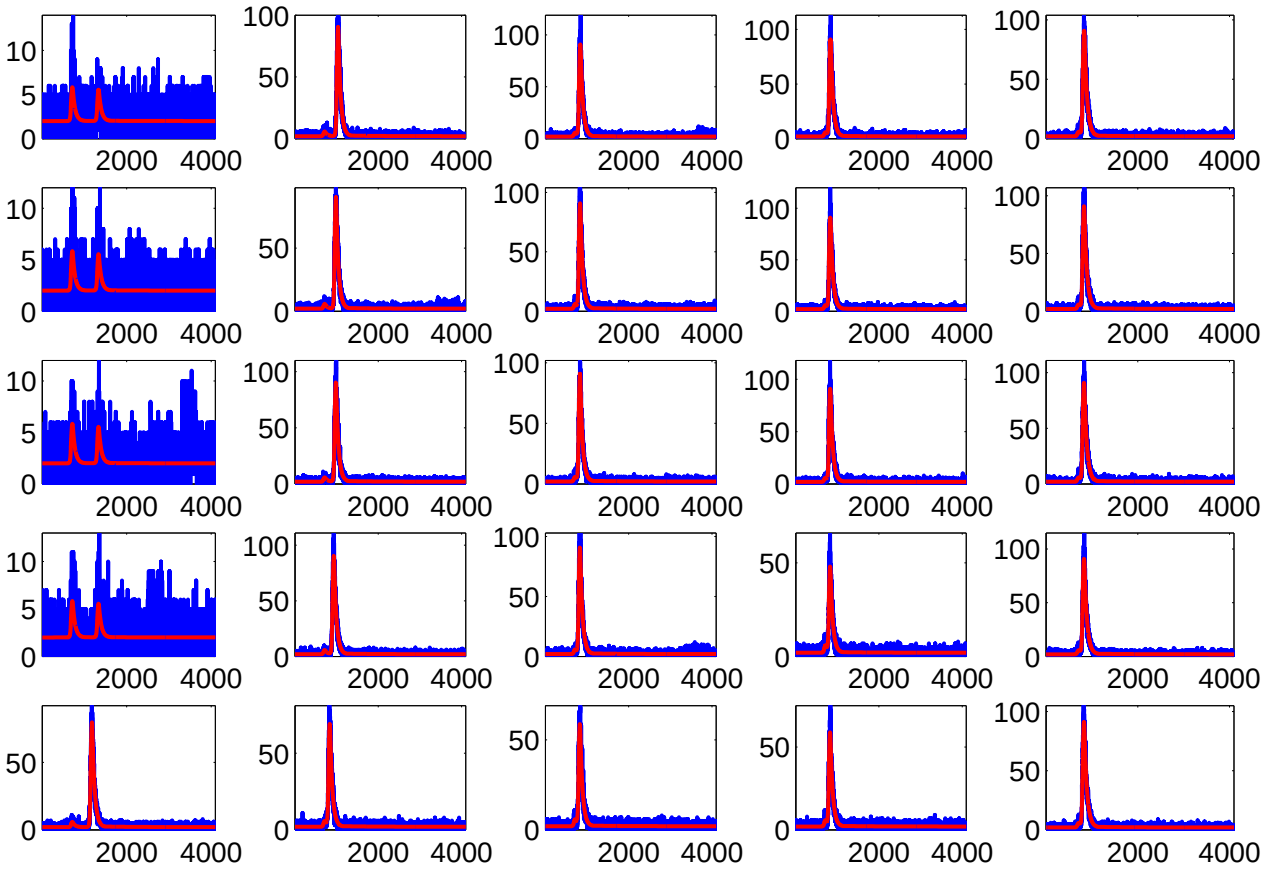



**Figure A.3:** Actual value of the mixture distribution (red) and noisy histogram (blue) with clutter corresponding image pixels with coordinates $rows \in \{6, 7, 8, 9, 10\}$ and $columns \in \{1, 2, 3, 4, 5\}$

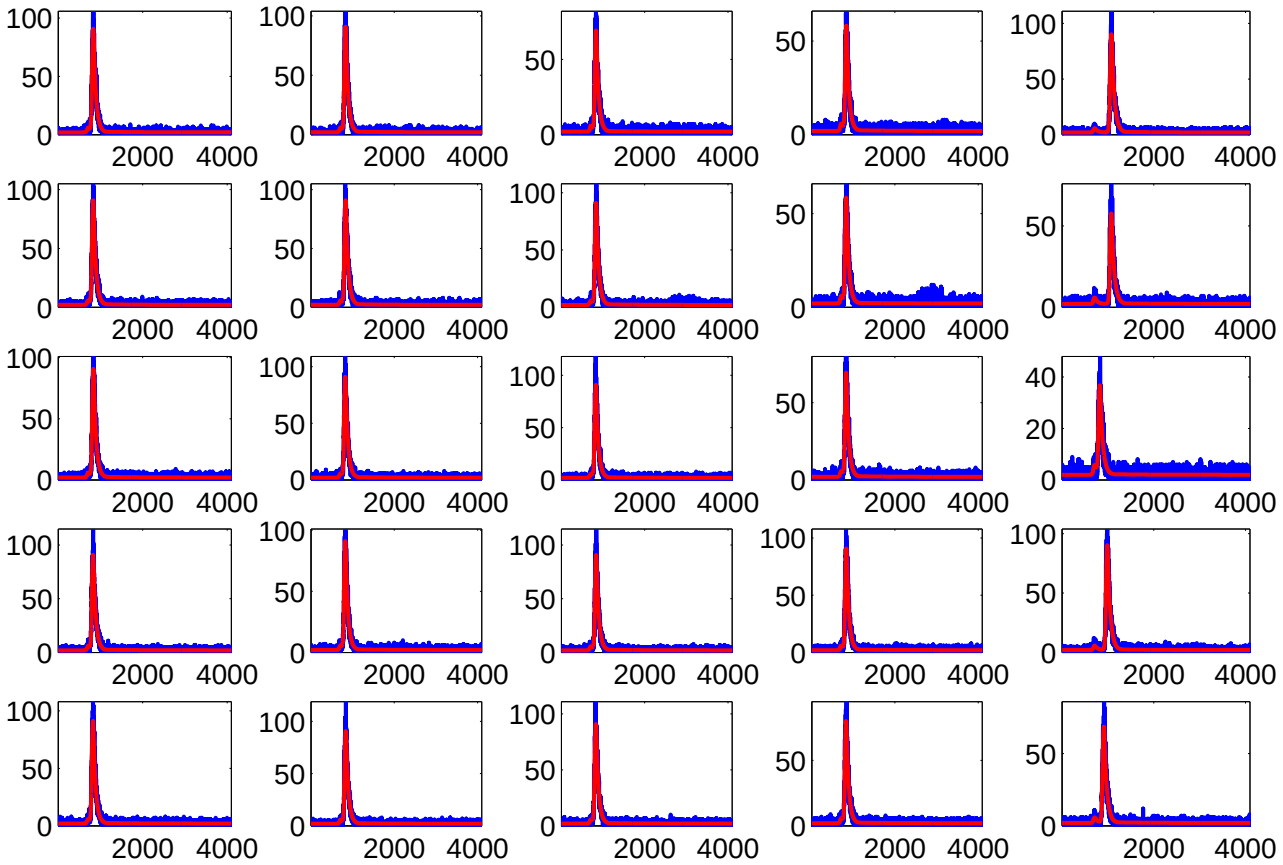



**Figure A.4:** Actual value of the mixture distribution (red) and noisy histogram (blue) with clutter corresponding image pixels with coordinates $rows \in \{6, 7, 8, 9, 10\}$ and $columns \in \{6, 7, 8, 9, 10\}$

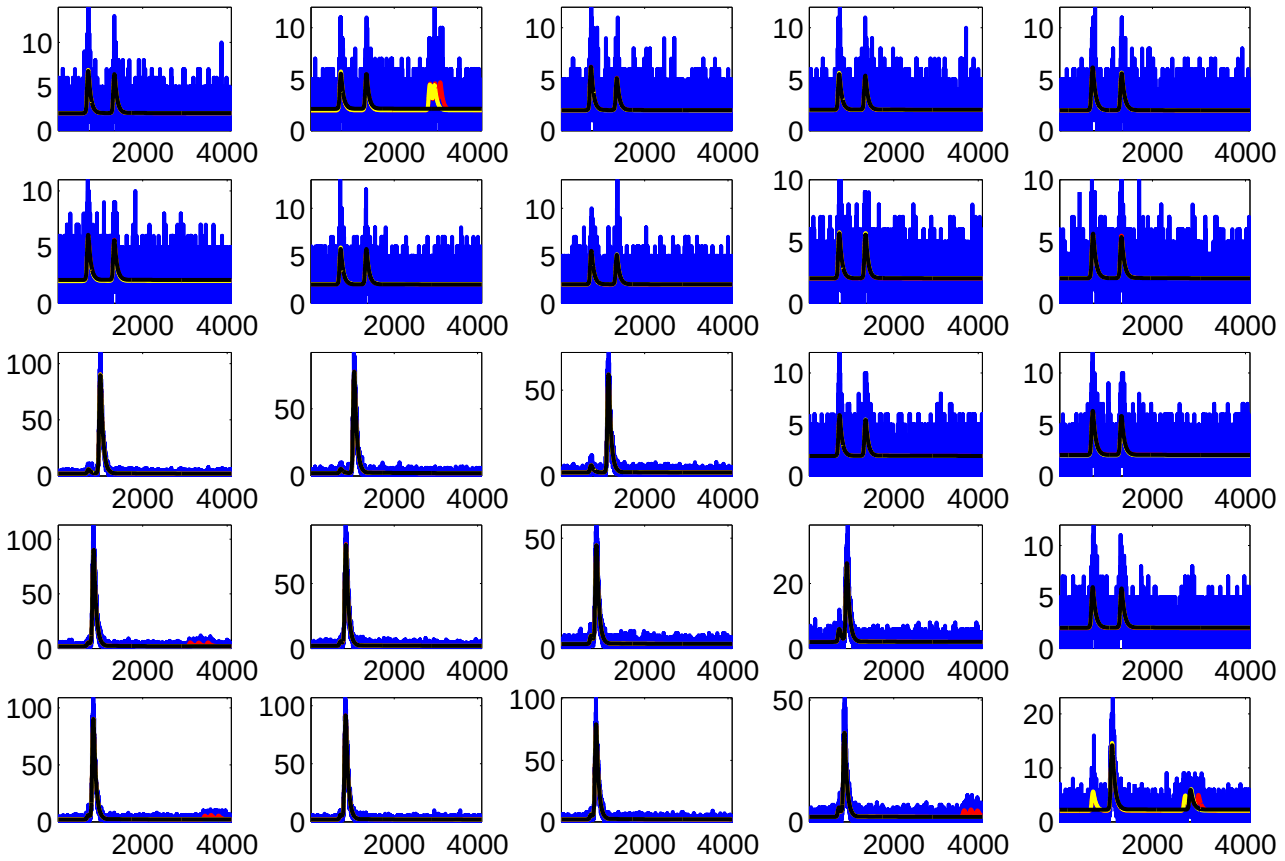



**Figure A.5:** Array of photon count histograms (in blue) corresponding to block 2 of figure 5.3 and final fit without spatial constraints(red), with $\psi = 1$ (yellow) and $\psi = 20$ (black).

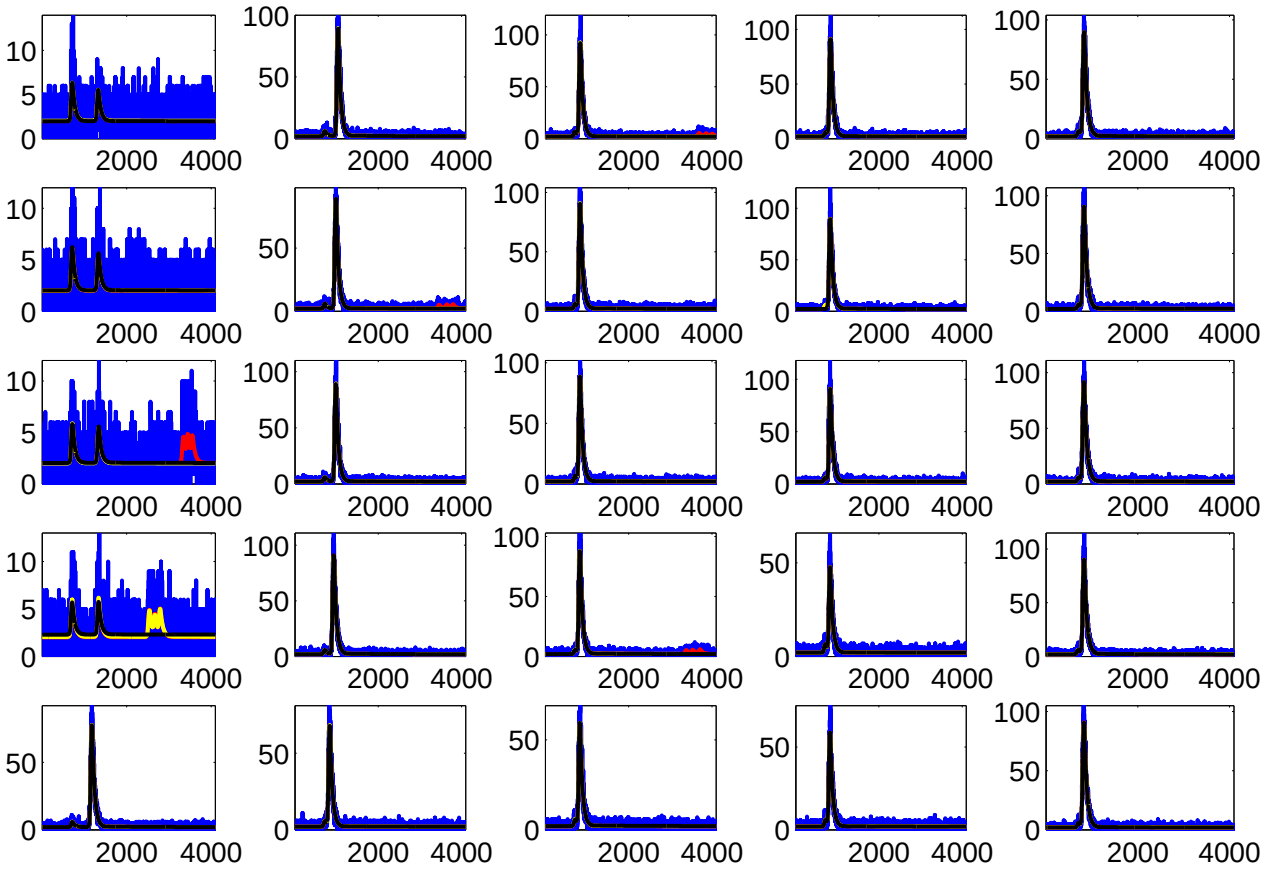



**Figure A.6:** Array of photon count histograms (in blue) corresponding to block 3 of figure 5.3 and final fit without spatial constraints(red), with $\psi = 1$ (yellow) and $\psi = 20$ (black).

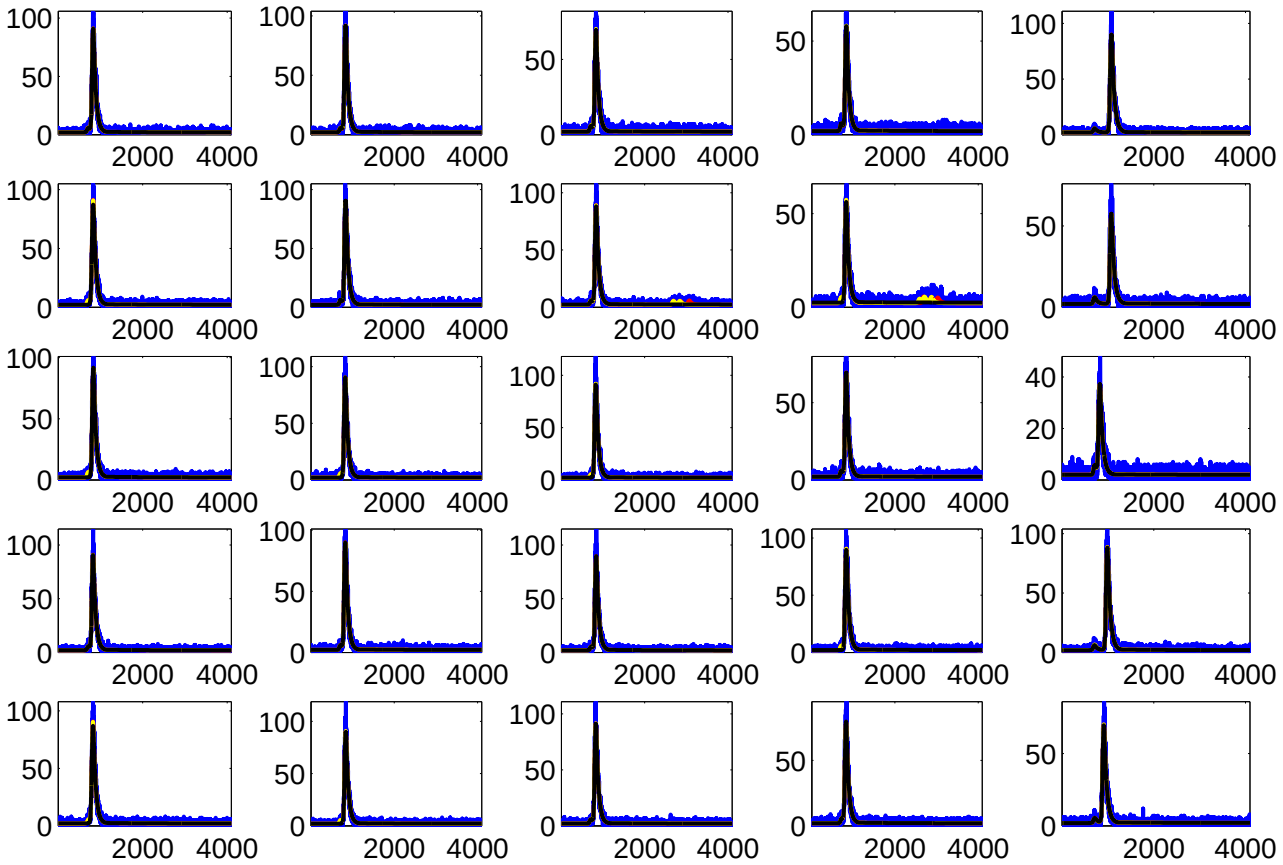



**Figure A.7:** Array of photon count histograms (in blue) corresponding to block 4 of figure 5.3 and final fit without spatial constraints(red), with $\psi = 1$ (yellow) and $\psi = 20$ (black).

Appendix B

# Analysis of the Shape Parameters of the TCSPC Operational Model

Some initial work has been carried out to estimate the shape parameters of the operational model described in section 4.3. Since only one peak is used the dimension of the space of solutions is fixed and the different unknowns to be analysed are therefore: the amplitude factor, $\beta$, the time of the peak maximum,$t_0$, the scale parameter of the peak gaussian-like distribution, $\sigma$, the points at which transitions between the different piecewise functions occur, i.e. $t_1$, $t_2$ and $t_3$, and the corresponding scale parameters $\tau_1$, $\tau_2$ and $\tau_3$. As in all the Bayesian analysis, these unknowns are regarded as drawn from appropriate prior distributions. We take for the parameters $\beta$ and $t_0$ the same independent priors that we have explained in previous chapters, that is, a gamma and a uniform distribution respectively. However, the assumption of independence is not valid for the rest of the parameters. For instance, the values $t_0$, $t_1$, $t_2$ and $t_3$ are constrained to be ordered as $t_1 < t_0 < t_2 < t_3$ as defined by the four-picewise equation defined in equation 4.1. The allowable constraints employed in this process are as follows

- Shape parameters: all parameters are positive, since negative values are not physically realisable.
- Breakpoint ordering: $t_1 < t_0 < t_2 < t_3$ as that corresponds to the piecewise definition of equation 4.1.
- Scale parameters: $\tau_2 < \tau_3$, to ensure that there is no abrupt transitions between the two exponentials.

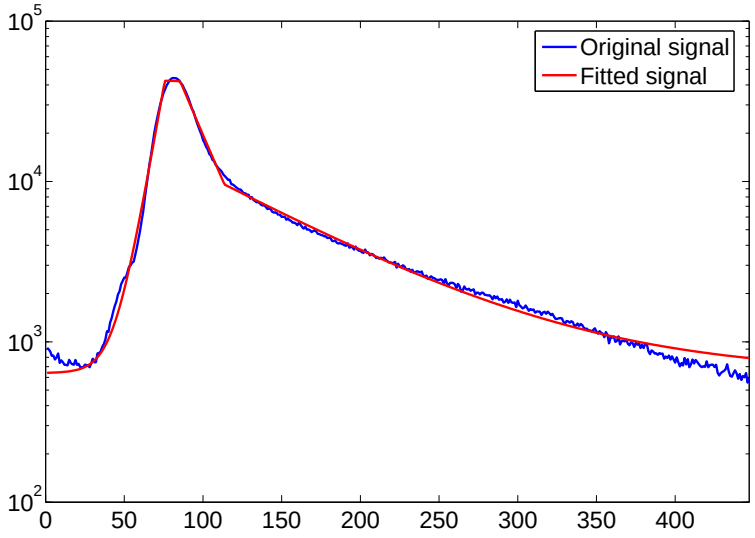
As we are using a Bayesian approach, these constraints will be represented in the form of a suitable prior. For the prior distribution of the parameters $\beta$, $\sigma$, $\tau_1$ and $\tau_3$ we assume

gamma distributions since their support is on $\Re^+$. The parameter $t_1$ is drawn from a flat uniform distribution with support in the interval $[1, t_0]$ whereas the parameter $t_3$ from a flat uniform distribution but this time with support on $[t_0, maximum\ channel\ number]$. The remaining parameters show dependencies with some of the previously mentioned parameters. Thus, the parameter $t_2$ is assumed to be drawn from a flat uniform distribution with support in $[t_0, t_3]$ and the parameter $\tau_2$ from a flat uniform distribution with support in $[0, \tau_3]$. We have chosen independent uniform priors so that we can easily assess the effect of the data when looking at the corresponding posterior distributions, however, other priors may be more suitable for this application. We do not claim any optimality in the choice of these priors since this only an attempt to better understand a fully Bayesian approach of the problem. More accurate analysis of this data is left as future work. The full joint prior distribution is therefore expressed as

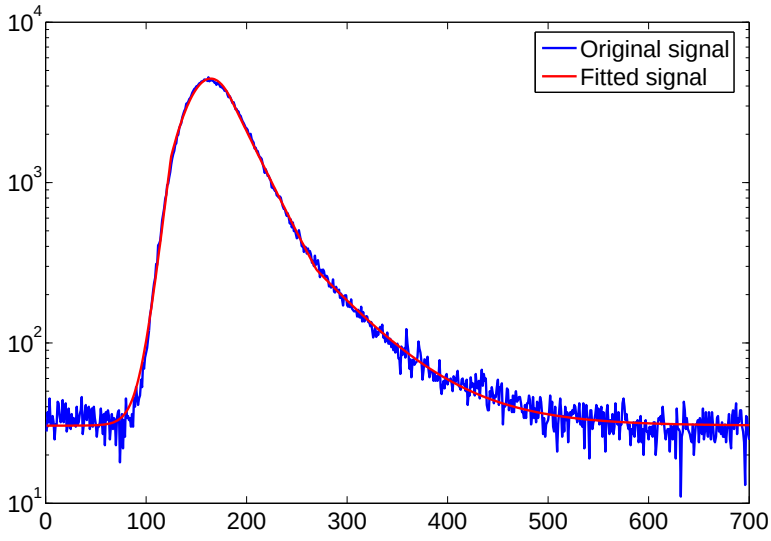
$$f(\beta, t_0, \sigma, t_1, t_2, t_3, \tau_1, \tau_2, \tau_3) = f(\beta)\cdot f(t_0)\cdot f(\sigma) f(t_1|t_0)\cdot f(t_3|t_0)\cdot f(t_2|t_0, t_3)\cdot f(\tau_1)\cdot f(\tau_3)\cdot f(\tau_2|\tau_3) \tag{B.1}$$

We consider the processing of two different lidar reference signals, to extract the different shape parameters. To obtain estimates of these parameters, we suggest a single-component Metropolis-Hastings algorithm in which the acceptance probability of a candidate parameter value involves ratios of the full conditional distribution of the updated parameter given all other parameters and the data and of course the proposal distribution. Figure B.1 displays some preliminary results using the priors detailed. Figure B.1(a) shows a relatively bad fitting of the reference signal whereas figure B.1(b) shows a good fitting of the signal. Figure B.2 and figure B.3 show the different time-series trace plots obtained for the different parameters for the reference signals shown in figure B.1(a) and B.1(b) respectively. Note that the first 1000 iterations are removed to facilitate the display of the different trace plots.

As can be observed in figure B.1(a) the peak of the signal seems to be flatten instead of having a sharp shape. This is due to the left-hand side tail of the 4-piecewise exponential function. As little prior information is available, the MCMC algorithm tries to fit this tail to the detriment of the peak. The contribution of this tail to likelihood is therefore bigger than the contribution of the peak. As a consequence, the parameter $t_1$ seems to




(a) Bad fitting




(b) Good fitting

**Figure B.1:** Histogram of photon counts of real data (blue) and final fit from MCMC(red) considering all the parameters as unknowns

approach the value of the parameter $t_0$ and the time-series trace plots they generated show a similar behaviour. A similar effect is seen on the right-hand side tail corresponding to the parameter $t_3$ which tends to approach to $t_0$.

In figure B.1(b) the algorithm seems to find appropriate estimates of the different shape parameters. It can also be observed the dynamic variation of the time-series trace plots which show that stronger dependencies between the different parameters may be observed in the posteriors, *e.g.* between the parameters $\sigma$ and $t_1$ or the amplitude factor $\beta$ and the position of the peak maximum $t_0$.

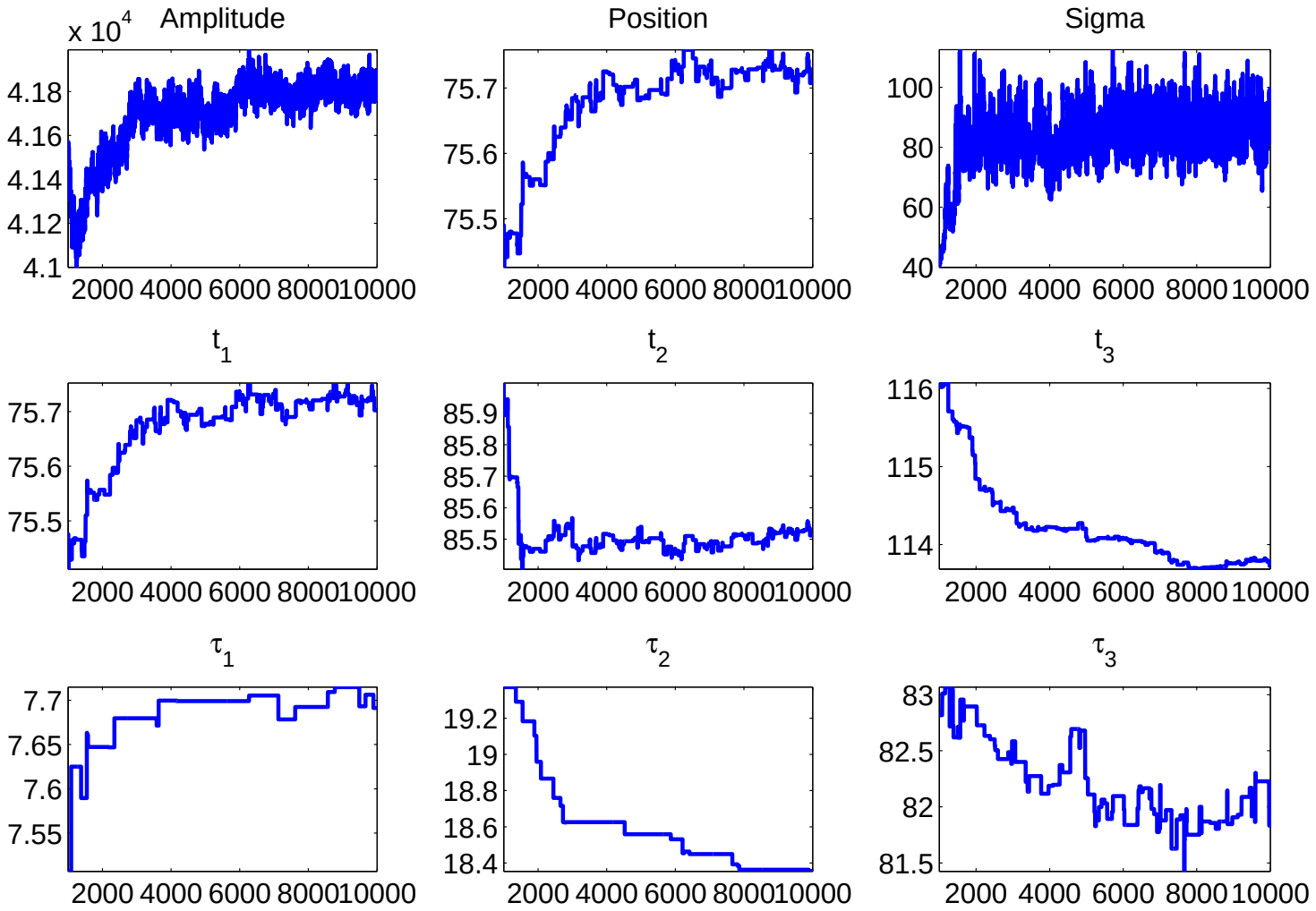



**Figure B.2:** Time-series trace plots obtained using the MCMC algorithm for the different parameters. Bad fitting.

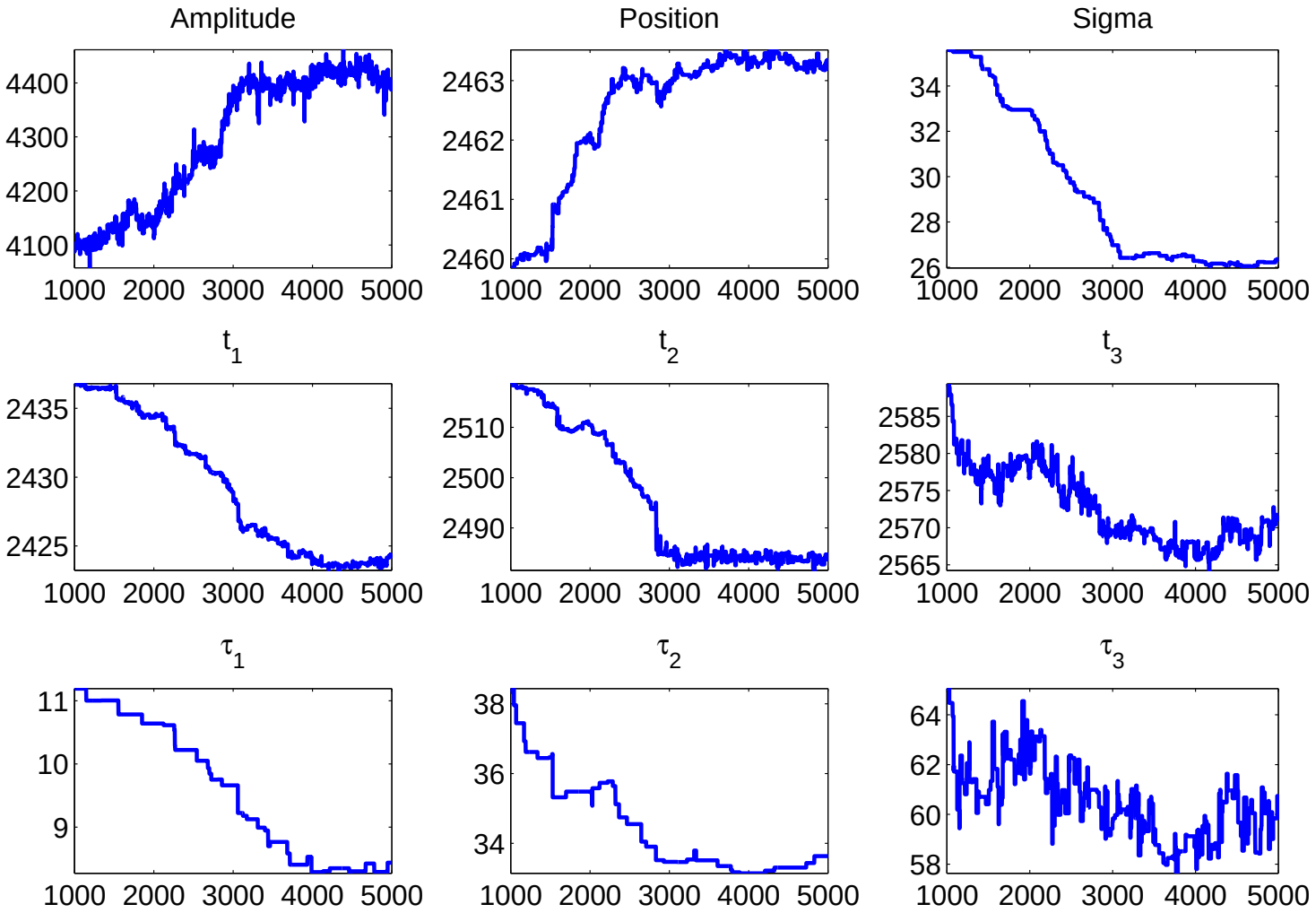



**Figure B.3:** Time-series trace plots obtained using the MCMC algorithm for the different parameters. Correct fitting.

# Bibliography


[1] H. Ailisto, V. Heikkinen, R. Mitikka, et al. Scannerless imaging pulsed-laser range finding. *J. Opt. A: Pure Appl. Opt.*, 4:S337–S346, 2002.

[2] F. Al-Awadhi, M. Hurn, and C. Jennison. Improving the acceptance rate of reversible jump MCMC proposals. *Statistics and Probability letters*, 69(2):189–198, 2004.

[3] M. A. Albota, R. M. Heinrichs, D. G. Kocher, et al. Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser. *Appl. Opt.*, 41(36):7671–7678, 2002.

[4] M. C. Amann, T. Bosch, M. Lescure, R. Myllyla, and M. Rioux. Laser ranging: a critical review of usual techniques for distance measurement. *Opt. Eng*, 40(1):10–19, 2001.

[5] P. Andersson. Long-range three-dimensional imaging using range-gated laser radar images. *Opt. Eng*, 45(3):1–10, 2006.

[6] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking. *IEEE Trans. on Signal Processing*, 50(2):174–188, 2002.

[7] B. F. Aull, A. H. Loomis, D. J. Young, et al. Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging. *Lincoln Laboratory Journal*, 13(2):335–348, 2002.

[8] I. Baker, S. Duncan, and J. Copley. A low noise, laser-gated imaging system for long range target identification. In *Proc. of SPIE. Infrared Technology and Applications XXX*, volume 5406, pages 133–144, 2004.

[9] W. Becker, A. Bergmann, H. Wabnitz, et al. High-count-rate multichannel TCSPC for optical tomography. In *Proc. of SPIE. Photon Migration, Optical Coherence Tomography, and Microscopy*, volume 4431, pages 249–254, 2001.

[10] W. Becker, A. Bergmann, H. Wabnitz, et al. Picosecond fluorescence lifetime microscopy by TCSPC imaging. In *Proc. of SPIE. Multiphoton Microscopy in the Biomedical Sciences*, volume 4262, pages 414–419, 2001.

[11] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* John Wiley, Chichester, 1994.

[12] K. K. Berthelsen and J. Moller. Bayesian analysis of Markov point processes. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Case Studies in Spatial Point Process Modeling*, volume 185 of *Springer Lecture Notes in Statistics*, pages 85 – 97. Oxford University Press, Oxford, 2003.

[13] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, 36:192–236, 1974.

[14] J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B*, 48:259–302, 1986.

[15] J. Besag. Digital image processing. Towards Bayesian image analysis. *J. Applied Stats.*, 16(3):395–407, 1989.

[16] J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. B*, 55(1):25–37, 1993.

[17] J. Besag, P. J. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10:3–66, 1995.

[18] J. Besag and D. Higdon. Bayesian analysis of agricultural field experiments. *J. Roy. Statist. Soc. Ser. B*, 61:691–745, 1999.

[19] J. Besag, J. York, and A. Mollie. Bayesian image-restoration, with 2 applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43(1):1–59, 1991.

[20] S. Brooks and P. Giudici. Convergence assessment for reversible jump MCMC simulations, 1998.

[21] S. P. Brooks, P. Giudici, and G. O. Roberts. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J. Roy. Statist. Soc. B*, 65(1):3–55, 2003.

[22] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference. A practical information-theoric approach.* Springer, 2002.

[23] J. Busck. Underwater 3-D optical imaging with a gated viewing laser radar. *Opt. Eng*, 44(11):1–7, 2005.

[24] J. Busck and H. Heiselberg. Gated viewing and high-accuracy three-dimensional laser radar. *Appl. Opt.*, 43(24):4705–4710, 2004.

[25] J. Busck and H. Heiselberg. High accuracy 3-D laser radar. In *Proc. of SPIE. Laser Radar Technology and Applications IX*, volume 5412, pages 257–263, 2004.

[26] O. Cappe, C. P. Robert, and T. Ryden. Reversible jump MCMC converging to birth-and-death MCMC and more general continuous time samplers. *J. Roy. Statist. Soc. B*, 65(3):679–700, 2003.

[27] B. P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B*, 57(3):473–484, 1995.

[28] R. Casella and R. Berger. *Statistical Inference.* Wadsworth & Brooks, 1990.

[29] G. Celeux, D. Chauveau, and D. Diebolt. On stochastic versions of the EM algorithm, 1995. Technical Report RR-2514. INRIA, France.

[30] P. Chaudhuri and J. S. Marron. SiZer for exploration of structure in curves. *J. Am. Stat. Assoc.*, 94:807–823, 1999.

[31] R. Chen and J. S. Liu. Mixture Kalman filters. *J. Roy. Statist. Soc. B*, 63(3):493–508, 2000.

[32] N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89:539–551, 2002.

[33] P. Churoux and C. Besson. Burst illumination laser imaging system for long range observation. In *Proc. of SPIE. Laser Radar Technology and Applications IV*, volume 3707, pages 582–589, 1997.

[34] P. Churoux, C. Besson, and J. P. Bouzinac. Model of a burst imaging lidar through the atmosphere. In *Proc. of SPIE. Laser Radar Technology and Applications V*, volume 4035, pages 324–331, 2000.

[35] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Stat. Assoc.*, 91:883–904, 1996.

[36] N. A. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Wiley, John & Sons, Incorporated, 1993.

[37] P. Dellaportas. Bayesian classification of neolithic tools. *Applied Statistics*, 47(2):279–297, 1998.

[38] P. Dellaportas and I. Papageorgiou. Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, 16(1):57–68, 2006.

[39] A. P. Dempster, N. M. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B*, 39:1–38, 1977.

[40] U. R. Dhond and J. K. Aggarwal. Structure from stereo: A review. *IEEE Trans. Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[41] J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B*, 56(2):363–375, 1994.

[42] A. Diplaros, T. Gevers, and N. Vlassis. Skin detection using the EM algorithm with spatial constraints. In *Proc. of IEEE International conference on Systems, Man and Cybernetics*, volume 4, pages 3071–3075, 2004.

[43] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001.

[44] R. G. Driggers, R. H. Vollmerhausen, and N. Devitt. Impact of speckle on laser range-gated shortwave infrared imaging system target identification performance. *Opt. Eng*, 42(3):738–746, 2003.

[45] I. L. Dryden, M. R. Scarr, and C. C. Taylor. Bayesian texture segmentation of weed and crop images using reversible jump Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. C*, 52(1):31–50, 2003.

[46] S. Duncan, P. Copley, G. Harvey, et al. Advances in laser gated imaging in airborne environment. In *Proc. of SPIE. Infrared Technology and Applications XXXII*, volume 6206, 2006.

[47] D. Edouart, P. Churoux, and P. H. Flamant. Burst illumination imaging lidar: intensity correlation function in the image plane. In *Proc. of SPIE. Laser Radar Technology and Applications VII*, volume 4723, pages 189–197, 2002.

[48] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:97–108, 1993.

[49] M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *J. American Statistical Association*, 90(430):577–588, 1995.

[50] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):406–417, 2005.

[51] P. Fearnhead. Particle filters for mixture models with an unknown number of components. *Statistics and Computing*, 14(1):11–21, 2004.

[52] C. Fernandez and P. J. Green. Modelling spatially correlated data via mixtures: a Bayesian approach. *J. Roy. Statist. Soc. Ser. B*, 64:805–826, 2002.

[53] B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.

[54] D. Gamerman. *Markov Chain Monte Carlo. Stochastic simulation for Bayesian inference.* Chapman & Hall, 1997.

[55] A. Gelbart, B. Redman, R. S. Light, et al. Flash lidar based on multiple-slit streak tube imaging lidar. In *Proc. of SPIE. Laser Radar Technology and Applications VII*, volume 4723, pages 9–18, 2002.

[56] A. Gelman and X. Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, 13(2):163–185, 1998.

[57] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511, 1992.

[58] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell*, 14:367–383, 1992.

[59] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell*, 6:721–741, 1984.

[60] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, 1995.

[61] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena*, volume 9 of *Topics in Applied Physics*, pages 9 – 75. Springer-Verlag, Heidelberg, 1975.

[62] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imaging*, 9:84–93, 1990.

[63] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[64] P. J. Green. Trans-dimensional Markov chain Monte Carlo. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, volume 27 of *Oxford Statistical Science Series*, pages 179–198. Oxford University Press, Oxford, 2003.

[65] P. J. Green and A. Mira. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika*, 88:1035–1053, 2001.

[66] P. J. Green and S. Richardson. Hidden Markov models and disease mapping. *J. American Statistical Association*, 97:1055–1070, 2002.

[67] U. Grenander. Tutorial in pattern theory, 1983. Technical Report, Division of Applied Mathematics, Brown University, Providence.

[68] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *J. Roy. Statist. Soc. B*, 56(4):549–603, 1994.

[69] P. Hall and D. M. Titterington. On some smoothing techniques used in image restoration. *J. Roy. Statist. Soc. B*, 48(3):330–343, 1986.

[70] M. J. Halmos, M. J. J. Asbrock, et al. 3-D flash ladar at raytheon. In *Proc. of SPIE. Laser Radar Technology and Applications VI*, volume 4377, pages 84–97, 2001.

[71] J. Harezlak. *Bump hunting in regression data revisited*. PhD thesis, Simon Fraser University, 1998.

[72] D. Hastie. *Towards Automatic Reversible Jump Markov chain Monte Carlo*. PhD thesis, University of Bristol, 1998.

[73] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:1317–1340, 1970.

[74] G. Hazel. Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. Geoscience and Remote Sensing*, 38(3):1199–1211, 2000.

[75] M. Hebbert and E. Krotkov. 3D measurements from imaging laser radars: How good are they? *Image and Vision Computing*, 10(3):170–178, 1992.

[76] M. Hebert. Active and passive range sensing for robotics. In *Proc. of 2000 IEEE International Conference on Robotics and Automation*, volume 1, pages 102–110, 2004.

[77] R. M. Heinrichs, B. F. Aull, R. M. Marino, et al. Three-dimensional laser radar with APD arrays. In *Proc. of SPIE. Laser Radar Technology and Applications VI*, volume 4377, pages 106–117, 2001.

[78] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson. Markov chain Monte Carlo Algorithms for 3D Ranging and Imaging. In *Proc. of IAPR conference on Machine Vision Applications*, pages 193–196, Tsukuba, Japan, May 2005.

[79] D. Higdon. *Spatial Aplications of Markov chain Monte Carlo for Bayesian inference*. PhD thesis, University of Bristol, 1998.

[80] D. M. Higdon. Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Stat. Assoc.*, 93(442):585–595, 1998.

[81] D. M. Higdon, J. E. Bowsher, V. E. Johnson, et al. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Trans. on Medical Imaging*, 16(5):516–526, 1997.

[82] P. A. Hiskett, G. Bonfrate, G. S. Buller, and P. D. Townsend. Eighty kilometre transmission experiment using an InGaAs/InP SPAD based quantum cryptography receiver operating at 1.55m. *J. Modern Optics*, 48:19571966, 2001.

[83] N. L. Hjort and H. Omre. Topics in spatial statistics. *Scand. J. Statist.*, 21:289–357, 1994.

[84] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.*, 20(1):50–67, 2005.

[85] K. Johnson, M. Vaidyanathan, et al. Adaptive LADAR receiver for multispectral imaging. In *Proc. of SPIE. Laser Radar Technology and Applications VI*, volume 4377, pages 98–105, 2001.

[86] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *J. Amer. Stat. Assoc.*, 91(435):1343–1370, 1996.

[87] M. R. Kitchin, S. Hernández Marín, A. M. Wallace, and G. J. Gibson. Processing and registration of range images from time of flight laser systems. In *Proc. of SPIE. Electro-Optical and Infrared Systems: Technology and Applications II*, volume 5988, pages 19–30, 2005.

[88] L. Knorr-Held and R. Gunter. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56:13–21, 2000.

[89] C. Lacoste, X. Descombes, and J. Zerubia. Point processes for unsupervised line network extraction in remote sensing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1568–1579, October 2005.

[90] L. Le Hors, P. Hartemann, and S. Breugnot. Multispectral polarization active imager in the visible band. In *Proc. of SPIE. Laser Radar Technology and Applications V*, volume 4035, pages 380–388, 2000.

[91] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. Springer, 2001.

[92] D. V. Lindley. A statistical paradox. *Biometrika*, 44:187–192, 1957.

[93] K. V. Mardia. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivar. Anal.*, 24(2):265–284, 1988.

[94] R. M. Marino, T. Stephens, R. E. Hatch, et al. A compact 3D imaging laser radar system using geiger mode APDs: system and measurements. In *Proc. of SPIE. Laser Radar Technology and Applications VIII*, volume 5096, pages 1–15, 2003.

[95] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Stat. Assoc.*, 82(397):76–89, 1987.

[96] J. L. Marroquin. Surface reconstruction preserving discontinuities, 1984. Technical Report. Massachusetts Institute of Technology. Cambridge, MA, USA.

[97] J. S. Massa, G. S. Buller, A. Walker, et al. A time-of-flight optical ranging system using time-correlated single-photon counting. *Applied optics*, 37(31):7298–7304, 1998.

[98] P. McCullagh and J. Nelder. *Generalized Linear Models, 2nd Edition*. Chapman and Hall, 1989.

[99] G. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons Ltd., 2000.

[100] G. J. McLachlan. On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. *Appl. Statist.*, 36(3):318–324, 1987.

[101] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.

[102] A. Mira. On Metropolis-Hastings algorithms with delayed rejection. *Metron.*, LIX:231–241, 2001.

[103] J. Møller. Markov chain Monte Carlo and spatial point processes. In O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry,*

*Likelihood, and Computation*, Proceedings Seminaire Européen de Statistique. Chapman and Hall, 1997.

[104] D. J. Murdoch and P. J. Green. Exact sampling for a continous state. *Scand. J. Statist.*, 25(3):483–502, 1998.

[105] J. T. Murray, S. E. Moran, N. Roddier, et al. Advanced 3D polarimetric flash ladar imaging through foliage. In *Proc. of SPIE. Laser Radar Technology and Applications VIII*, volume 5086, pages 84–95, 2003.

[106] C. Niclass, A. Rochas, P. Besse, and E. Charbon. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits*, 40(9):1847–1854, 2005.

[107] A. Nobile and A. Fearnside. Bayesian finite mixtures with an unknown number of components: the allocation sampler, 2005. Technical Report 05-4. Department of Statistics, University of Glasgow.

[108] K. Patorski. *Handbook of the Moire Fringe Technique*. Elsevier, 1993.

[109] S. Pellegrini, G. Buller, J. Smith, A. Wallace, and S. Cova. Laser-based distance measurement using picosecond resolution TCSPC. *Meas. Sci. Technol.*, 11:712–716, 2000.

[110] P. H. Peskun. Optimum Monte Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.

[111] G. Petris and L. Tardella. A geometric approach to transdimensional Markov chain Monte Carlo. *Can. J. Stat.*, 31(4):469–482, 2003.

[112] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, 1985.

[113] S. C. Popescu, R. H. Wynne, and R. F. Nelson. Measuring individual tree crown diameter with lidar and assesing its influence on estimating forest volume and biomass. *Can. J. Remote Sensing*, 29(5):564–577, 2003.

[114] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1&2):223–252, 1996.

[115] D. Pyburn, R. Leon, B. Haji-Saeed, et al. Low-power portable scanning imaging ladar system. In *Proc. of SPIE. Laser Radar Technology and Applications VIII*, volume 5086, pages 96–103, 2003.

[116] A. E. Raftery and S. Lewis. How many iterations in the Gibbs sampler? In J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4 of *Oxford Statistical Science Series*, pages 763–773. Oxford University Press, Oxford, 1992.

[117] B. Redman, W. Ruff, and M. Giza. Photon counting chirped AM ladar: Concept, simulation, and initial experimental results. In *Proc. of SPIE. Laser Radar Technology and Applications XI*, volume 6214, pages 1–9, 2006.

[118] H. Ren and C. I. Chang. A generalized orthogonal subspace projection approach to unsupervised multispectral image classification. *IEEE Trans. Geoscience and Remote Sensing*, 38(6):2515–2528, 2000.

[119] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B*, 59:731–792, 1997.

[120] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, 2004.

[121] C. P. Robert and G. Casella. *Monte Carlo statistical methods.* Springer texts in statistics, 1999.

[122] C. P. Robert, T. Rydn, and D. M. Titterington. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. Roy. Statist. Soc. B*, 62(1):57–75, 2000.

[123] G. O. Roberts and N. Polson. On the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. B*, 56(2):377–384, 1994.

[124] G. O. Roberts and A. F. M. Smith. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their applications*, 49:207–216, 1994.

[125] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[126] S. H. Rowe and W. T. Welford. Surface topography of non-optical surfaces by projected interference fringes. *Nature*, 216(5177):786–787, 1967.

[127] J. Salvi. *An Approach to Coded Structured Light to Obtain 3D Information.* PhD thesis, University of Girona, 2001.

[128] B. W. Schilling, B. N. Dallas, G. C. Templeton, et al. Multiple-return laser radar for three-dimensional imaging through obscurations. *Applied Optics*, 41(15):2791–2799, 2002.

[129] R. Schneider, P. Thrmel, and M. Stockmann. Distance measurement of moving objects by frequency modulated laser radar. *Opt. Eng*, 40(1):33–37, 2001.

[130] Z. Shun and P. McCullagh. Laplace approximation of high dimensional integrals. *J. Roy. Statist. Soc. B*, 57(4):749–760, 1995.

[131] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B*, 55(1):3–23, 1993.

[132] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc. B*, 64(4):583–639, 2002.

[133] A. Srivastava, U. Grenander, G. R. Jensen, and M. I. Miller. Jump diffusion Markov processes on orthogonal groups for object pose estimation. *J. Statist. Planning and Inference*, 103(1-2):15–37, 2002.

[134] B. L. Stann, A. Abou-Auf, K. Aliberti, et al. Research progress on a focal plane array ladar system using a laser diode transmitter and FM/cw radar principles. In

*Proc. of SPIE. Laser Radar Technology and Applications VII*, volume 4723, pages 19–30, 2002.

[135] B. L. Stann, A. Abou-Auf, K. Aliberti, et al. Research progress on a focal plane array ladar system using chirped amplitude modulation. In *Proc. of SPIE. Laser Radar Technology and Applications VIII*, volume 5086, pages 47–57, 2003.

[136] I. Steinsland. Parallel sampling of GMRFs and geostatistical GMRF models, 2003. Technical Report Preprint statistics 7/2003. Norwegian University of Science and Technology (NTNU).

[137] O. Steinvall and T. Carlsson. Three-dimensional laser radar modelling. In *Proc. of SPIE. Laser Radar Technology and Applications VI*, volume 4377, pages 23–33, 2001.

[138] D. A. Stephens and R. D. Fisch. Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics*, 54(4):1334–1347, 1998.

[139] M. Stephens. Bayesian analysis of mixture models with an unknown number of components - an alternative to reversible jump methods. *The Annals of Statistics*, 28(1):40–74, 2000.

[140] M. Stephens. Dealing with label switching in mixture models. *J. Roy. Statist. Soc. Ser. B*, 62:795–809, 2000.

[141] R. H. Swendsen and J. S. Wang. Non-universal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[142] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *The Annals of Applied Probability*, 8(1):1–9, 1998.

[143] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *J. Amer. Stat. Assoc.*, 81(393):82–86, 1986.

[144] L. Tierney and A. Mira. Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, 18:2507–2515, 1999.

[145] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical analysis of finite mixture distributions*. Wiley series in probability and mathematical statistics, 1985.

[146] H. Tjelmeland and J. Besag. Markov random fields with higher-order interactions. *Scand. J. Statist.*, 25:415–433, 1998.

[147] H. Tjelmeland and B. K. Hegstad. Mode jumping proposals in MCMC. *Scand. J. Statist.*, 28:205–223, 2001.

[148] B. Tso and P. M. Mather. *Classification methods for remotely sensed data*. Taylor & Francis New York, 2001.

[149] M. Umasuthan, A. M. Wallace, J. S. Massa, et al. Processing time-correlated single photon data to acquire range images. *IEE Proc. Vis. Image Signal Process*, 145(4):237–243, 1998.

[150] M. Vaidyanathan, A. Joshi, B. Song Xue Hanyaloglu, et al. High performance ladar focal plane arrays for 3d range imaging. In *Proc. of IEEE Aerospace Conference*, volume 3, pages 1776–1781, 2004.

[151] V. Viallefont, S. Richardson, and P. J. Green. Bayesian analysis of poisson mixtures. *J. of Nonparametric Statistics*, 14:181–202, 2002.

[152] R. Waagepetersen and D. Sorensen. A tutorial on reversible jump MCMC with a view toward QTL-mapping. *International Statistical review*, 69:49–61, 2001.

[153] A. M. Wallace, G. S. Buller, R. C. W. Sung, et al. Multi-spectral laser detection and ranging for range profiling and surface characterization. *J.Opt.A: Pure Appl. Opt*, 7:S438–S444, 2005.

[154] A. M. Wallace, G. S. Buller, and A. C. Walker. 3D imaging and ranging by time-correlated single photon counting. *Computing & Control Eng. Journal*, 12(4):157–168, 2001.

[155] A. M. Wallace, P. Csakany, G. S. Buller, and A. C. Walker. 3D imaging of transparent objects. In *Proc. of British Machine Vision Conference*, 2000.

[156] A. M. Wallace, R. C. W. Sung, G. S. Buller, et al. Detecting and characterising returns in a pulsed ladar system. *IEE Proc. Vis. Image Signal Process*, 153(2):160–172, 2006.

[157] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on acoustics, speech and signal processing*, ASSP-33(2):387–392, 1985.

[158] M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, and S. M. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fmri data. *IEEE Trans. on Medical Imaging*, 24(1):1–11, 2005.

[159] W. Y. and A. E. H. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. In *Proc. of IEEE conference on Computer Vision and Pattern Recognition*, pages 321–326, 1996.

[160] M. Zahid, J. S. Smith, and L. J. High-frequency phase measurement for optical rangefinding system. *IEE Proc. Sci. Meas. Technol.*, 144(3):141–148, 1997.

[161] F. Zappa, S. Tisa, S. Cova, et al. Single-photon avalanche diode arrays for fast transients and adaptive optics. *IEEE Trans. Instr. and Meas.*, 55(1):365–374, 2006.

[162] G. Zhang and A. M. Wallace. Physical modeling and combination of range and intensity edge data. *CVGIP: Image Understanding*, 58(2):191–220, 1993.

[163] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survery. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.