WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/141041

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

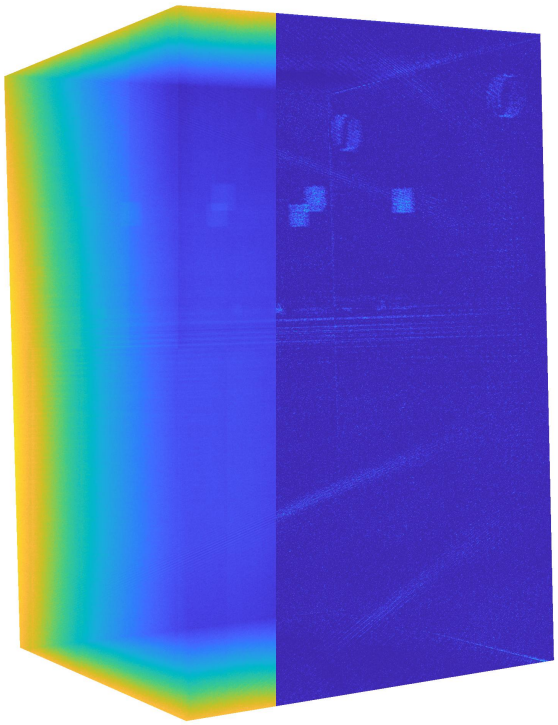
For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Characterisation of Computed Tomography Noise in Projection Space with Applications to Additive Manufacturing

Sherman Edea Lo

Doctor of Philosophy in Statistics



University of Warwick
Department of Statistics

April 2020

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors Julia Brettschneider and Tom Nichols for their guidance throughout my PhD. I would also like to thank my calibrators who are part of the Inside-Out group, in Warwick Statistics, Clair Barnes, Wilfred Kendall and Audrey Kueh, and the Warwick Manufacturing Group, Greg Gibbons, Jay Warnett and Mark Williams.

This work is funded by the EPSRC and MRC Centre for Doctoral Training in Next Generation Statistical Science: The Oxford-Warwick Statistics Programme (EP/L016710/1). I would like to thank my cohort for the friendship and collaboration during the programme: Nathan Cunningham, Giuseppe di Benedetto, Beniamino Hadj-Amar, Jack Jewson, Ella Kaye, Leon Law, Kaspar Märtens, Marcin Mider, Xenia Miscouridou, Paul Vanetti, Andi Wang.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- The fabrication, x-ray computed tomography acquisition and simulation of test samples were done by Greg Gibbons and Jay Warnett, from the Warwick Manufacturing Group, as part of Inside-out: Statistical methods for Computed Tomography validation of complex structures in Additive Layer Manufacturing funded by EPSRC (EP/K031066/1).

Abstract

X-ray computed tomography can be used for defect detection in additive manufacturing. Typically, several x-ray projections of the product at hundreds of angles are used to reconstruct the object in 3D to look for any defects. The process can be time-consuming. This thesis aims to investigate if it is possible to conduct defect detection from a single projection to speed up the process. An additive manufacturing test sample was created with voids to see if they can be detected.

The uncertainty of the projection was modelled using a compound Poisson distribution. This arises from x-ray photon arrivals being a Poisson process and each photon has random energy. This results in a linear relationship between the mean and variance of the grey values in the projection. Fitting of the compound Poisson distribution using the expectation-maximisation algorithm was unsuccessful due to identifiability issues with the model. Instead, a gamma-distributed generalised linear model was fitted onto sample variance-mean data and used for variance prediction to quantify the uncertainty.

Software, called *aRTist*, was used to simulate the projection and compared with the experimental projection in the face of uncertainty by treating each pixel as a hypothesis test. To overcome the imperfections of the simulation, the empirical null filter was used to cater for model misspecification so that sensible inference was achieved. This was done by locally normalising the test statistics using the mode. Voids with diameters in the order of millimetres were detectable.

This thesis is a contribution to real-time quality control in additive manufacturing.

Chapter 1

Introduction

In the field of engineering, additive manufacturing is an emerging technology and has uses in producing bespoke products. Because it is a new form of technology, the process is not well understood and tolerances are not as precise as other forms of manufacturing. As a result, there exist methods for quality control of additive manufactured products. Typically, this is done using x-ray computed tomography and requires hundreds of x-ray projections, making this a slow process.

The main aim of this thesis is to investigate if it is possible to speed up the quality control process of additive manufactured products by using only a few x-ray projections. By using fewer x-ray projections, uncertainty is introduced. This can be tackled by using statistics because it enables the sensible handling of uncertainty from sources of error such as random error and systematic error.

The analysis was done by comparing the experimental projection with a simulated projection to look for areas with disagreement in the face of uncertainty. Random error can arise from how x-ray photons are produced in an x-ray tube and how they interact with the additive manufactured product and the x-ray detector. This was modelled by using the compound Poisson distribution. Incorrect simulation of the projection contribute to systematic error and this was corrected using the empirical null filter. These sources of error were considered in the comparison of the projections with

the use of hypothesis testing.

The front cover shows the before and after of the statistical analysis. The left-hand side shows an x-ray projection of an additive manufactured cuboid. Its edges appeared curved due to spot and panel effects and this can be fixed using shading correction. The right-hand side shows the p-values of the resulting inference. Lighter colours show evidence of a defect and they successfully highlighted voids put in there purposefully.

In Chapter 2, x-ray computed tomography and additive manufacturing is reviewed. Sources of error were investigated and it was discussed how they were handled. In Chapter 3, a test sample was additively manufactured and the chapter describes how experimental x-ray projections were obtained. Shading correction is explained here and it was used to remove sources of systematic error in the projections. In Chapter 4, the compound Poisson distribution is studied so that it can be used to model the detection of x-rays. In Chapter 5, the uncertainty in the projection was quantified using the variance and generalised linear models were used to predict it. In Chapter 6, novel statistical techniques were developed and implemented to look for defects in the test sample in the face of uncertainty. This was done by comparing the experimental projection with a simulated projection and looking for any disagreement. The empirical null filter was used to cater for any model misspecification so that sensible conclusions were made. The thesis ends with an evaluation in Chapter 7.

The results presented in this thesis can be reproduced using the source code in the *GitHub* repository `https://github.com/shermanlo77/oxwasp_phd`.

Chapter 2

Literature Review

One of the first methods of additive manufacturing (AM) is stereolithography (Kodama, 1981; Hull, 1986; *3D Systems Inc.*, 2019a) which involves the curing of a photosensitive resin using an ultraviolet laser. The technology has evolved and AM is capable of manufacturing objects with complicated internal and external geometries, some examples are shown in Figure 2.1. However, there is a need for product inspection and in particular assessing the quality of the internal structures.

Imaging using x-rays (Röntgen, 1896) has been used in the medical field. In x-ray computed tomography (Cormack, 1973; Hounsfield, 1973, 1980), the patient has an x-ray image taken at multiple angles. These x-ray images are used to reconstruct what was taken in 3D to make a diagnostic.

X-ray computed tomography (XCT) can be used as a non-destructive test for AM products. Various reviews on AM exist such as Kruth (1991); Kruth et al. (1998); Pham and Gault (1998); Gibson et al. (2010); Wong and Hernandez (2012); Ngo et al. (2018). For XCT used in manufacturing, there are Cantatore and Müller (2011); Kruth et al. (2011); Sun et al. (2012). Thompson et al. (2016) reviewed the applications of XCT on AM.

In this chapter, AM is reviewed followed by XCT. The latest research for the use of XCT on AM is reviewed at the end of the chapter.

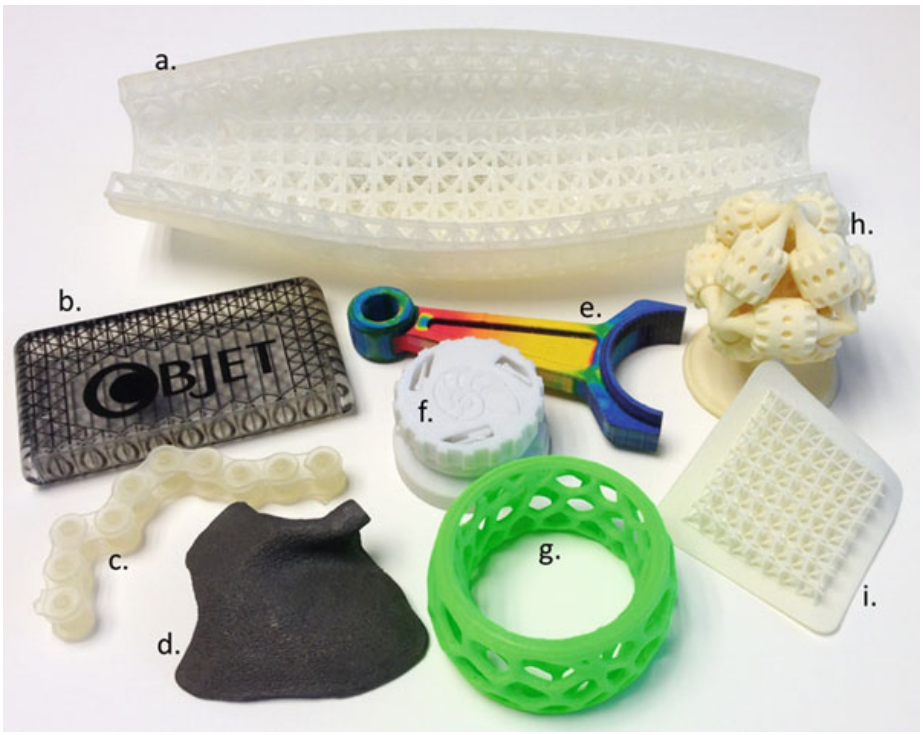


Figure 2.1: Examples of additive manufactured parts: a) lattice structure, b) toy, c) chain, d) model of a facial implant, e) spanner, f) ratchet mechanism, g) toy, h) series of rotatable gears, i) lattice structure. Republished with permission of Springer New York, from Gibson et al. (2010); permission conveyed through Copyright Clearance Center, Inc.

2.1 Additive Manufacturing

Loosely, AM involves solidifying material onto a moving platform so that the object is manufactured layer by layer. Typically, this is a slow and expensive method compared to destructive methods such as computer numerical control (CNC) machining for example. An advantage of AM is that the setup cost is low, in particular, destructive methods require planning and setting up various apparatus before the manufacturing stage (Gibson et al., 2010). This makes it suitable to manufacture bespoke items which achieves the goals of AM's predecessor called rapid prototyping (Kruth, 1991).

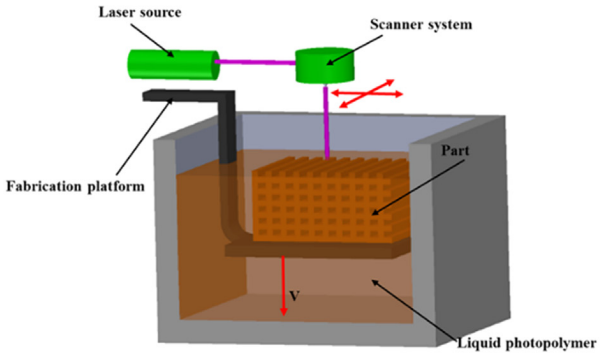
Various AM technologies were invented during the advancement of AM. Because of this, there are various applications of AM, for example in medical and biomedical sciences (Kang et al., 2016; Kourra et al., 2018), engineering (Cooper et al., 2015), food engineering (Godoi et al., 2016) and art (Ornes, 2013; Grossman, 2019).

Additive Manufacturing Technologies

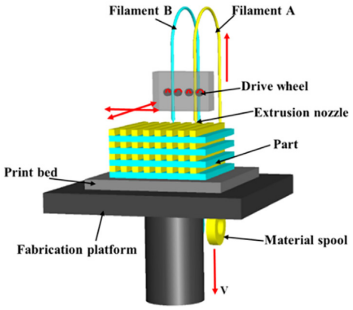
The different AM technologies can be classified based on the apparatus, for example, liquid-based or powder-based, and/or on the method of manufacturing, for example, point by point or layer by layer (Kruth, 1991). The liquid-based AM technologies presented here are stereolithography (Kodama, 1981; Hull, 1986; *3D Systems Inc.*, 2019a) and fused deposition modelling (Crump, 1991, 1992; *Stratasys Ltd.*, 2019). The following powder-based technologies are presented here: 3D printing (Sachs et al., 1990), selective laser sintering (Deckard, 1989; *DTM Corp.*, 1990; *3D Systems Inc.*, 2019a), electron beam melting (Larsson and Larsson, 2004; *Arcam AB*, 2019), laser engineered net shaping (Atwood et al., 1998). Illustrations of these technologies are shown in Figure 2.2.

Stereolithography is a liquid-based AM technology. It consists of a container containing a liquid photo-hardening monomer or polymer as well as a piston and platform which holds and moves the manufactured product up and down. A laser with a specific wavelength, typically 300 nm to 400 nm (Kodama, 1981), is emitted onto a point of the surface of the liquid and solidifies. The laser is controlled by a computer to solidify specific parts of the liquid surface. The platform is lowered and the cycle repeats, manufacturing the object layer by layer. Laser absorption happens a few tenths of a millimetre which corresponds to the thickness of each layer (Kruth, 1991; Pham and Gault, 1998).

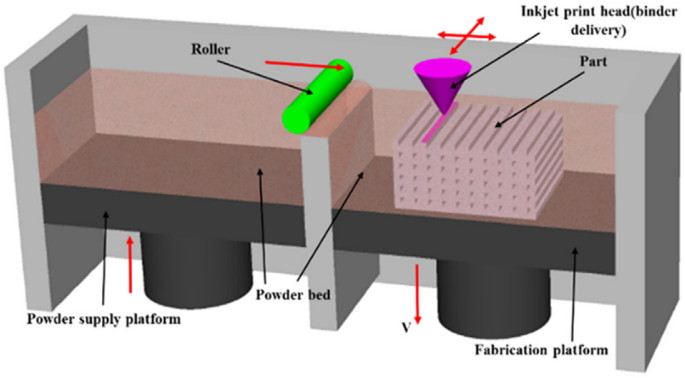
Fused deposition modelling is another liquid-based AM technology. A jetting head, or nozzle, deposits the molten material onto a platform or on top of the previous layer. The material is usually plastic in the form of a thin filament. It is heated to just above its melting points, typically



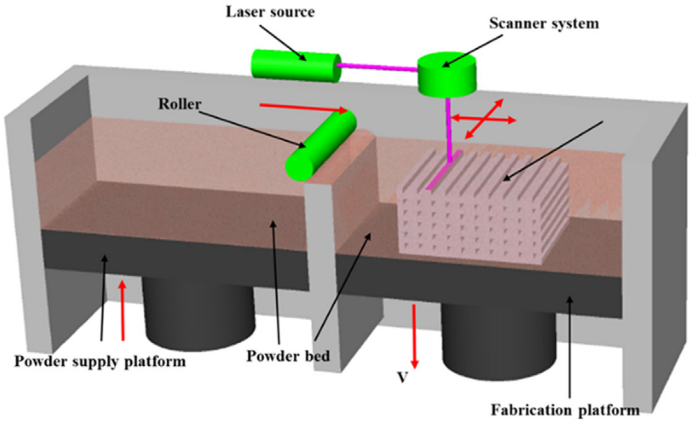
(a) Stereolithography



(b) Fused deposition modelling



(c) 3D printing



(d) Selective laser sintering

Figure 2.2: Diagrams of various AM technologies. Reprinted from Wang et al. (2017a)©, with permission from Elsevier.

1 °C (Crump, 1992), so that it cools down within 0.1 s (Kruth, 1991). The platform moves, and controlled by a computer, in the xy-plane, or left to right and front to back, to produce a layer. The jetting head can move in the z-axis, or up and down, to manufacture the next layer. In the original patent by Crump (1992), the thickness can be as thin as 0.000 1 inches (0.003 mm).

3D printing is a powder-based AM technology. A jetting head deposits a binding agent in droplets onto a bed of powder of ceramic, metal or polymer. The binding agent is cured via evaporation or heating which glues the powder particles together. The jetting head can move in the xy-plane and the object is manufactured layer by layer by moving the platform in the z-axis and renewing the powder using a roller. The binding agent must have a low viscosity so it can be deposited and may also be charged so that it can be deflected using an electric field for precise deposition (Sachs et al., 1990). The thickness of each layer is determined by the size of the droplets of the binding agent, which can be as small as 15 µm in diameter (Sachs et al., 1990). Sachs et al. (1990) reported a tolerance of 0.001 inches (0.03 mm).

Selective laser sintering, electron beam melting and laser engineered net shaping are powder-based AM technology. Selective laser sintering is similar to 3D printing, but instead, a laser is used to sinter or fuse the powder particles in a chamber heated just below the melting point of the material (Wong and Hernandez, 2012). Various materials such as metals and plastics can be used (Wong and Hernandez, 2012). Electron beam melting is similar, but instead of a laser, an electron beam is used. This is done in a high vacuum chamber to avoid oxidation (Wong and Hernandez, 2012). In laser engineered net shaping, a powder bed is not used; the powder is deposited on the desired location and then melted using a laser, as shown in Figure 2.3. This is a popular method to manufacture metal objects (Gibson et al., 2010).

There are many more AM technologies but they can be found in numerous review literature. A comparison of the mentioned AM technologies available at the time was done by Pham and Gault (1998); Kim and Oh

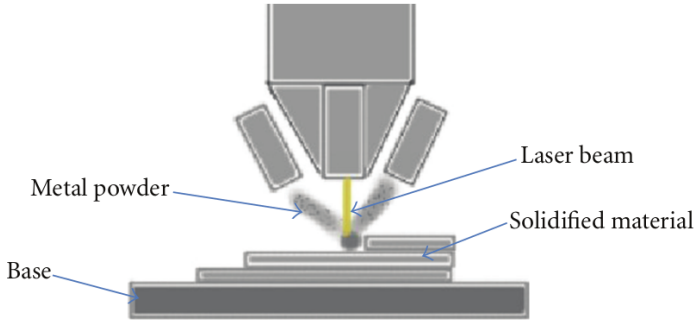


Figure 2.3: Laser engineered net shaping. Reprinted from (Wong and Hernandez, 2012) under the CC BY 3.0 license.

(2008). Factors such as material cost, mechanical properties and the resolution of the manufacturing were considered. There are also safety aspects to assess, for example, powder in powder-based methods can escape into the environment and the liquid used in stereolithography is toxic, sticky and has spilling risk (Kim and Oh, 2008). This makes fused deposition modelling a popular choice and can be used in an office environment (Ngo et al., 2018).

The strength of the manufactured object varies from geometry to geometry but also from direction to direction. Because the manufactured object is made layer by layer, the strength varies if the load was applied in the building direction (vertical) or the scanning direction (horizontal) (Kim and Oh, 2008). Experimental results have shown that fused deposition modelling has superior strength in the scanning direction but weak in the building direction (Kim and Oh, 2008).

The strongest manufacturing methods were found to be powder-based methods and stereolithography, however, they are slow and material costs are high (Kim and Oh, 2008). Fused deposition modelling has low costs and high speeds but suffers from weak mechanical properties (Ngo et al., 2018).

The materials available for each AM technology varies. The materials used in stereolithography is limited because of the use of liquids with photo-hardening properties (Ngo et al., 2018). Fused deposition modelling

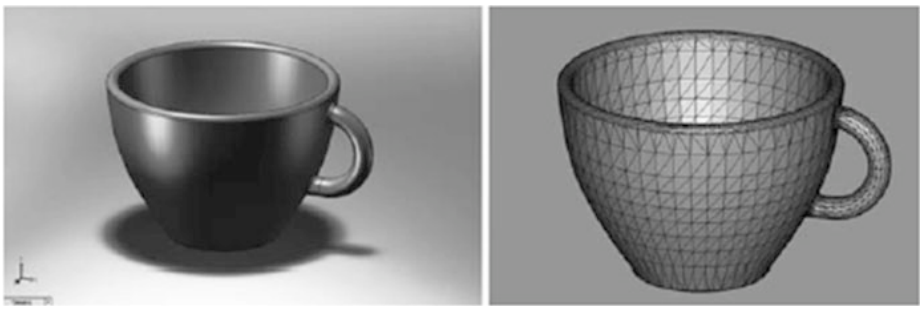


Figure 2.4: An example of a CAD model (left) converted to an STL file (right). Republished with permission of Springer New York, from Gibson et al. (2010); permission conveyed through Copyright Clearance Center, Inc.

is limited to plastics (Ngo et al., 2018). Selective laser sintering and laser engineered net shaping can manufacture objects using metals such as aluminium alloys, steel, titanium and titanium alloys (Herzog et al., 2016).

Pre/Post Processing

The blueprint of the object to be manufactured is called a computer-aided design (CAD) model. For it to be processed by an AM apparatus, the CAD model is converted to an STL file (*3D Systems Inc.*, 1989, 2019b) which represent surfaces by a series of triangles, an example is shown in Figure 2.4. STL stands for stereolithography but could also be called standard tessellation language (Wong and Hernandez, 2012). Some accuracy is lost here as the surface of the CAD model is represented approximately by triangles (Gibson et al., 2010). The STL file is then sliced into layers (Jamieson and Hacker, 1995; Vatani et al., 2009) so that the AM apparatus knows what to build for each layer.

When the AM object is manufactured, post-processing techniques can be done at this stage. For example, sanding may be done to smooth the surfaces (Gibson et al., 2010). The manufactured object may be inspected for pores or defects by comparing the x-ray projection of the object with the CAD model (Lee and Tarbutton, 2015; Villarraga-Gómez et al., 2015;

Kim et al., 2016).

As with any apparatus, regular maintenance is required (Bell, 2014).

Defects and Quality Control

There are various discontinuities in AM. In fused deposition modelling, a staircase effect on the surface of the product arises from poor slicing methods of the CAD model (Weeren et al., 1995). Internal voids can be formed due to insufficient material flow (Weeren et al., 1995). Other factors which can cause defects include misalignment of the platform or nozzle, depletion of material and lack of adhesion due to low temperatures (Günaydin and Türkmen, 2018).

There are also problems in the manufacturing of metal parts (Everton et al., 2016), for example, gas can become trapped during the manufacturing process forming gas pores in the manufactured object (Thijs et al., 2010; Tammas-Williams et al., 2015). These gas pores can be 5 µm to 20 µm in diameter (Everton et al., 2016).

Layers may not fuse and form elongated pores. This can be fixed by increasing the energy of the beam but increasing it too much will cause evaporation of the AM part (Mumtaz et al., 2008). These pores can be 50 µm to 500 µm in size (Everton et al., 2016) and can be observed using a scanning electron microscope as shown in Figure 2.5.

Low wetting ability of the melt pool can cause balling which is where the sintered powder has poor contact on the existing layer causing spherical particles to form on the surface of the AM part (Li et al., 2012; Gu and Shen, 2009). The spherical particles can vary in size of 10 µm to 500 µm (Li et al., 2012). The balling effect can be reduced by ensuring low oxygen content in the environment (Niu and Chang, 1999) and using higher energy beams (Gu and Shen, 2009). Some examples are shown in Figure 2.6 using an electron scanning microscope.

Cracks can form due to extreme temperature changes and gradients (Mercelis and Kruth, 2006; Zaeh and Branner, 2010).

The manufacturing process can be monitored, this is called online or

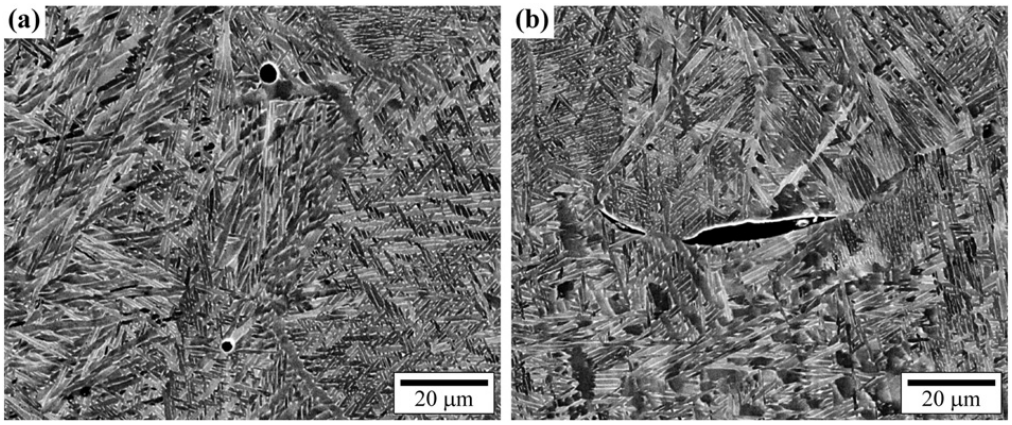


Figure 2.5: A scanning electron microscope image of a) pores and b) elongated pores from an electron beam melting manufactured object. Reprinted from Tammas-Williams et al. (2015) under the CC BY 4.0 license.

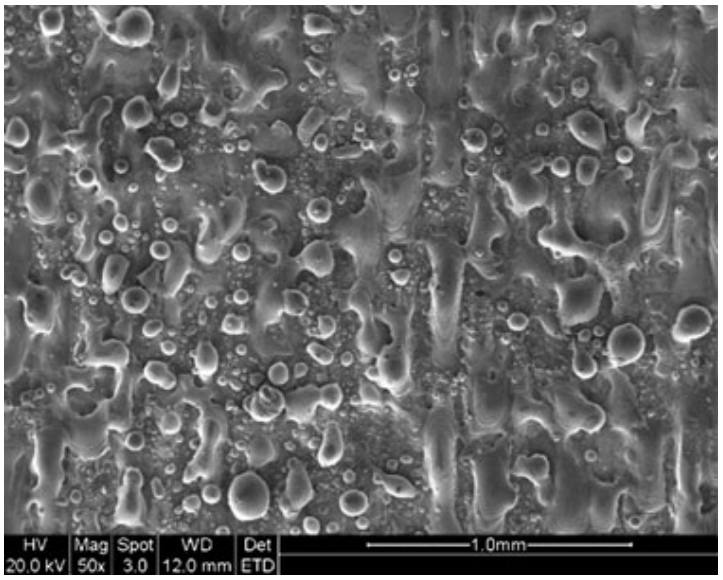


Figure 2.6: A scanning electron microscope image of balling on a selective laser melting manufactured object. Reprinted by permission from Springer Nature: Li et al. (2012)©.

in-situ process monitoring (Everton et al., 2016). The idea is that problems during the manufacturing process are found as soon as possible before the final product is spoiled (Cerniglia et al., 2015). Various methods are used for in-situ process monitoring, for example, a high-speed camera can be installed to capture the various wavelengths in the electromagnetic spectrum emitted by the melt pool (Berumen et al., 2010; Craeghs et al., 2011; Lott et al., 2011). Various discontinuities and errors can be detected (Clijsters et al., 2014) and be used to give feedback to the AM apparatus (Herzog et al., 2013). Other methods include measuring the surface using a laser (Cerniglia et al., 2015) and using an infrared camera to measure the temperature of the melt pool (Rodriguez et al., 2012).

2.2 X-ray Computed Tomography

XCT started its use in the medical field but the advancement of the technology saw its use in manufacturing and metrology, the science of measurement. Applications of XCT include the examination of acetabular hip prosthesis cups (Kourra et al., 2018), skeletons (Appleby et al., 2014), batteries (Taiwo et al., 2017) and materials (Zhang et al., 2016; Wang et al., 2017b). XCT can be used to reverse engineer existing products and improvements can be fabricated using AM, for example, it was used for improving existing hollow engine valves (Cooper et al., 2015). However, the use of XCT in metrology is not yet firmly established compared to other methods of measurement (Thompson et al., 2016). This is because there are a lot of inconsistencies in the setup of XCT apparatuses and on controlling the sources of error.

Concepts from the Medical Field

The setup of XCT (Cormack, 1973; Hounsfield, 1973, 1980) in the medical field involves the patient laying on a flatbed. An x-ray source and x-ray detector pair rotate around and translate along the patient to get readings of the x-rays after attenuating through the patient via different paths. X-

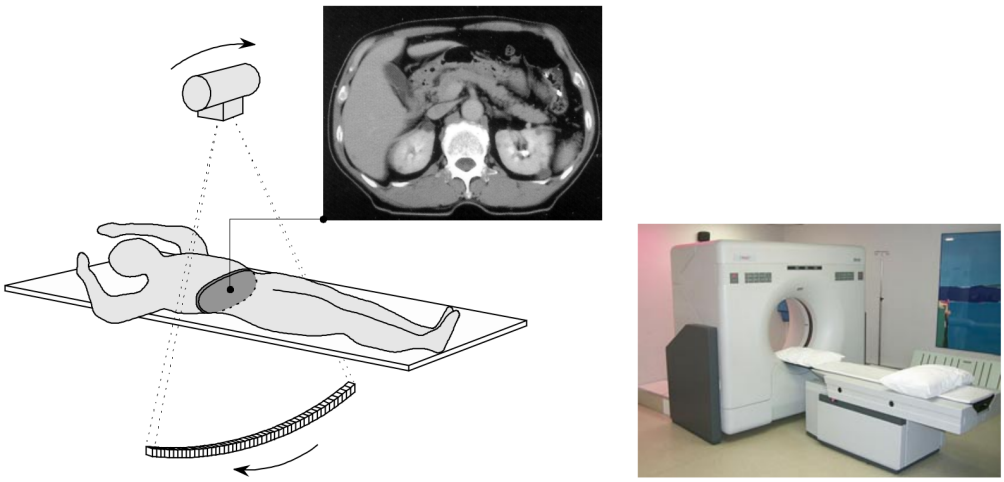


Figure 2.7: In medical XCT, a fan-shaped x-ray beam is emitted and attenuate through the patient and detected by a detector. The x-ray source and detector rotate around and translate along the patient. By collecting readings at different angles, the image of the patient can be reconstructed. Reprinted from Michael (2001). © IOP Publishing. Reproduced with permission. All rights reserved.

ray beams were pencil beams in the early versions of XCT (Michael, 2001). To reduce scanning times, fan-shaped beams and arrays of detectors were used and they can move in a spiral fashion along and around the patient (Cierniak, 2011). These multiple x-ray readings can be used to reconstruct a representation of the patient in 3D (Zeng, 2010). This is illustrated in Figure 2.7.

The patient cannot be exposed to too much radiation, therefore the x-rays used are of low power which can cause noisy readings from the detector. The sources of noise are from the behaviour of the x-rays and the electronics in the detector (Yang et al., 2010). In this realm of low signal to noise ratio, the noise has a compound Poisson element to it (Whiting, 2002; Whiting et al., 2006). Many reconstruction algorithms have been proposed to consider the compound Poisson noise (Elbakri and Fessler, 2002, 2003; Elbakri, 2003; Lasio et al., 2007; Xie, 2008).

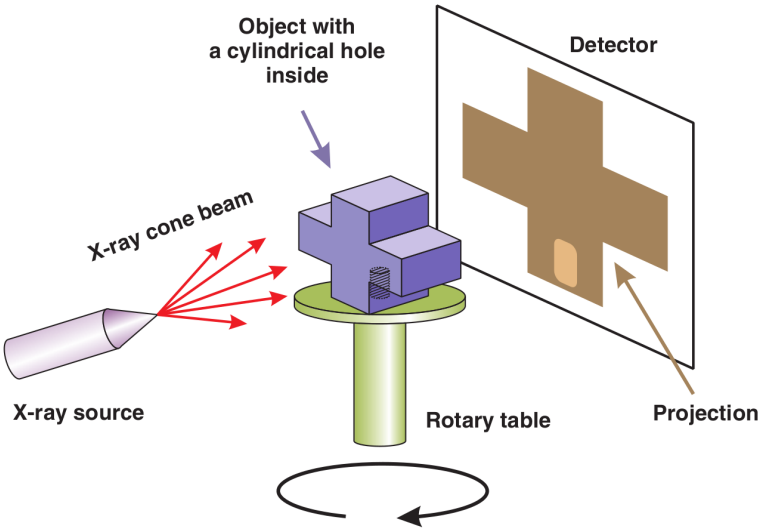


Figure 2.8: The setup of XCT used in metrology. Reprinted from Warnett et al. (2016) under the CC BY 3.0 license.

Acquisition Process in Manufacturing

In manufacturing and metrology, high power x-rays can be used in XCT because there is no consequence of the manufactured object absorbing the radiation. As a result, the XCT setup is different. The object is held by foam on a turntable and placed between an x-ray source and an x-ray detector. X-ray projections are taken while the object rotates. Typically, the x-ray is a cone-beam (Kruth et al., 2011). This is illustrated in Figure 2.8.

The acquisition process consists of the production of x-rays, x-rays attenuating the object, the detection of x-rays and the reconstruction process.

X-rays (Röntgen, 1896) are produced in an x-ray tube, a diagram shown in Figure 2.9. It consists of a vacuum tube containing a cathode and an anode. Electrons are fired from the cathode to the anode due to an electric potential. The cathode is usually tungsten and the anode contains a small amount of tungsten, molybdenum or copper (Sun et al., 2012).

The electrons can interact with the anode in many ways. The electrons can be deflected or decelerated due to the electric field from the nucleus of the target anode material. The energy lost by the electrons is emitted

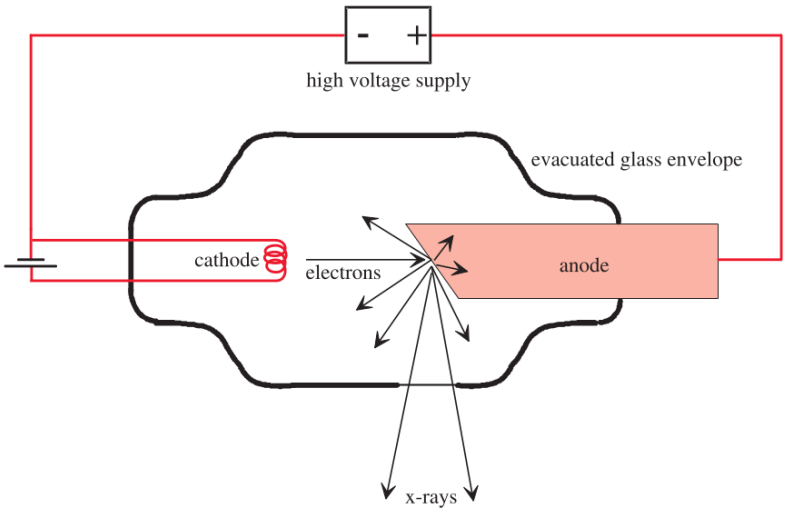


Figure 2.9: An x-ray tube. Reprinted from Michael (2001). © IOP Publishing. Reproduced with permission. All rights reserved.

as bremsstrahlung radiation. The energy of the radiation depends on the potential difference in the x-ray tube, as this determines the energy of the fired electrons, and also the proton number of the anode target because this affects the electric field produced by the nucleus in the anode target (Sun et al., 2012). Another interaction is when the electrons may collide with the nucleus in the anode target, exciting an inner shell electron and ionising it. This produces a vacancy in the electron shell and emits a photon when the excited electron drops down back to the ground state. This is known as characteristic radiation and the energy emitted is discrete and depends on the material in the anode target (Sun et al., 2012).

The efficiency of an x-ray tube is poor. Over 99% of the energy from electrons is converted to heat, the rest to x-rays (Kruth et al., 2011).

Photons, making up the radiation, are emitted from the x-ray tube which can be modelled as a Poisson process (Whiting et al., 2006; Cierniak, 2011). The rate of x-ray emission depends on the current, that is the rate of charge between the cathode and anode. Sources of energy of each photon come from bremsstrahlung radiation and characteristic radiation, making the distribution of x-ray photons energy a mix of continuous and discrete

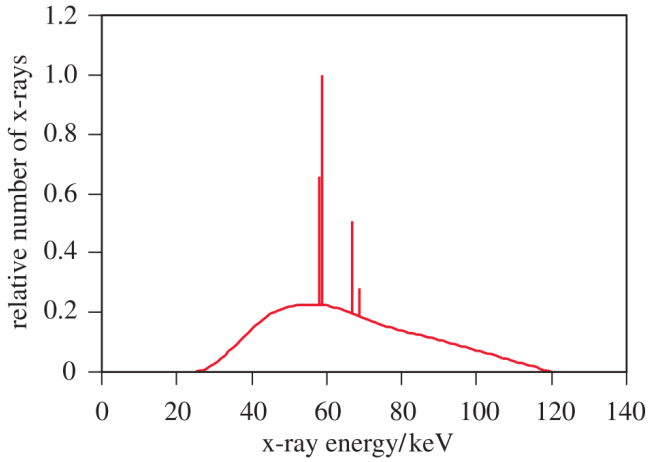


Figure 2.10: An example of the distribution of energies a photon can have emitted from an x-ray tube. Bremsstrahlung and characteristic radiation contribute to the continuous and discrete components of the distribution. Reprinted from Michael (2001). © IOP Publishing. Reproduced with permission. All rights reserved.

energies (Sun et al., 2012). An example is shown in Figure 2.10.

The scanned object is exposed to x-ray photons which undergo attenuation when interacting with the object in several ways (Cantatore and Müller, 2011). The object can absorb the photons via the photoelectric effect. In the photoelectric effect, a photon transfer all of its energy to a bounded electron and ejects it from the atom in the object (Millikan, 1916). Photons can be scattered by the object by colliding inelastically with and transfers its energy to an electron. This process is known as Compton scattering (Compton, 1923). The photoelectric effect and Compton scattering cause several photons to be undetectable. If some of the photons avoid these processes, they are detected with their energy unaffected.

Beer's law simplifies these quantum mechanistic process. Suppose the x-ray beam with a rate of emission I_0 is mono-energetic and travels in a straight line in the x-axis. Let $\mu(x)$ be the attenuation coefficient of the object and the x-ray beam has a rate of emission I after attenuation. A differential equation can be set up to model the decay of photons as it

attenuates through the object such that

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -I\mu(x) \tag{2.1}$$

which can be solved

$$I = I_0 \exp\left[\int_{x\in\text{path of photon}} -\mu(x)\mathrm{d}x\right] . \tag{2.2}$$

However, the photoelectric effect and Compton scattering, thus the attenuation coefficient as well, depends on the energy of the photons (Elbakri and Fessler, 2002). Therefore $\mu(x, E)$ should be made dependent on the energy of the photons (Cantatore and Müller, 2011) and can cause some inaccuracies in Beer's law. In general, low energy photons are more likely to be absorbed and scattered than high energy photons, which increases the average energy of the detected photons (Sun et al., 2012). This is called beam hardening.

After attenuation, the x-ray photons are detected by the x-ray detector. The detectors used in XCT are typically flatbed scanners made up of a scintillator material (Curran, 1953; Greskovich and Duclos, 1997) and photodiodes. The x-ray photons interact with the scintillator material and produce visible light pulses (Rossner et al., 1993). These pulses are detected by photodiodes and converted into an electrical signal (Nikl, 2006; Ren et al., 2018). The electrical signal can be a quantum counter, counting the number of photons detected, or an energy integrating detector, adding up all of the energies of each detected photon (Nikl, 2006; Whiting et al., 2006; Kruth et al., 2011; Ren et al., 2018). The electrical signals are subject to sampling and quantisation to store these signals as an image (Cierniak, 2011). This image is known as a projection.

Not all of the visible light pulses are detected by the photodiodes, thus not all the x-ray photons are detected. The ratio between the number of x-ray photons detected by the detector and the number of x-ray photons arriving at the detector is called the quantum efficiency (Cierniak, 2011; Ren et al., 2018). This makes the detection a two-stage process, converting the x-ray photons into visible light which are then detected (Cierniak, 2011). There exist equipment which detects x-ray directly such as a xenon gas

ionisation detector (Fuchs and Kachelrie, 2000) but this is unrivalled by solid-state CT systems, such as scintillator-photodiodes detectors, which have a high quantum efficiency of about 98% to 99.5% (Hsieh et al., 2000).

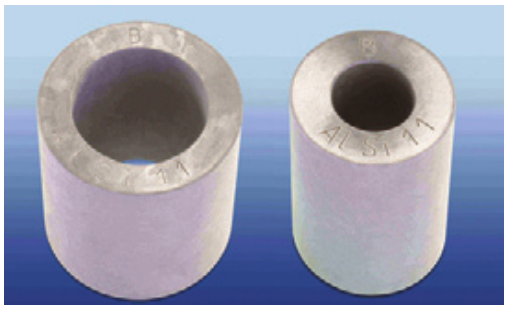
Once projections of the object have been acquired at multiple angles, the reconstruction process can start. The objective of reconstruction is to estimate the attenuation coefficient of the object at each point in space $\mu(x, y, z)$ using the x-ray projections. This is done using the fact that the projections are based on the line integral of the attenuation coefficient along the path of photons. This problem was formed by Radon (1986) as the 'determination of functions from their integral values along certain manifolds'.

A number of reconstruction algorithms in XCT have been developed (Smith, 1990) such as the filtered back-projection (Brooks and Di Chiro, 1976) and the FDK algorithm (Feldkamp et al., 1984). Once the reconstruction has been done, the shape or surface can be extracted by the use of thresholding (Kruth et al., 2011). There are many software packages available for the reconstruction stage of XCT (Reinhart, 2008; Sun et al., 2012).

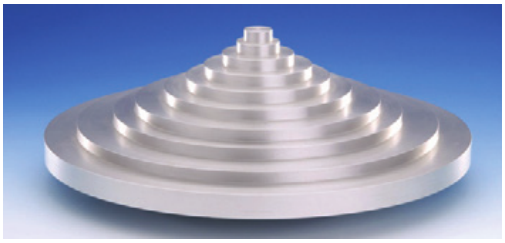
Metrology in Practice

XCT can be used to measure lengths and distances, making it useful for measuring the dimensions of AM objects internally and externally. *Nikon* offer products and services for XCT including features such as direct comparison to the CAD model (*Nikon Metrology NV*, 2015b, 2018b) and automated production line inspection (*Nikon Metrology NV*, 2015a, 2018a).

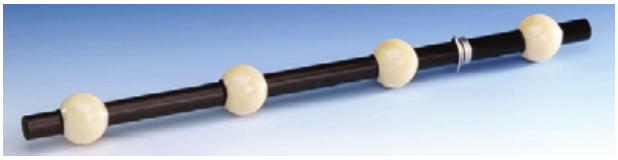
As with a lot of measurement apparatus, calibration is required. In XCT, the scale of each voxel in the reconstruction can be obtained by using XCT on an object with pre-determined lengths, these are known as reference standards (Bartscher et al., 2007) but can have similar names. Reference standards can vary in geometry such as a sphere on a cylinder (Lifton et al., 2013), two spheres on a cylinder (Sun et al., 2016a), a cube with cut-outs (Kiekens et al., 2011), a hollow cylinder, a step-cylinder and



(a) Hollow cylinder



(b) Step-cylinder



(c) Ball-bar

Figure 2.11: Various reference standards: a) aluminium hollow cylinder, outer diameters 30 mm and 20 mm, b) aluminium step-cylinder with diameter 300 mm, c) ceramic balls of diameter 30 mm on a carbon fibre rod, the balls are separated by 100 mm. Reprinted from Bartscher et al. (2007)© with permission from Elsevier.

a ball-bar (Bartscher et al., 2007); the latter three are shown in Figure 2.11.

There are many variables in XCT and a lot of them have to be controlled, for example, XCT should be done in room temperature to avoid any thermal variation (Bryan, 1990), however, this can be hard to do when the x-ray tube is a heat source (Kruth et al., 2011).

The potential difference and current of the x-ray tube can be adjusted to control the contrast and brightness of the x-ray projection. The exposure time is also a factor. These settings should be set high enough to avoid beam extinction but low enough that there is a contrast where less material is present (Kruth et al., 2011).

The magnification can be modified by altering the distances between the x-ray tube, the object and the x-ray detector. Increasing the magnification increases the image resolution but can cause blurry images, this is the result of using an x-ray source with a finite spot size as shown in Figure 2.12

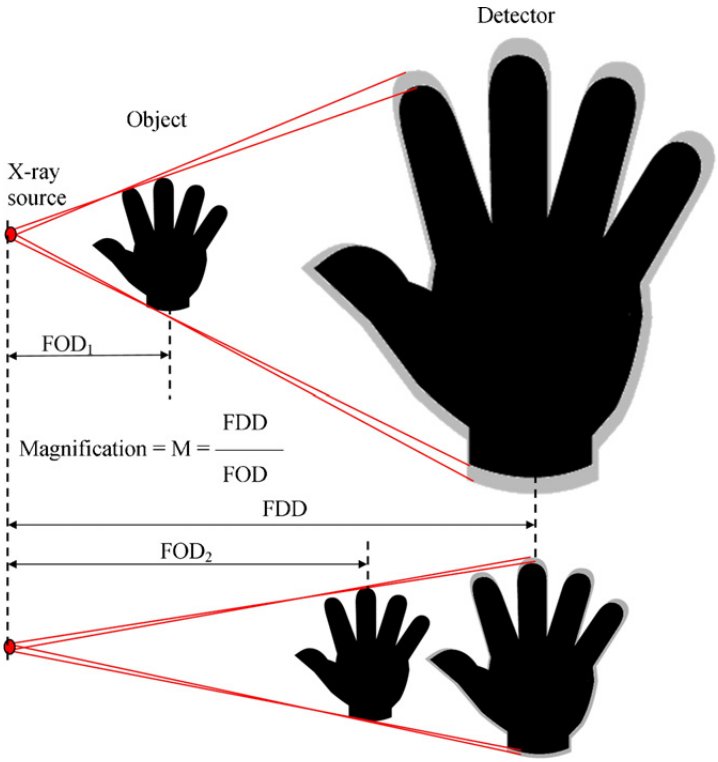


Figure 2.12: The magnification can be tuned by adjusting the distances between the x-ray source, the object and the x-ray detector. Because of a finite x-ray spot size, blurry effects are produced using a magnification too large. Reprinted from Kruth et al. (2011)© with permission from Elsevier.

(Kruth et al., 2011). Larger spot sizes cause more blurry results, this is known as the penumbra effect (Kueh et al., 2016). However, spot sizes too small can produce concentrated heat (Welkenhuyzen et al., 2009) and can damage the x-ray tube.

There is also the question on how to orient the object on the turntable (Corcoran et al., 2016) as well as how many angles to use (Kruth et al., 2011). More angles produce a more accurate reconstruction but require more acquisition time. Figure 2.13 shows an example of a reconstruction using various numbers of angles.

All of the parameters of XCT discussed can be determined before the XCT process by use of simulations (Reisinger et al., 2011; Reiter et al.,

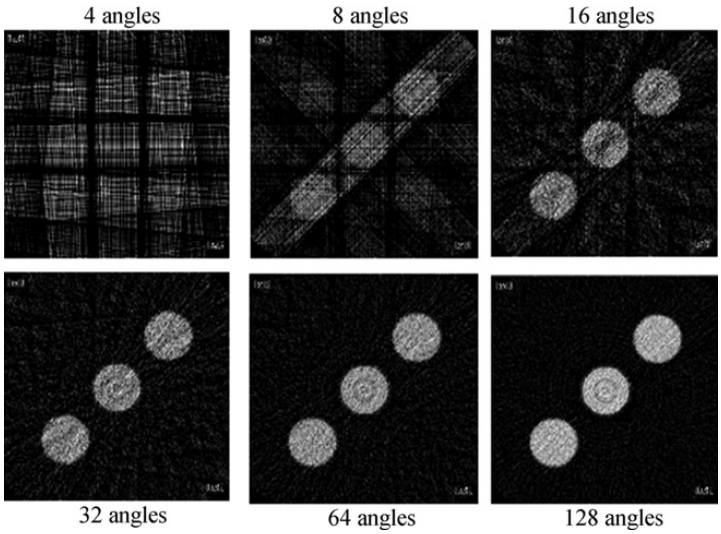
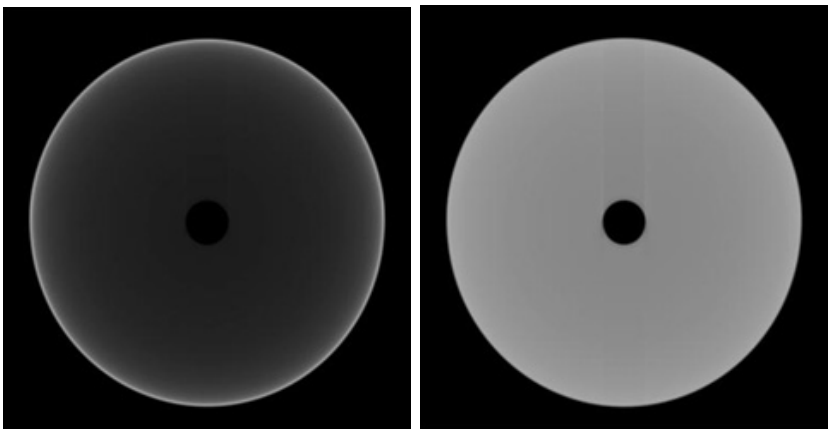


Figure 2.13: A reconstruction when scanning three aligned balls using a different number of angular projections. Reprinted from Kruth et al. (2011)© with permission from Elsevier.

2011), however, there may be inconsistencies. For example, even though the target material of the anode and power is specified, the energy spectrum can still vary (Stumbo et al., 2004).

Problems can occur in the detector, for example, pixels in the acquired projection can be defective or dead (Brettschneider et al., 2014), the panel structure of the detector can be observed (Yang et al., 2009) and the cone-beam appears as a spot. The x-ray spot could be fitted by using a mixture of a Gaussian spot and a uniform spot (Kueh et al., 2016). Another problem is that there exist spatially correlated noise within a projection which can be detected experimentally (Sun et al., 2016c) as well as a correlation between acquisitions, known as image lag (Yang et al., 2009). Precautions can be taken to reduce the impact from image lag such as waiting for 20 minutes between acquisitions (Yang et al., 2010).

Errors due to beam hardening can occur. Low energy photons are more likely to be absorbed or scattered, which causes a few millimetres of the surface of the object to absorb or scatter more photons than the interior. This can cause artefacts (Sun et al., 2016b) such as flat surfaces to be



(a) No filter (b) Al/Cu filter

Figure 2.14: A reconstruction of a hollow cylinder, outer diameter 6.0 mm and inner diameter 0.6 mm. In a), no filter was used. In b) a filter was placed in front of the x-ray tube. Reprinted from (Kruth et al., 2011)© with permission from Elsevier

barrelled and edges rounded off (Kruth et al., 2011). Beam hardening can be tackled by eliminating the low energy photons by placing a filter, a thin metal plate, in front of the x-ray tube (Welkenhuyzen et al., 2009), for example, copper. Figure 2.14 shows an example of reconstruction with and without a filter. Without the filter, the interior of the object appears less dense than it should be. A filter reduces the rate of photon emission but this can be compensated by increasing the exposure time (Kruth et al., 2011). Early reconstruction algorithms ignored beam hardening but modern methods can take beam hardening into account (Elbakri and Fessler, 2001; Sun et al., 2016b).

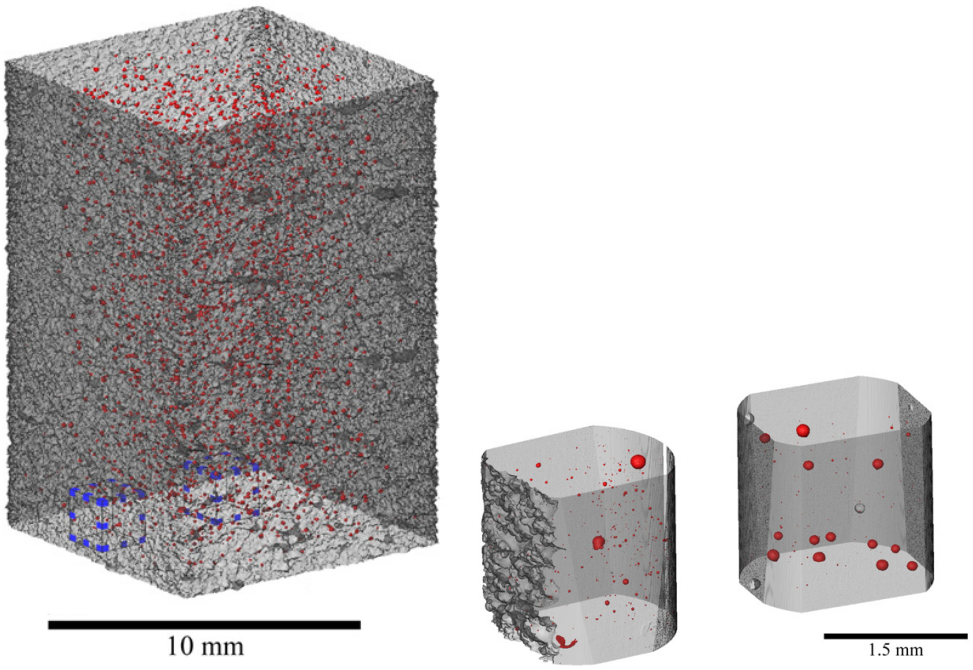
The most common reconstruction method is the FDK (Feldkamp et al., 1984) algorithm because it caters for cone beams. However, it assumes a circular trajectory from the source which can cause artefacts if the trajectory is not circular (Sun et al., 2016b).

Latest Research

The most common use of XCT in AM is the investigation of pores in the manufactured object (Thompson et al., 2016). Pores can be classified as defects if the pores are larger than some volume threshold. This threshold controls the probability of the detection of defects (Gandossi and Annis, 2010; Amrhein et al., 2014).

The porosity is defined by dividing the volume of all of the pores by the volume of solid material (Taud et al., 2005). This can be used to quantified the material's strength and can be measured accurately using Archimedes' method (Spierings et al., 2011). Studies have been done to link porosity to stress concentration (Leuders et al., 2015; Siddique et al., 2015; Carlton et al., 2016) and it was found the location of the pores is a good predictor of fatigue strength (Leuders et al., 2015). XCT can be used to measure porosity and has an advantage over Archimedes' method because the location of the pores can be visualised in XCT. An example of visualising pores is shown in Figure 2.15 and the pores can be compared to the CAD model (Lee and Tarbutton, 2015; Villarraga-Gómez et al., 2015; Kim et al., 2016). In addition to pores, any surface deviation can be measured by aligning the reconstruction with the CAD model and measuring any discrepancies (Lee and Tarbutton, 2015; Villarraga-Gómez et al., 2015; Kim et al., 2016), an example is shown in Figure 2.16.

One of the disadvantages of XCT is that it is a slow process. XCT is not an instantaneous process so progress bars are usually featured in XCT marketing such as *Nikon Metrology NV* (2015a)'s inline quality control. The reconstruction can take between 5 minutes to several hours (Warnett et al., 2016). More angular projections would take more time but will improve the accuracy of the reconstruction (Kruth et al., 2011). Warnett et al. (2016) improved the speed of XCT by sacrificing the accuracy of the reconstruction. This was done by placing the object on a conveyor belt surrounded by multiple x-ray source and detector pairs as shown in Figure 2.17. Fewer angular projections were taken but they can be obtained in one go, speeding up the process.

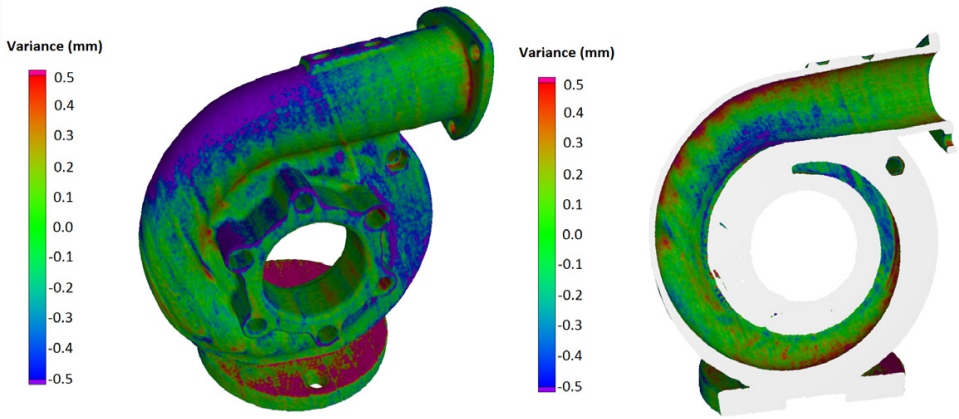


(a) Reconstruction (b) Sample from the reconstruction

Figure 2.15: The reconstruction can show pores in the manufactured object. b) shows the reconstructed samples from the blue cubes in a). Reprinted from Tammas-Williams et al. (2015) under the CC BY 4.0 license.

Instead of reconstructing the object, the analysis can be done on the projections itself, or in projection space, by comparing it to a simulated projection produced by a software called *aRTist* (Bellon and Jaenisch, 2007; Jaenisch et al., 2008; Bellon et al., 2012). It can simulate projections of the object given the specifications of the CT apparatus, such as the x-ray source and the x-ray detector, and the CAD of the object (Bellon et al., 2011; Deresch et al., 2012).

An algorithm was developed to adjust the parameters of the simulation as well as aligning it so that it fits with the x-ray acquisition (Brierley et al., 2018). However, it is very complicated as it is optimising over a very large dimensional space (Brierley et al., 2018). Studies have been conducted comparing simulated projections with each other, one with defects and the other without, by looking at the contrast to noise ratio of the defects. This



(a) External (b) Internal

Figure 2.16: The reconstruction was aligned and compared to the CAD model. The surface heat map shows the surface deviation (or 'variance' in the literature) externally (a) and internally (b). Reprinted from Warnett et al. (2016) under the CC BY 3.0 license.

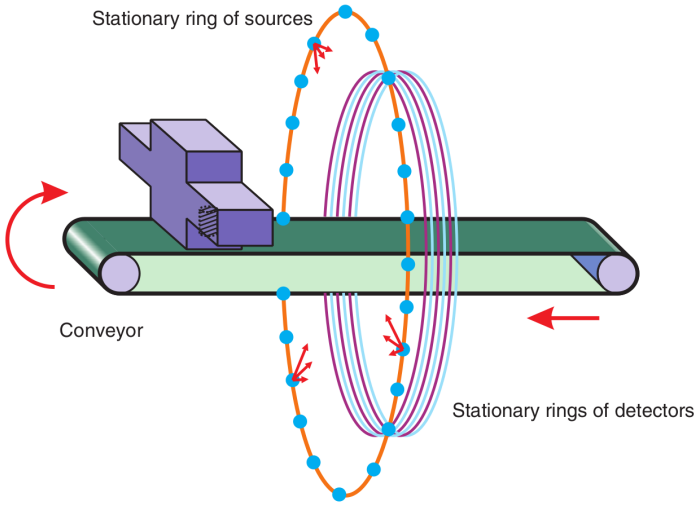


Figure 2.17: XCT can be done on a conveyor belt surrounded by x-ray source and detector pairs. Reprinted from Warnett et al. (2016) under the CC BY 3.0 license.

is done at various projection angles as part of a large optimisation problem (Brierley et al., 2018). Another method is to use machine learning methods to classify defects from a projection (Rale et al., 2009).

It is however inevitable that accuracy is lost from the transition from reconstruction space to projection space, for example, in the diagnostic of pneumonia, a CT scan has superior performance compared to a chest radiograph (Hayden and Wrenn, 2009).

Chapter 3

Data Collection

The objective is to investigate if voids can be detected using a single projection. The quality control procedure can be sped up if defect detection can be done in projection space rather than reconstruction space. An experiment was conducted where a test sample was additively manufactured with purposely designed voids. This was done by comparing a projection of the test sample with the simulation of that projection as if the voids were not there. The simulated projections were produced by using software called *aRTist* (Bellon and Jaenisch, 2007; Jaenisch et al., 2008; Bellon et al., 2012). Any disagreement in the comparison can suggest evidence of a defect.

This chapter describes the apparatus used to manufacture the test sample and obtaining the projections. There is also a discussion, at the end of the chapter, on shading correction which was used to remove panel and x-ray spot effects from the projections.

Many figures presented here were given by engineers concerning the experiment or by the manufacturer. Figures with no error bars were rounded to an appropriate number of significant figures.

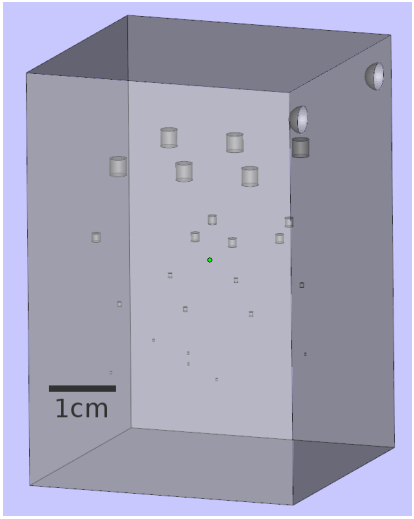


Figure 3.1: The CAD model of the test sample. The scale is approximate.

3.1 Apparatus

The test sample is a cuboid (40.0 mm × 40.0 mm × 60.0 mm) with voids. The voids were of diameters 2.4 mm, 1.2 mm, 0.6 mm and 0.3 mm. 6 voids for each diameter were designed in the CAD model. Voids with diameters 2.4 mm and 0.6 mm were regularly arranged, the other ones were arranged irregularly. The CAD model of the test sample is shown in Figure 3.1.

The *Fortus 400mc (Stratasys, US)* was used to manufacture the test sample made of plastic (acrylonitrile butadiene styrene or ABS). The precision of the manufacturing was in the order of ±0.1 mm (Hanssen, 2013).

X-ray projections were obtained using the *Nikon XT H LC 225/320* x-ray CT scanner (*Nikon Metrology, UK*) together with a *Perkin Elmer XRD 1621 (Perkin Elmer, US)* detector. The target material in the x-ray tube was tungsten. A 0.35 mm copper filter was used to tackle beam hardening. The detector was made up of 2 rows and 16 columns of panels and together has dimensions of 409.6 mm × 409.6 mm which produced a 16-bit projection of 2048 px × 2048 px in size (*PerkinElmer Optoelectronics*, 2006). Therefore, the scale of each pixel is 200.0 µm px^{-1}. The projections were cropped to 2000 px × 2000 px to remove boundary effects. The gain and offset were adjusted by the engineers to produce a projection with good contrast and

negligible penumbra effect. Each pixel has a grey value in units of analogue to digital units (ADU).

Greyscale projections were taken in addition to the projection of the test sample. These are projections with nothing between the source and the detector, obtained with the x-ray tube at different powers. The power was varied by fixing the potential difference and varying the current. The greyscale projections were used for calibration such as shading correction. The greyscale projection with the x-ray turned off is called the black image. The greyscale projection with the x-ray set up the same when obtaining the test sample projection is called the white image.

Replicate test sample projections and greyscale projections were obtained by repeating the acquisition. These replicated projections were used to study the noise observed in the projections.

aRTist was used to simulate the test sample projection and all greyscale projections except for the black image. The black image was simulated by producing a uniform image with a grey value the mean over the obtained black image. The engineers used numerical methods to align the simulated projection with the obtained projection.

3.2 Datasets

Two datasets were collected and named `AbsNoFilter` and `AbsFilter`. They contain replicate projections of the test sample at two different angles, named 30° and 120°, as well as replicate greyscale projections. To investigate the effects of beam hardening, no x-ray filter was used in `AbsNoFilter` and a filter was used in `AbsFilter`.

The properties of each dataset are shown in Table 3.1. These include the properties of the x-ray tube, the XCT apparatus and what powers were used in the greyscale projections. A sample of the obtained, simulated and greyscale projections from the datasets `AbsNoFilter` and `AbsFilter` are shown in Figures 3.2 and 3.3 respectively.

The projections show the test sample, but, with panel and spot effects. The structure of the 32 panels are predominate in the black image, in par-

	Dataset name	
	`AbsNoFilter`	`AbsFilter`
potential difference (kV) of x-ray tube	80	80
power (W) of x-ray tube	36	20
filter	no filter	0.35 mm copper
time exposure (ms)	708	500
distance from source to object (mm)	217	168
distance from source to detector (mm)	1178	876
number of replications for each angle	100	20
powers used for greyscale projections (W)	0.0	0.0
⋮	10	5.0
⋮	18	10
⋮	28	15
⋮	36	20
number of replications for each power	20	20

Table 3.1: Properties of the two datasets obtained for the experiment.

ticular, this was observed by Yang et al. (2009) as well. This is concerning because systematic errors could be introduced as a result of the panel effects in the black image. A black image should be flat because the detector is exposed only to background radiation. The x-ray spot can be observed, in particular, in the white image, and this is the result of using a cone beam.

3.3 Shading Correction

Shading correction, also known as flat field correction, aims to eliminate any spatial variation in sensitivity, observed in the projections, as a result from panel effects, spot effects and other artefacts. Shading correction is done by using the greyscale projections and examining how the grey values respond to different powers for different pixels. For example, the black and white images can be used to correct the projections using

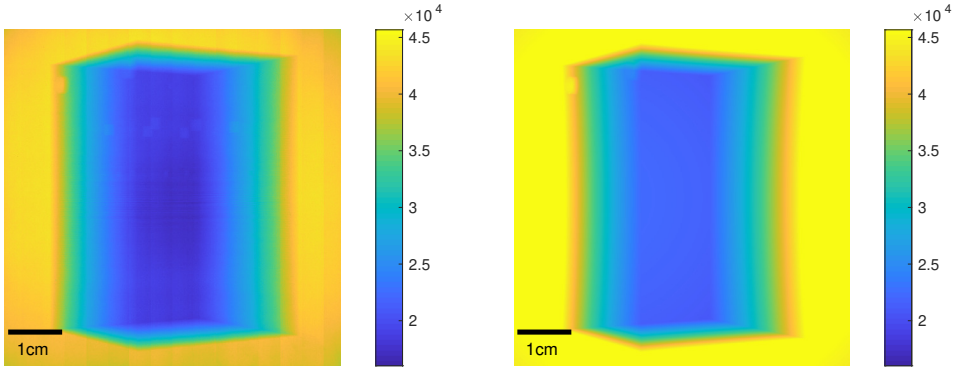
$$U_{x,y} = \frac{N_{x,y} - \text{black}_{x,y}}{\text{white}_{x,y} - \text{black}_{x,y}} \times B + A \tag{3.1}$$

where $N_{x,y}$ is the obtained projection, $U_{x,y}$ is the shading corrected projection and A and B are some user defined constants (Young, 2000; Münzenmayer et al., 2003). This can be extended to include more greyscale projections by modelling the grey value to respond linearly to the power of the x-ray source (Seibert et al., 1998). More generally, shading correction can be expressed as

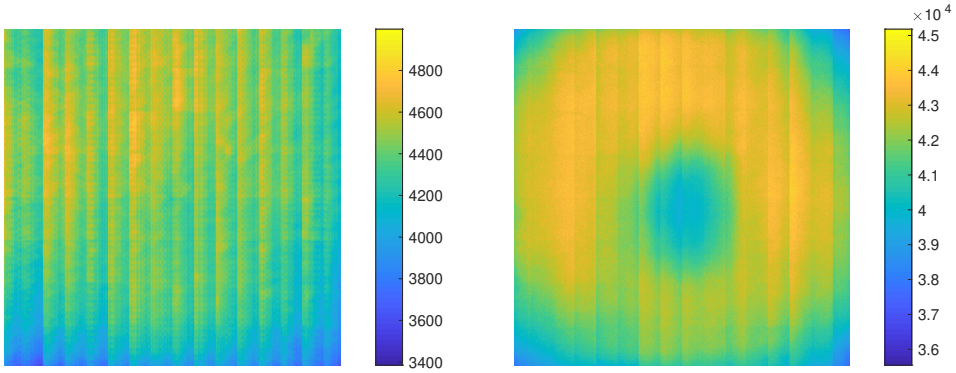
$$U_{x,y} = \beta_{x,y} N_{x,y} + \alpha_{x,y} \tag{3.2}$$

where $\alpha_{x,y}$ and $\beta_{x,y}$ are some spatially varying functions (Münzenmayer et al., 2003). This model has limitations because $\alpha_{x,y}$ and $\beta_{x,y}$ may depend on the energy of each photon, thus beam hardening could cause inaccuracies in shading correction (Davidson et al., 2003). Other methods include minimising the entropy of the projection while constraining $\alpha_{x,y}$ and $\beta_{x,y}$ to be some parametric function (Likar et al., 2000) and using a low pass filter to remove low spatial frequencies from the projections (Young, 2000; Münzenmayer et al., 2003).

In this section, the shading correction in Seibert et al. (1998) is presented in a form without any user-defined constants. The shading correction

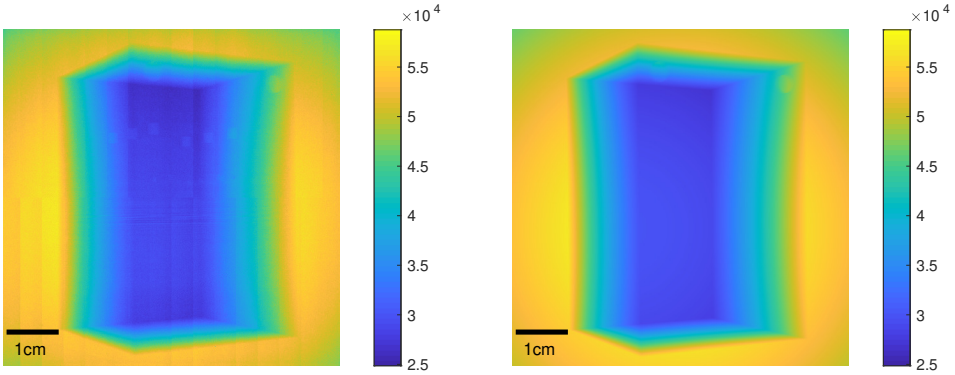


(a) Obtained projection at 36 W (b) Simulated projection at 36 W

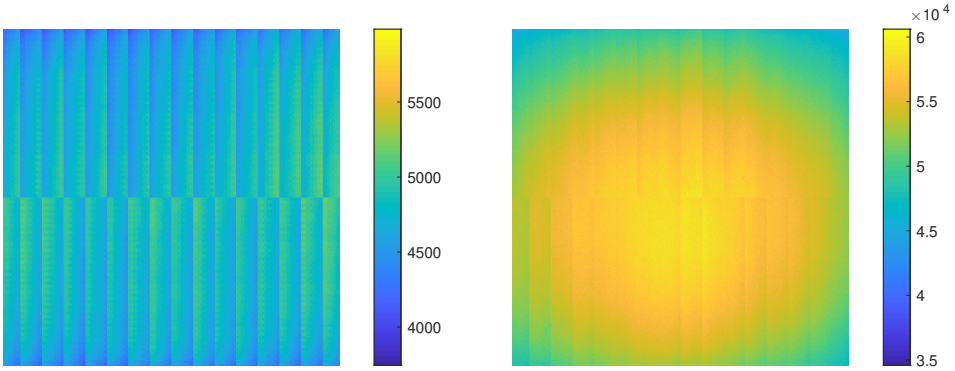


(c) Greyscale projection at 0 W (d) Greyscale projection at 36 W

Figure 3.2: `AbsNoFilter` projections at 30°. The colour scales are in units of ADU.



(a) Obtained projection at 20 W (b) Simulated projection at 20 W



(c) Greyscale projection at 0 W (d) Greyscale projection at 20 W

Figure 3.3: `AbsFilter` projections at 30°. The colour scales are in units of ADU.

was experimented to investigate its performance when shading correcting greyscale projections.

Proposed Shading Correction

Let $S_{x,y}(P)$ be the greyscale projection when exposed to x-rays produced by an x-ray tube with power P for some fixed time exposure τ. The projection, $S_{x,y}(P)$, may be averaged over replications. Let $x = 1, 2, 3, \ldots, W$ and $y = 1, 2, 3, \ldots, H$. Let $N_{x,y}$ be the obtained projection of the test sample when exposed to x-rays produced by an x-ray tube with power P_{proj} for some fixed time exposure τ. The black and white images can be expressed as $\text{black}_{x,y} = S_{x,y}(0)$ and $\text{white}_{x,y} = S_{x,y}(P_{\text{proj}})$ respectively. The power was varied by fixing the potential difference and varying the current of the x-ray tube.

The shading free image is not known, but it is expected that the shading corrected greyscale projection should be flat with some noise. In other words, all pixels in a shading corrected greyscale projection should have grey values with the same expectation $\mu_S(P)$ and same variance $\sigma_S^2(P)$. Suppose $\mu_S(P)$ was estimated using the within projection mean

$$\overline{S}(P) = \frac{1}{\text{WH}} \sum_{x=1}^{W} \sum_{y=1}^{H} S_{x,y}(P) \ . \tag{3.3}$$

Consider a pixel at (x, y), shading correction was done by fitting a linear regression on

$$\overline{S}(P) = \beta_{x,y} S_{x,y}(P) + \alpha_{x,y} \quad (+\varepsilon) \tag{3.4}$$

for $P \in \mathbb{P}$ where $\mathbb{P} = \{0, P_1, P_2, \cdots, P_{\text{proj}}\}$ are the powers used for the greyscale projections. ε is a random variable and an error term, it is included for formality purposes. Let $b_{x,y}$ and $a_{x,y}$ be the estimated parameters of $\beta_{x,y}$ and $\alpha_{x,y}$ from the linear regression respectively. Given a projection $N_{x,y}$, the shading corrected projection $U_{x,y}$ is

$$U_{x,y} = b_{x,y} N_{x,y} + a_{x,y} \ . \tag{3.5}$$

In full, the equations for $b_{x,y}$ and $a_{x,y}$ are given as

$$b_{x,y} = \frac{\sum_{P\in\mathbb{P}}(S_{x,y}(P) - \overline{S}_{x,y})(\overline{S}(P) - \overline{S})}{\sum_{P\in\mathbb{P}}(S_{x,y}(P) - \overline{S}_{x,y})^2} \tag{3.6}$$

and

$$a_{x,y} = \overline{S} - b_{x,y}\overline{S}_{x,y} \tag{3.7}$$

where $\overline{S}_{x,y}$ is the between projection mean

$$\overline{S}_{x,y} = \frac{1}{|\mathbb{P}|}\sum_{P\in\mathbb{P}} S_{x,y}(P) \tag{3.8}$$

and $\overline{S}$ is the global mean

$$\overline{S} = \frac{1}{|\mathbb{P}|}\sum_{P\in\mathbb{P}} \overline{S}(P)\ . \tag{3.9}$$

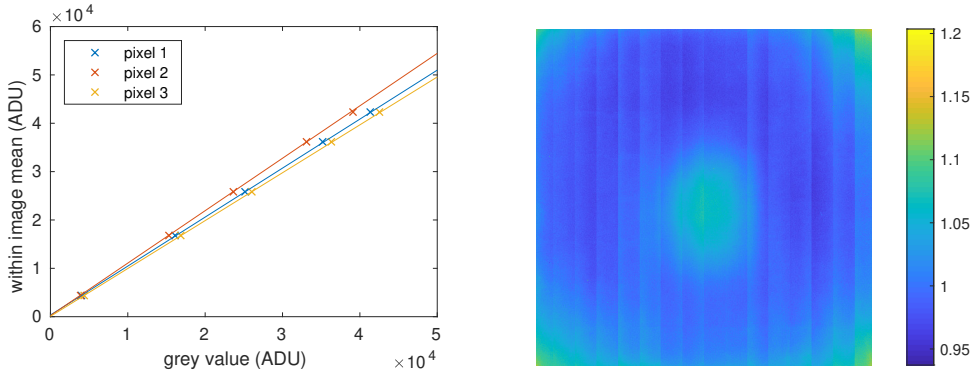
This type of shading correction shall be referred to as linear shading correction. Expressing shading correction in this way has the advantage that there are no user defined constants.

An example of the linear regression is shown in Figure 3.4 where 3 random pixels were chosen for illustration. By plotting the within projection mean versus the grey value for a particular pixel, a linear relationship can be observed. The gradient varied for different pixels which correspond to different sensitivities. The resulting shading correction for the `AbsNoFilter` projection is shown in Figure 3.5. It can be observed that the shading correction removed the panel and spot effects from the background and test sample.

A variation of the shading correction which uses only the black and white images such that $\mathbb{P} = \{0, P_{\text{proj}}\}$ shall be known as black/white (BW) shading correction.

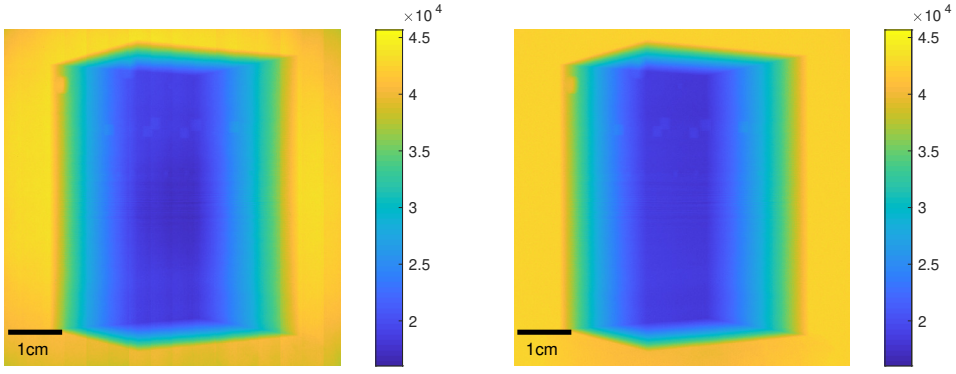
Exploratory Analysis

It was investigated if shading correction on a greyscale projection results in an image which is flat with some noise. To avoid overfitting, one greyscale projection from each power was held out and used to fit the parameters of



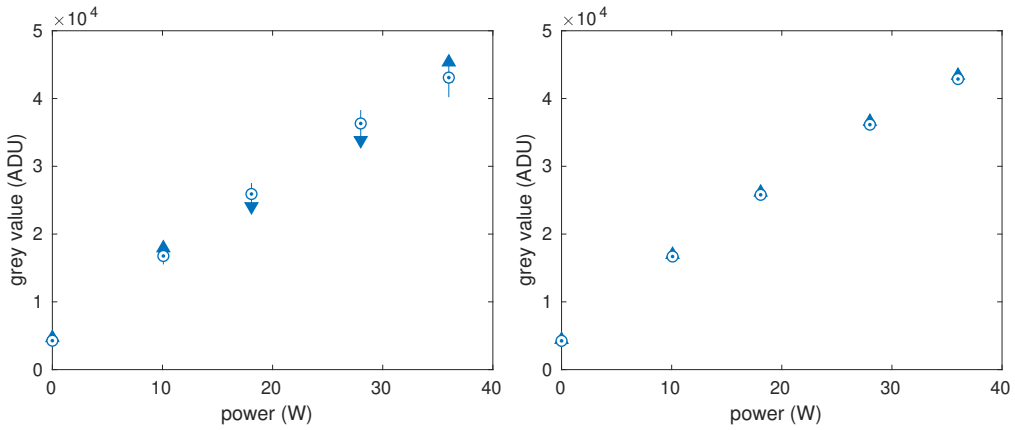
(a) Linear regression (b) $b_{x,y}$

Figure 3.4: For each pixel, a linear regression was fitted on the within projection mean versus the pixel's grey value. The `AbsNoFilter` dataset was used. 3 random pixels were used to illustrate this in a). b) shows the fitted gradient for all pixels.



(a) No shading correction (b) Linear shading correction

Figure 3.5: Projection of `AbsNoFilter` at 30° with and without shading correction. The colour scales are in units of ADU.



(a) No shading correction (b) Linear shading correction

Figure 3.6: Grey values in the greyscale projections before and after shading correction using the `AbsNoFilter` dataset. The boxplots summarise all $2\,000 \times 2\,000$ pixels in a projection.

the shading correction. Shading correction was then used on the unused greyscale projections in this exploratory analysis.

Figure 3.6 shows the grey values in each greyscale projections before and after shading correction. The sensitivity of the pixels, which corresponds to the gradient in units of ADU W^{-1}, became more consistent with shading correction. This should imply that pixels should respond similarly to each other for varying power with shading correction

The `AbsNoFilter` black and white images before and after shading correction are shown in Figure 3.7. The flatness of the image can be shown using a profile plot, this is a plot of the grey values along a column or row, an example shown in Figure 3.8. The figure shows that the shading uncorrected black image was not flat but a remarkable structure was observed by plotting the odd rows and even rows separately. Such a plot shows that the grey values depend on neighbouring pixels, where the majority of pixels on even rows have grey values larger than pixels above and below it. Perhaps this is caused by the read-out structure in the detector.

The BW shading corrected black and white images appeared uniform but there is some structure in the linear shading correction. By giving linear

shading correction various greyscale projections, it attempts to generalise to a range of powers, thus may struggle at shading correcting the black and white images.

ANOVA

An experiment was conducted to quantify the performance of the different types of shading correction. Applying shading correction on the greyscale image should remove effects from the panels and the spot, leaving a flat noisy image with no spatial structure. The variance within a pixel and across replications should be similar to the variance between pixels within a replication if shading correction flattens the greyscale images.

In each dataset, there are 20 replicated greyscale images for each power. One randomly selected replication from each power was assigned to the training set and used to calibrate the shading correction. The remaining images were assigned to the test set and the shading correction was applied to each image.

For each power, the within and between pixel variance was calculated. Let $U_{x,y}^{(j)}(P)$ be the shading corrected greyscale image in the test set with power P for $j = 1, 2, \ldots, m$ replicates. The within and between pixel variance are

$$s_{\mathrm{w}}^2(P) = \frac{1}{WH(m-1)} \sum_{x=1}^{W} \sum_{y=1}^{H} \sum_{j=1}^{m} \left(U_{x,y}^{(j)}(P) - \overline{U}_{x,y}(P) \right)^2 \tag{3.10}$$

and

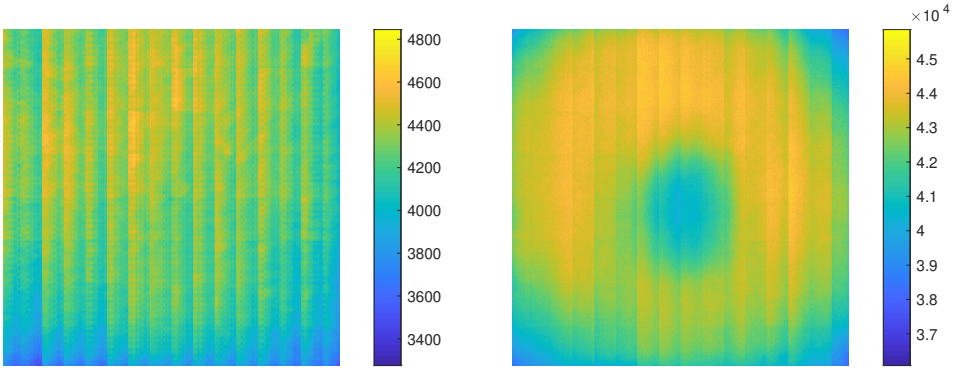
$$s_{\mathrm{b}}^2(P) = \frac{m}{WH-1} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(\overline{U}_{x,y}(P) - \overline{U}(P) \right)^2 \tag{3.11}$$

respectively where

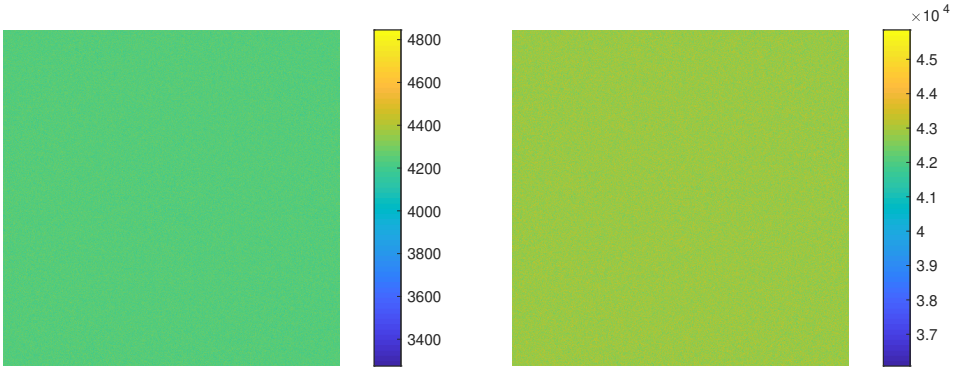
$$\overline{U}_{x,y}(P) = \frac{1}{m} \sum_{j=1}^{m} U_{x,y}^{(j)}(P) \tag{3.12}$$

and

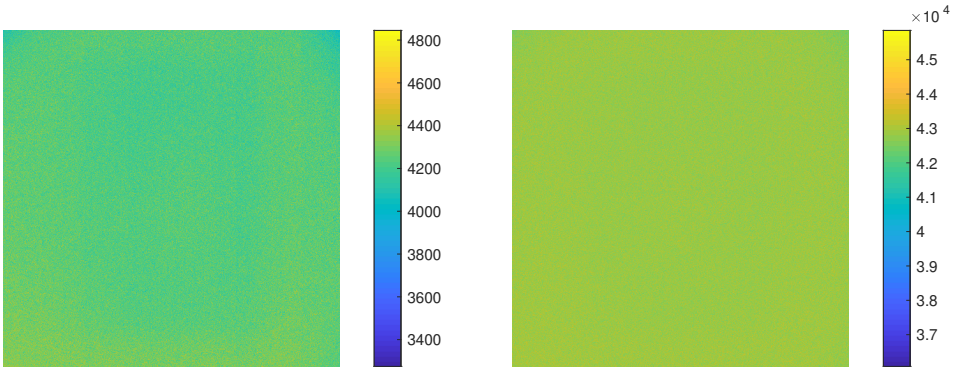
$$\overline{U}(P) = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \overline{U}_{x,y}(P) \ . \tag{3.13}$$



(a) Black - No shading correction (b) White - No shading correction

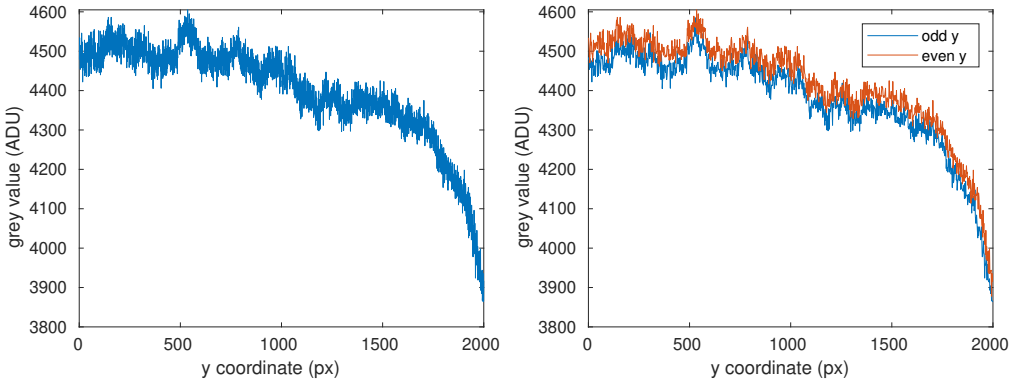


(c) Black - BW shading correction (d) White - BW shading correction

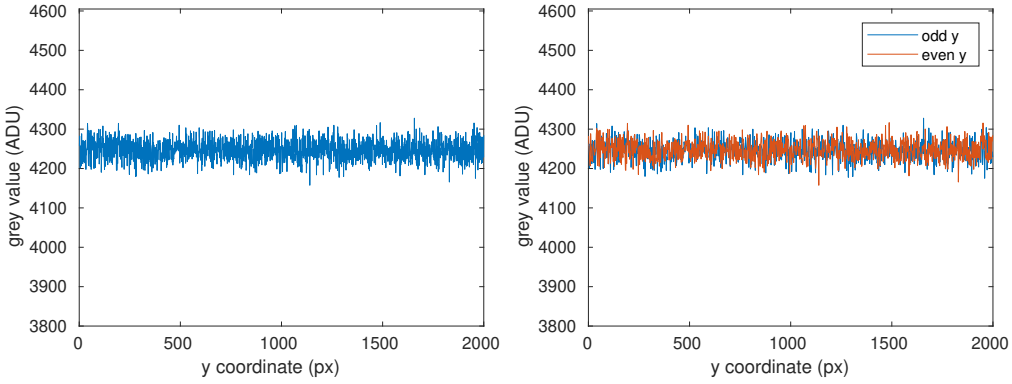


(e) Black - Linear shading correction (f) White - Linear shading correction

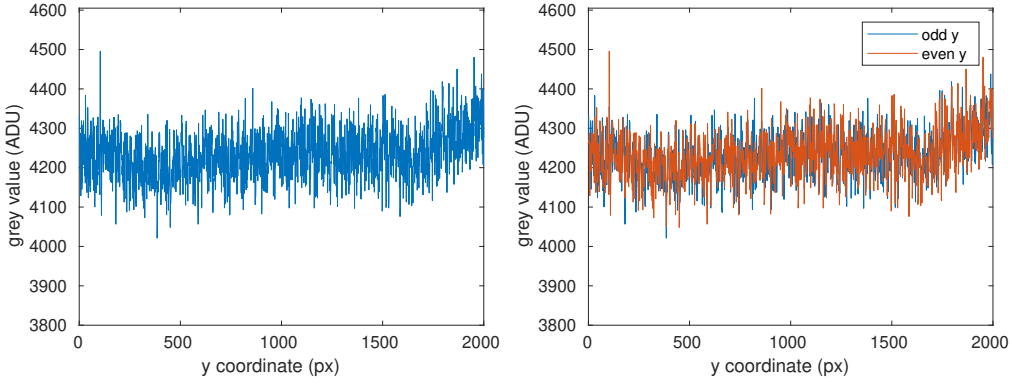
Figure 3.7: The black and white images before and after shading correction from the `AbsNoFilter` dataset. The colour scales are in units of ADU.



(a) No shading correction



(b) BW shading correction



(c) Linear shading correction

Figure 3.8: Left shows the profile plot of an `AbsNoFilter` black image at $(879, y)$ using various shading corrections. Right shows two curves for odd and even y.

In this specific experiment, $W = 2000$, $H = 2000$ and $m = 19$. The F statistic is

$$F(P) = \frac{s_{\mathrm{b}}^2(P)}{s_{\mathrm{w}}^2(P)} \tag{3.14}$$

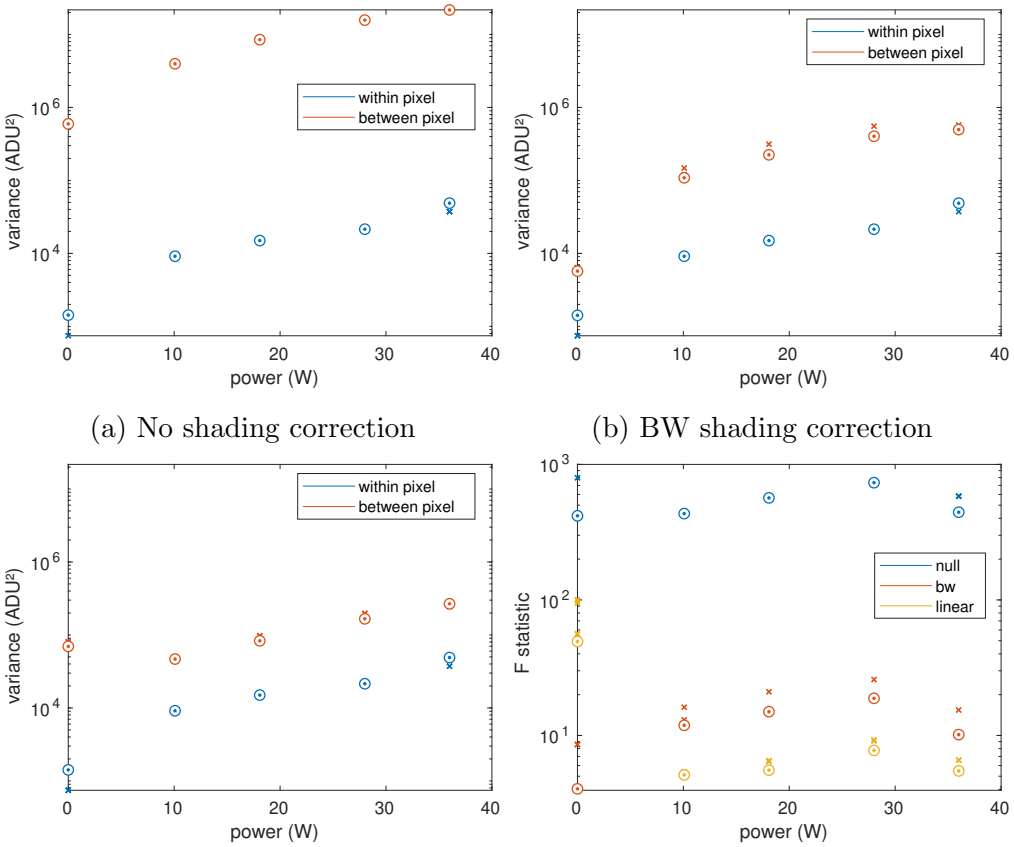
and it should be about one if the within and between pixel variances are similar. The experiment was repeated 100 times by reallocating the training and test set.

The results are shown in Figures 3.9 and 3.10 for the datasets `AbsNoFilter` and `AbsFilter` respectively. With shading correction, the within and between pixel variance became similar. A difference between BW and linear shading correction was that for the black image, the F statistic is closer to one for BW shading correction compared with linear shading correction. For the rest of the powers, linear shading correction outperformed BW shading correction. This shows that linear shading correction generalises to powers between zero and P_{proj}. Since beam extinction is avoided, black grey values in the obtained projection should not be possible. As a result, the linear shading correction is recommended for its good performance for various powers.

Using the F test from ANOVA was found to be too strict in this experiment. Under the hypothesis that the grey values all have the same mean and assume they are Normally distributed and i.i.d., then $F(P) \sim F_{WH-1,WH(m-1)}$. In this experiment, the 5% critical value is 1.001 to 3 decimal places, this is too small for this analysis. This may be due to the grey values of the greyscale projections not satisfying the assumptions for the F test, for example, they may not be i.i.d.

Conclusion

Shading correction is important because it removes any panel and spot effects from projections which may cause systematic errors in any statistical analysis. Shading correcting the greyscale projections visually produced a flat image but there was some spatial variation which was picked up by the between pixel variances in ANOVA.



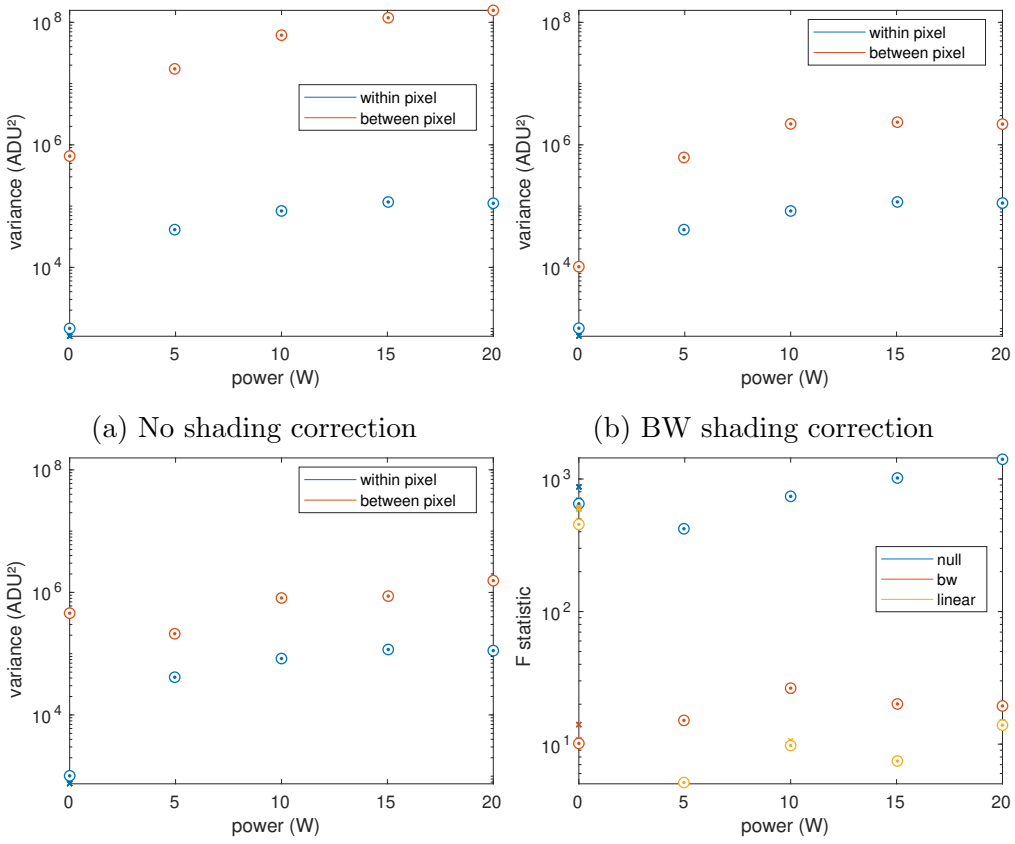
(a) No shading correction

(b) BW shading correction

(c) Linear shading correction

(d) F statistic

Figure 3.9: One randomly selected greyscale image from each power was used to train the shading correction which was then applied to the remaining of the greyscale projections. The within and between pixel variance were estimated and used to calculate a F statistic for each power. The experiment was repeated by reselecting the greyscale projections used for training the shading correction. The boxplots represent the 100 repeats. The `AbsNoFilter` dataset was used here.



(a) No shading correction

(b) BW shading correction

(c) Linear shading correction

(d) F statistic

Figure 3.10: One randomly selected greyscale image from each power was used to train the shading correction which was then applied to the remaining of the greyscale projections. The within and between pixel variance were estimated and used to calculate a F statistic for each power. The experiment was repeated by reselecting the greyscale projections used for training the shading correction. The boxplots represent the 100 repeats. The `AbsFilter` dataset was used here.

Linear shading correction outperformed BW shading correction, in terms of ANOVA, on all greyscale projections except for the black image. This is because linear shading correction trains on greyscale projections of various powers and generalises to these powers. Linear shading correction is used throughout this thesis.

The variance of the greyscale projections was studied in this chapter. This is straightforward because the greyscale projections are flat images. The noise of a projection of a test sample is studied in the next chapter and this was done by modelling the grey value of each pixel as a compound Poisson random variable.

Chapter 4

Compound Poisson

The grey value of each pixel in the detector can be modelled as a random variable, due to the random behaviour of photons being produced, interacting with the test sample and the scintillator in the detector. By modelling using a random variable, the uncertainty can be quantified and considered when conducting inference about any detected defects.

The compound Poisson distribution is studied here because of the compound Poission-like behaviour from the detection of photons (Whiting, 2002; Elbakri and Fessler, 2003; Whiting et al., 2006). It is defined by defining a latent variable $Y \sim \text{Poisson}(\lambda)$ with probability mass function (p.m.f.) $\mathbb{P}(Y = y) = \mathrm{e}^{-\lambda}\frac{\lambda^y}{y!}$ for $y = 0, 1, 2, \ldots$, where $\lambda > 0$ is the Poisson rate parameter. Let U_i be some independent and identically distributed (i.i.d.) latent random variables with probability density function (p.d.f.) $p_U(u)$ for $i = 1, 2, 3, \ldots$. Let X be a compound Poisson random variable where

$$X|Y = \sum_{i=1}^{Y} U_i \ . \tag{4.1}$$

The p.d.f. of X can be obtained by marginalising the joint p.d.f.

$$p_X(x) = \sum_{y=0}^{\infty} p_{X|Y}(x|y)\,\mathbb{P}(Y = y) \quad \text{for } x \geqslant 0 \ . \tag{4.2}$$

It should be noted that $X = 0$ if and only if $Y = 0$ and this happens with probability $\mathbb{P}(Y = 0) = \mathrm{e}^{-\lambda}$. This implies that X has probability mass at

zero and probability density for positive numbers which results in the p.d.f.

$$p_X(x) = \begin{cases} \delta(x)\mathrm{e}^{-\lambda} & \text{for } x = 0 \\ \sum_{y=1}^{\infty} p_{X|Y}(x|y)\mathrm{e}^{-\lambda}\frac{\lambda^y}{y!} \quad \text{for} & \text{for } x > 0 \end{cases} \tag{4.3}$$

where $\delta(x)$ is the Dirac delta function.

The compound Poisson distribution has applications in, for example, modelling rainfall (Revfeim, 1984) and insurance claims (Jørgensen and Paes De Souza, 1994; Smyth and Jørgensen, 2002).

This chapter starts with a literature review on the compound Poisson distribution, how it is derived from the behaviour of photons, how its likelihood is evaluated and methods for fitting it onto data. A model was proposed for the grey values and the expectation-maximisation (EM) algorithm was implemented to fit the model onto data. It was found that for high photon rates, there were identifiability issues. The chapter is concluded on a discussion on why the EM algorithm failed.

4.1 Literature Review

Compound Poisson in X-ray Detection

In an x-ray tube, photons are emitted as a Poisson process (Whiting et al., 2006; Cierniak, 2011) and each photon has some random energy due to bremsstrahlung and characteristic radiation (Sun et al., 2012). This is similar to the compound Poisson distribution. Let Y be the number of photons emitted for some time exposure τ, then $Y \sim \text{Poisson}(\lambda)$. Each photon is assumed to be i.i.d. with random energy U_i for $i = 1, 2, 3, \ldots$ with p.d.f. $p_U(u)$. The random variables discussed here cover all the latent variables in the compound Poisson.

Photons emitted from the x-ray tube undergo attenuation when propagating through the test sample. Assuming no beam hardening, some photons are either absorbed or scattered, making them undetectable. Scattered photons may be detected but it is very rare (Cantatore and Müller, 2011). The energy of the attenuated photons remain unaffected so attenuation

decreases the parameter λ. The amount it decreases by depends on the attenuation coefficient of the material and the amount of material the x-ray attenuates. The density of the random variable U_i remains unchanged because the energy of each photon remains the same after attenuation, assuming no beam hardening.

When the photons interact with the scintillator in the detector, they are converted into visible light. The visible light photons are then detected and converted into a digital signal or grey value. A quantum counter set the digital signal to be linear with the number of photons detected (Whiting et al., 2006). Let X be the grey value observed, then

$$X = bY + \epsilon \tag{4.4}$$

where $\epsilon \sim \mathrm{N}(a, \kappa)$, b and a are some constant and κ is the variance of electronic noise. The mean and variance of the grey value are

$$\mathbb{E}[X] = b\lambda + a \tag{4.5}$$

and

$$\mathbb{V}\mathrm{ar}[X] = b^2\lambda + \kappa \tag{4.6}$$

respectively. By eliminating λ

$$\mathbb{V}\mathrm{ar}[X] = b\,\mathbb{E}[X] + \kappa - ab\ , \tag{4.7}$$

a linear relationship between the variance and expectation of the grey value (Ma et al., 2012) is obtained.

In an energy integrating detector, the recorded grey value is linear to the energy detected (Whiting et al., 2006). The grey value X is

$$X|Y = \sum_{i=1}^{Y} U_i + \epsilon\ . \tag{4.8}$$

This is the compound Poisson distribution with Normal noise added to it. The scale factor b is not included as this can be absorbed into U. Using the result that $\mathbb{E}[X] = \mathbb{E}\,\mathbb{E}[X|Y]$ and $\mathbb{V}\mathrm{ar}[X] = \mathbb{V}\mathrm{ar}\,\mathbb{E}[X|Y] + \mathbb{E}\,\mathbb{V}\mathrm{ar}[X|Y]$, the mean and variance of the grey value are

$$\mathbb{E}[X] = \lambda\,\mathbb{E}[U] + a \tag{4.9}$$

and

$$\mathbb{V}\mathrm{ar}\left[X\right] = \lambda\,\mathbb{E}\left[U^2\right] + \kappa \tag{4.10}$$

respectively. Eliminating λ obtains

$$\mathbb{V}\mathrm{ar}\left[X\right] = \frac{\mathbb{E}\left[U^2\right]}{\mathbb{E}\left[U\right]}\,\mathbb{E}\left[X\right] + \kappa - a\frac{\mathbb{E}\left[U^2\right]}{\mathbb{E}\left[U\right]}\ . \tag{4.11}$$

By assuming no beam hardening, $\mathbb{E}\left[U\right]$ and $\mathbb{E}\left[U^2\right]$ remains constant, thus there is a linear relationship between the variance and expectation of the grey value (Yang et al., 2009). There are other types of detection schemes (Whiting et al., 2006) but it shall not be considered here.

Experiments have been done to verify the compound Poisson nature of the detector. This was done by investigating the variance of radiographs of air (Hsieh et al., 2015) and a polyethene cylinder (Yang et al., 2009, 2010) using different exposures. It was found there were two components in the noise, one was signal-dependent and comes from the compound Poisson distribution, the other was signal independent and maybe electronic noise. The electronic noise can be modelled using a Normal distribution (Xu and Tsui, 2009).

Moment Generating Function

Returning to the compound Poisson distribution with no electronic noise $X|Y = \sum_{i=1}^{Y} U_i$, let the moment generating function (m.g.f.) of X be $M_X(\theta) = \mathbb{E}\left[\mathrm{e}^{X\theta}\right]$. It can be shown that the m.g.f. is

$$M_X(\theta) = \exp\left[\lambda\left(M_U(\theta) - 1\right)\right] \tag{4.12}$$

(Gatto, 2010). The derivation is shown in Appendix B.1. Moments of X can be obtained from the m.g.f. by differentiating it and setting θ to zero. In other words, $\mathbb{E}[X^r] = M_X^{(r)}(0)$. Then it can be shown that

$$\mathbb{E}\left[X\right] = \lambda\,\mathbb{E}\left[U\right] \tag{4.13}$$

$$\mathbb{V}\mathrm{ar}\left[X\right] = \lambda\,\mathbb{E}\left[U^2\right] \tag{4.14}$$

$$\mathbb{E}\left[(X - \mathbb{E}[X])^3\right] = \lambda\,\mathbb{E}\left[U^3\right] \tag{4.15}$$

which agrees with the expectation and variance results in the previous section. The derivation is shown in Appendix B.1.

Compound Poisson-Gamma Distribution

A special case of the compound Poisson distribution is when $U \sim \text{Gamma}(\alpha, \beta)$ where $\alpha > 0$ is the gamma shape parameter and $\beta > 0$ is the gamma rate parameter. This was used for example in Xu and Tsui (2009). The distribution is known as the compound Poisson-gamma distribution and is denoted by $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$. The p.d.f. is

$$p_X(x) = \begin{cases} \delta(x)\mathrm{e}^{-\lambda} & \text{for } x = 0 \\ \sum_{y=1}^{\infty} \dfrac{\beta^{y\alpha}}{\Gamma(y\alpha)} x^{y\alpha-1}\mathrm{e}^{-\beta x}\mathrm{e}^{-\lambda}\dfrac{\lambda^y}{y!} & \text{for } x > 0 \end{cases} . \tag{4.16}$$

Recall that $X|Y = \sum_{i=1}^{Y} U_i$ which involves a sum of gamma random variables. It can be shown that

$$X|Y \sim \text{Gamma}(Y\alpha, \beta) . \tag{4.17}$$

The m.g.f. of U is $M_U(\theta) = \left(\dfrac{\beta}{\beta - \theta}\right)^{\alpha}$, then the m.g.f. of X is

$$M_X(\theta) = \exp\left[\lambda\left(\left(\frac{\beta}{\beta - \theta}\right)^{\alpha} - 1\right)\right] \tag{4.18}$$

and moments can be obtained from it such as

$$\mathbb{E}[X] = \frac{\alpha\lambda}{\beta} \tag{4.19}$$

$$\mathbb{V}\text{ar}[X] = \frac{\alpha(\alpha+1)\lambda}{\beta^2} \tag{4.20}$$

and

$$\mathbb{E}\left[(X - \mathbb{E}[X])^3\right] = \frac{\alpha(\alpha+1)(\alpha+2)\lambda}{\beta^3} . \tag{4.21}$$

Generalised Linear Model

It can be shown that the compound Poisson-gamma distribution is in the exponential family for fixed α (Jørgensen, 1987). To show this, the compound Poisson-gamma distribution was parametrised using the following:

$$p = \frac{2 + \alpha}{1 + \alpha} , \tag{4.22}$$

$$\mu = \frac{\lambda\alpha}{\beta} \ , \tag{4.23}$$

$$\phi = \frac{\alpha + 1}{\beta^{2-p}(\lambda\alpha)^{p-1}} \ . \tag{4.24}$$

The parameters p, μ and ϕ are called the index, mean and dispersion parameters respectively and take the values of $1 < p < 2$, $\mu > 0$ and $\phi > 0$. It can be shown that the p.m.f. at zero is

$$\mathbb{P}(X = 0) = \exp\left[-\frac{\mu^{2-p}}{\phi(2-p)}\right] \tag{4.25}$$

and the p.d.f. for $x > 0$ is

$$p_X(x) = \exp\left[\frac{1}{\phi}\left(x\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right]\frac{1}{x}\sum_{y=1}^{\infty} W_y(x, p, \phi) \tag{4.26}$$

where

$$W_y = W_y(x, p, \phi) = \frac{x^{y\alpha}}{\phi^{y(1+\alpha)}(p-1)^{y\alpha}(2-p)^y y!\Gamma(y\alpha)} \ . \tag{4.27}$$

The derivation is shown in Appendix B.2. This is in the form of a distribution in the dispersive exponential family (Nelder and Wedderburn, 1972; Nelder and Baker, 1972; McCullagh, 1984) for fixed p.

Parameter estimation for known p can be done via the generalised linear model framework and can be extended to include linear mixed models (Zhang, 2013). Estimating p is difficult and various methods were discussed (Zhang, 2013). One way is to estimate μ and ϕ on a grid of p's and then select the p which maximises the likelihood (Dunn and Smyth, 2005).

One special property of the compound Poisson-gamma distribution is that it is in the Tweedie dispersion exponential family (Jørgensen, 1987). It can be shown that it has a special variance-mean relationship

$$\mathbb{V}\mathrm{ar}[X] = \phi\mu^p \tag{4.28}$$

where $1 < p < 2$. This is derived in Appendix B.2. It should be noted that this relationship is for fixed p and ϕ. This is different from the linear variance-mean relationship found at the start of the chapter which was for fixed α and β from assuming no beam hardening.

Method of Moments

The method of moments is a simpler method to estimate the parameters of a compound Poisson-gamma distribution. Suppose $\widehat{\mu}_1$ is an estimator of $\mathbb{E}[X]$ and $\widehat{\mu}_j$ is an estimator of $\mathbb{E}\left[(X - \mathbb{E}[X])^j\right]$ for $j = 2, 3$, then the estimators

$$\widehat{\lambda} = \frac{\widehat{\mu}_1^2 \widehat{\mu}_2}{2\widehat{\mu}_2^2 - \widehat{\mu}_1 \widehat{\mu}_3} \tag{4.29}$$

$$\widehat{\alpha} = \frac{2\widehat{\mu}_2^2 - \widehat{\mu}_1 \widehat{\mu}_3}{\widehat{\mu}_1 \widehat{\mu}_3 - \widehat{\mu}_2^2} \tag{4.30}$$

$$\widehat{\beta} = \frac{\widehat{\mu}_1 \widehat{\mu}_2}{\widehat{\mu}_1 \widehat{\mu}_3 - \widehat{\mu}_2^2} \tag{4.31}$$

are method of moments estimators of λ, α and β respectively (Withers and Nadarajah, 2011), this is shown in Appendix B.3. These estimators suffer because estimation is not done through the sufficient statistics and can be negative. This is a problem because the parameters do not take non-positive values.

Normal Approximation

The evaluation of the p.d.f. of a compound Poisson-gamma distribution is useful so that the likelihood can be obtained. The likelihood can be used to find, for example, maximum likelihood estimators. A problem occurs when dealing with the infinite sum in the p.d.f. because it cannot be simplified. There are several approximations or computational methods to evaluate the p.d.f. such as Fourier inverting the characteristic function (Dunn and Smyth, 2008), using the saddlepoint approximation (Daniels, 1954) or cleverly sum over certain terms in the infinite sum (Dunn and Smyth, 2005). Monte Carlo methods can be used to evaluate the p.d.f. by simulating compound Poisson-gamma random variables.

The m.g.f. provides a starting point to what limiting distributions the compound Poisson-gamma distribution converges to for large parameters. These limiting distributions can be used to approximate the p.d.f. of the compound Poisson-gamma distribution.

It can be shown for large λ, the Normal approximation (Shevtsova, 2014) is

$$X \sim \mathrm{N}\left(\frac{\lambda\alpha}{\beta}, \frac{\lambda\alpha(\alpha+1)}{\beta^2}\right) . \tag{4.32}$$

The solution is shown in Appendix B.4.

Saddlepoint Approximation

The saddlepoint approximation (Daniels, 1954) uses the approximate solution to inverting the Laplace transformation or the m.g.f. This has been used in XCT (Elbakri, 2003; Elbakri and Fessler, 2003). Inverting the Fourier transformation of the characteristic function also gives the p.d.f. using computational methods (Dunn and Smyth, 2008). The saddlepoint approximation will be studied here.

For a given m.g.f. $M_X(\theta)$, the saddlepoint approximation (Daniels, 1954; Butler, 2007) finds an approximate p.d.f. $p_X(x)$. The saddlepoint approximation is given as

$$p_X(x) \approx \left(2\pi K_X''(s)\right)^{-1/2} \exp\left[K_X(s) - sx\right] \tag{4.33}$$

where $K_X(\theta) = \ln\left(M_X(\theta)\right)$ and $s = s(x)$ is the solution to the saddle point equation $K_X'(s) = x$.

For the compound Poisson-gamma distribution, the saddle point approximation (Jensen, 1991) is given as

$$p_X(x) \approx \frac{(\lambda\alpha\beta^\alpha)^{\frac{1}{2(\alpha+1)}}\,\mathrm{e}^{-\lambda}}{\sqrt{2\pi(\alpha+1)}} x^{-\frac{\alpha+2}{2(\alpha+1)}} \mathrm{e}^{-x\beta} \exp\left[x^{\frac{\alpha}{\alpha+1}} \frac{(\lambda\beta^\alpha)^{\frac{1}{\alpha+1}}(\alpha+1)}{\alpha^{\frac{\alpha}{\alpha+1}}}\right] \quad \text{for } x > 0 \tag{4.34}$$

with the derivation shown in Appendix B.5. The approximation is not well defined for $x = 0$.

The integral of the density approximation over the support may not equal to one and it can be numerically re-normalised if necessary. Thus, it may be more sensible to write the approximation up to a constant

$$p_X(x) \propto x^{-\frac{\alpha+2}{2(\alpha+1)}} \mathrm{e}^{-x\beta} \exp\left[x^{\frac{\alpha}{\alpha+1}} \frac{(\lambda\beta^\alpha)^{\frac{1}{\alpha+1}}(\alpha+1)}{\alpha^{\frac{\alpha}{\alpha+1}}}\right] . \tag{4.35}$$

Series Evaluation

The infinite sum, $\sum_{y=1}^{\infty} W_y$, can be computationally summed cleverly to evaluate the p.d.f. This was done by summing only large terms in the sum and ignoring small terms (Dunn and Smyth, 2005). Dunn and Smyth (2005) approximated the sum by truncation

$$\sum_{y=1}^{\infty} W_y \approx \sum_{y=y_{\mathrm{l}}}^{y_{\mathrm{u}}} W_y \tag{4.36}$$

where $y_{\mathrm{l}} < y_{\max} < y_{\mathrm{u}}$ and $y_{\max}$ is the value of y which maximises W_y. Dunn and Smyth (2005) used Stirling's approximation to find that

$$y_{\max} \approx \frac{x^{2-p}}{\phi(2-p)} \tag{4.37}$$

by treating W_y as a continuous and differentiable function of y. The derivation is shown in Appendix B.6. Because values of y are positive integers, it would be appropriate to round $y_{\max}$ accordingly

$$y_{\max} = \max\left[1, \operatorname{round}\left(\frac{x^{2-p}}{\phi(2-p)}\right)\right] . \tag{4.38}$$

The limits, y_{l} and y_{u}, can be chosen such that $W_{y_{\mathrm{l}}}$ and $W_{y_{\mathrm{u}}}$ are less than $\epsilon W_{y_{\max}}$ where ϵ is some small constant, for example, $\epsilon = \mathrm{e}^{-37}$ will be better than machine precision in 64 bits (Dunn and Smyth, 2005). To prevent overflow problems, it is advised to calculate each term in the summation in log scale (Dunn and Smyth, 2005) by using the equation

$$\ln\left[\sum_{y=y_{\mathrm{l}}}^{y_{\mathrm{u}}} W_y\right] = \ln\left(W_{y_{\max}}\right) + \ln \sum_{y=y_{\mathrm{l}}}^{y_{\mathrm{u}}} \exp\left[\ln\left(W_y\right) - \ln\left(W_{y_{\max}}\right)\right] . \tag{4.39}$$

4.2 Simulation Studies on Density Evaluation

Simulations of a compound Poisson-gamma random variable were conducted to compare how well these density evaluation methods performed. This was done by comparing the evaluated densities using the histogram

of simulations and a Q-Q plot. The following compound Poisson-gamma distributions were used in the simulations to capture the variety in the compound Poisson-gamma family: $\text{CP}\Gamma(1,1,1)$, $\text{CP}\Gamma(1,100,1)$, $\text{CP}\Gamma(10,1,1)$, $\text{CP}\Gamma(100,100,1)$. For each of these distribution, 1 000 random variables were simulated. Varying β is not interesting as this only scales the random variable.

There were a few technical problems with the histogram because the compound Poisson has probability mass at zero and probability density for positive numbers. To correctly represent the empirical density of a compound Poisson random variable, a bar chart was used to show the frequency of zeros and a histogram to show the frequency density of positive numbers. However, the Normal approximation and the saddlepoint approximation does not have mass support at zero. Therefore, to compare these approximate densities to the empirical distribution fairly, a histogram containing both zero and positive samples was used when appropriate.

The evaluation of the p.d.f. using the saddlepoint approximate required a bit of caution to avoid over/underflow problems. Suppose realisations of X were simulated $\{x_1, x_2, x_3, \ldots, x_n\}$. The saddlepoint approximation was computed up to a constant using

$$p_X(x) \propto \exp\left[-\frac{\alpha+2}{2(\alpha+1)}\ln(x) - x\beta + \left(\frac{x\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+1}} \lambda^{\frac{1}{\alpha+1}}(\alpha+1) - k\right] \tag{4.40}$$

for 10 000 equally spaced points from and including the minimum non-zero simulated value to the maximum simulated value. k is some constant which was chosen to be

$$k = \max_{i\in\{1,2,3,\ldots,n\}}\left[-\frac{\alpha+2}{2(\alpha+1)}\ln(x_i) - x_i\beta + \left(\frac{x_i\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+1}} \lambda^{\frac{1}{\alpha+1}}(\alpha+1)\right]. \tag{4.41}$$

The density was then normalised by numerically integrating it using the trapezium rule using the 10 000 evaluated points.

A Q-Q plot is a plot which compares the empirical quantiles with the theoretical quantiles. Let

$$F_X(x) = \mathbb{P}(X \leqslant x) \tag{4.42}$$

and

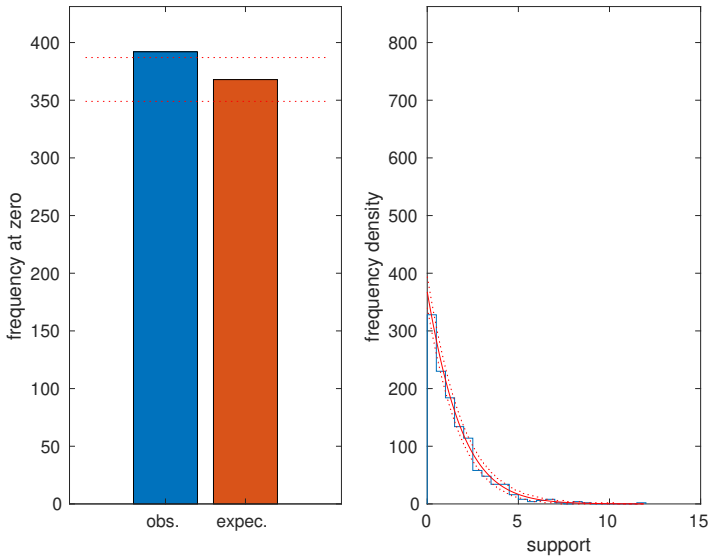
$$\widehat{F}_X(x) = \frac{1}{n} \cdot \max\left[\left(\sum_{i=1}^{n} \mathbb{I}(x_i \leqslant x)\right) - 0.5, 0\right], \tag{4.43}$$

then a Q-Q plot is a parametric plot which plots $\widehat{F}_X^{-1}(p)$ against $F_X^{-1}(p)$ for $p = \frac{0.5}{n}, \frac{1.5}{n}, \frac{2.5}{n}, \ldots, \frac{n-0.5}{n}$. If $F_X(x)$ and $\widehat{F}_X(x)$ are similar, a Q-Q plot should be a straight line with gradient 1, intercepting the origin. For the exact method and the saddlepoint approximation, $F_X(x)$ was found numerically by evaluating the p.d.f. at 10 000 equally spaced points from and including the minimum to the maximum of the simulated samples, and then summing the required trapeziums. For the saddlepoint approximation, it does not support zero, thus the numerical integration started at a non-zero value. $\widehat{F}_X^{-1}(p)$ was then calculated by interpolation.

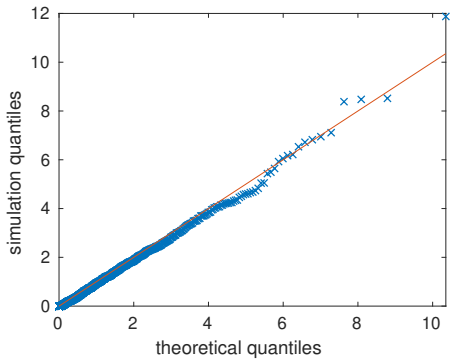
When plotting the probability of obtaining a zero or the p.d.f. for positive values, confidence intervals were plotted as well. Consider a bin in a histogram, the confidence intervals were obtained by assuming that the frequency in a bin $\sim$ Poisson(p.d.f. evaluated at the bin $\times$ 1 000 $\times$ bin width). The 68% confidence interval was chosen because this is a typical choice in physics (Cowan, 1998).

For low λ (Figures 4.1 to 4.4) there was a chance of simulating zeros. The exact method has an advantage here because it can evaluate probability mass at zero. The Normal approximation failed to capture the probability mass at zero because the Normal distribution is symmetric and supports negative values. The saddlepoint approximation improves on the Normal approximation by capturing the skewness and kurtosis of the distribution (Bedrick and Hill, 1992; Butler, 2007). As a result, it captured the probability mass at zero as shown by an increase in probability density towards zero.

For CP$\Gamma(1, 100, 1)$, the compound Poisson-gamma p.d.f. contained multiple peaks. Figure 4.4 shows that the saddlepoint approximation was not flexible enough to capture them. In Figure 4.2, the Q-Q plot for the exact method was quite sensitive at the tails of each peak, perhaps there were a few inaccuracies in the exact evaluation of the p.d.f. or in the calculation of the numerical calculation of the c.d.f. and its inversion.

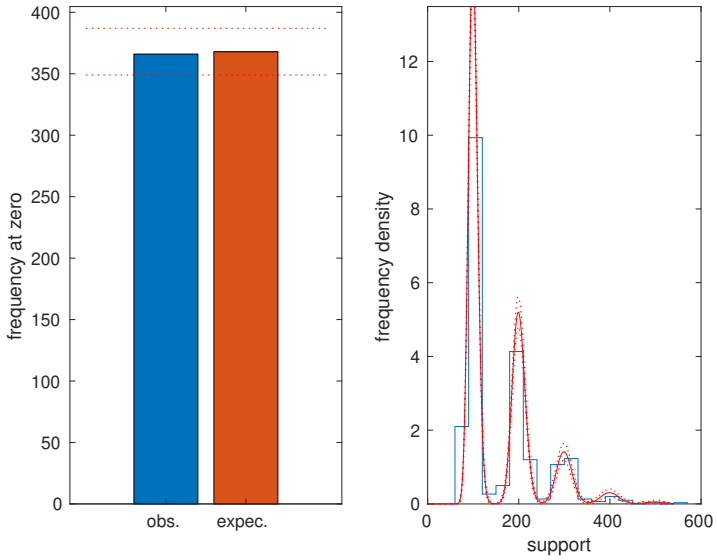


(a) Left: Observed and expected frequency of a zero. Right: Histogram and p.d.f. of non-zero values.

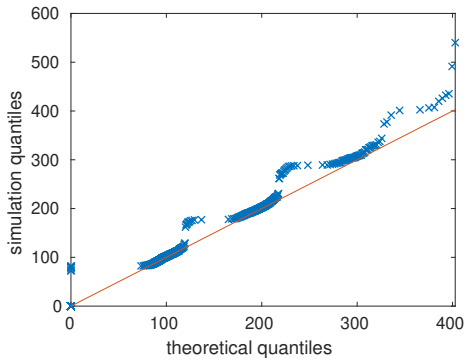


(b) Q-Q plot

Figure 4.1: 1 000 $\mathrm{CP\Gamma}(1, 1, 1)$ random variables were simulated and its empirical density is compared to the p.d.f. evaluated using the exact method. In a), the dotted red line shows the 68% confidence interval of the expected frequency or frequency density.

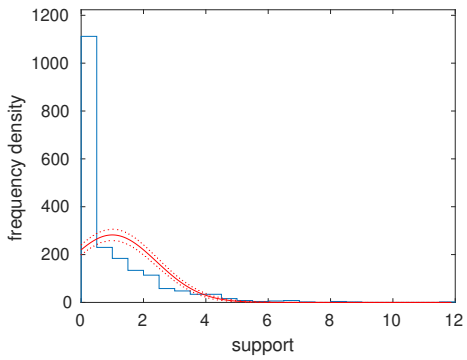


(a) Left: Observed and expected frequency of a zero. Right: Histogram and p.d.f. of non-zero values.

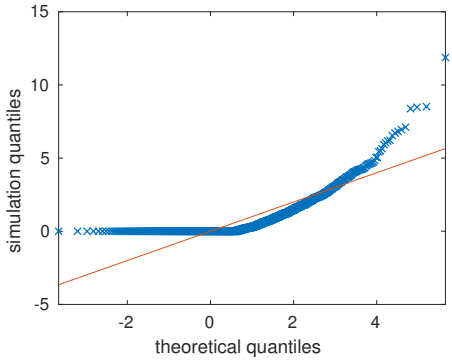


(b) Q-Q plot

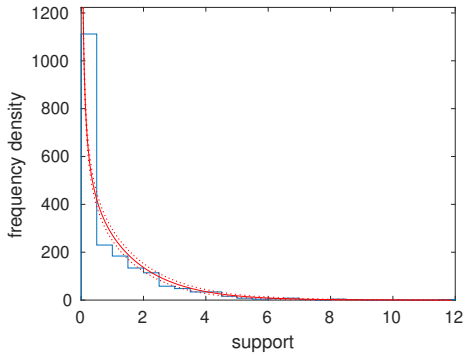
Figure 4.2: 1 000 $\mathrm{CP}\Gamma(1, 100, 1)$ random variables were simulated and its empirical density is compared to the p.d.f. evaluated using the exact method. In a), the dotted red line shows the 68% confidence interval of the expected frequency or frequency density.



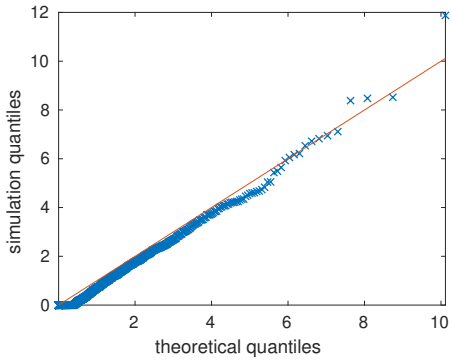
(a) Histogram - Normal approx.



(b) Q-Q plot - Normal approx.



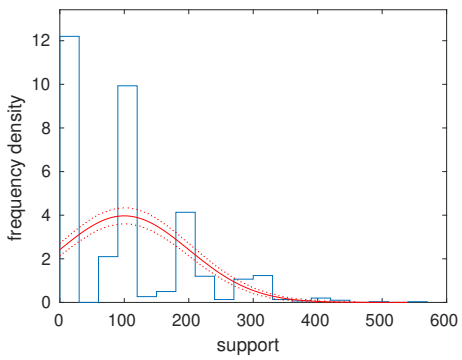
(c) Histogram - Saddlepoint approx.



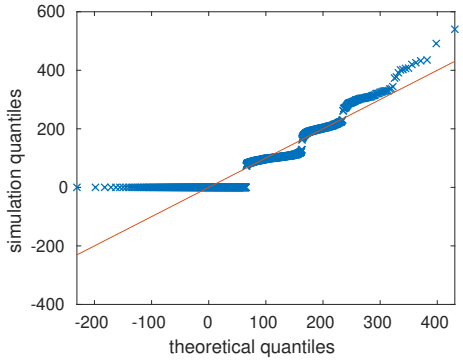
(d) Q-Q plot - Saddlepoint approx.

Figure 4.3: 1 000 CPΓ(1, 1, 1) random variables were simulated and its empirical density is compared to the approximate p.d.f. On the left, the dotted red line shows the 68% confidence interval of the expected frequency density.

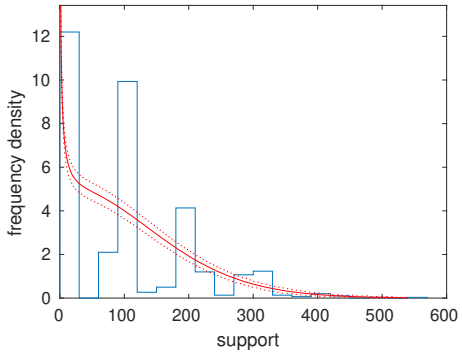
As λ increased (Figures 4.5 and 4.6), all 3 density evaluation methods performed quite well. For CPΓ(10, 1, 1) in Figure 4.5, the Normal approximation did not capture the skewness. In Figure 4.6, λ was high enough where the compound-Poisson distribution started to converge to a Normal distribution. All methods performed well in this realm.



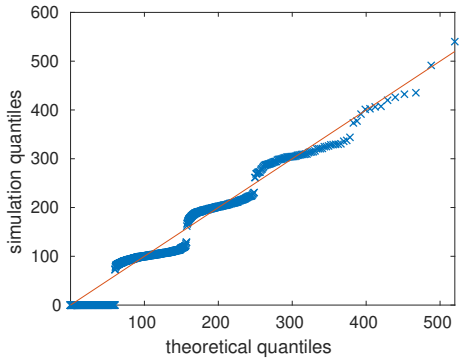
(a) Histogram - Normal approx.



(b) Q-Q plot - Normal approx.



(c) Histogram - Saddlepoint approx.



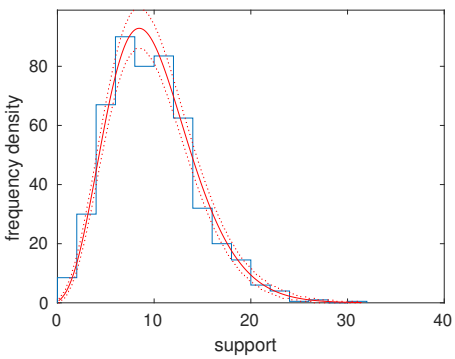
(d) Q-Q plot - Saddlepoint approx.

Figure 4.4: 1 000 CPΓ(1, 100, 1) random variables were simulated and its empirical density is compared to the approximate p.d.f. On the left, the dotted red line shows the 68% confidence interval of the expected frequency density.

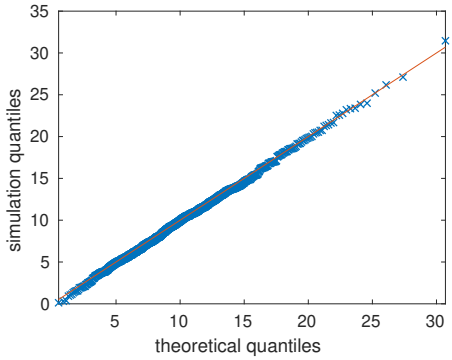
4.3 Proposed Model

The compound Poisson-gamma distribution can be used to model the grey value of each pixel in a projection. Suppose a projection has N pixels and m replicate projections were obtained. Let $X_{i,j}$ and $Y_{i,j}$ be the grey value and photon count, respectively, of the ith pixel in the jth replicate projection.

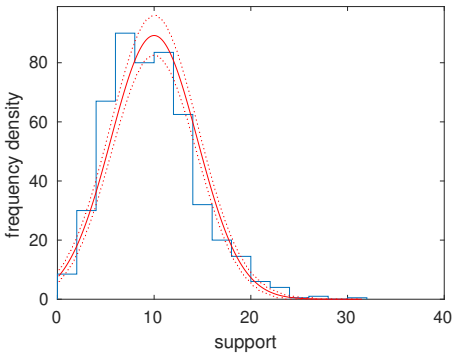
By assuming no beam hardening, the distribution of the photon energy does not change with attenuation. As a result, all pixels will detect photons



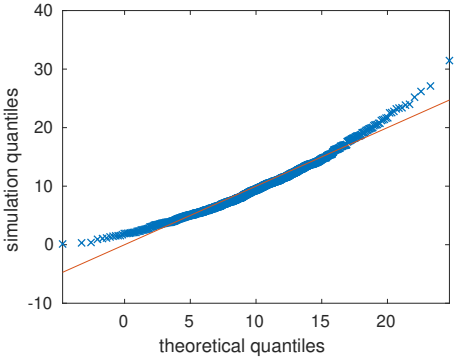
(a) Histogram - Exact method



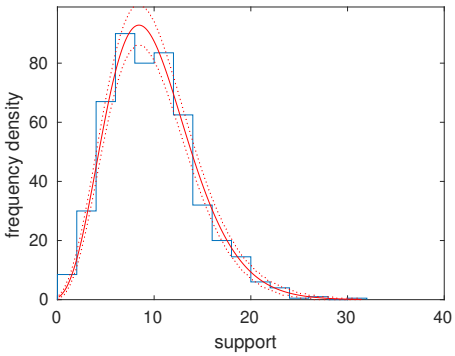
(b) Q-Q plot - Exact method



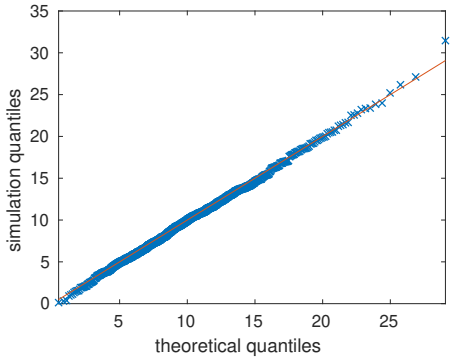
(c) Histogram - Normal approx.



(d) Q-Q plot - Normal approx.

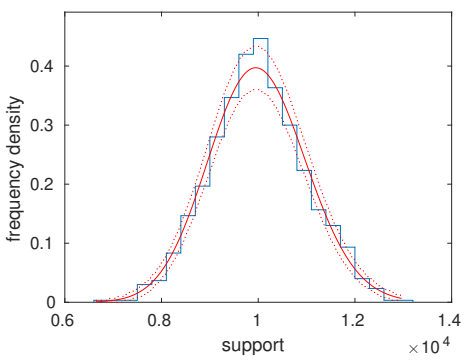


(e) Histogram - Saddlepoint approx.

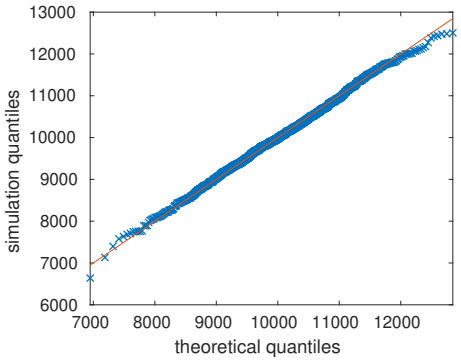


(f) Q-Q plot - Saddlepoint approx.

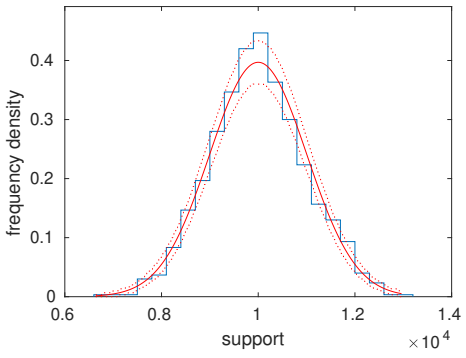
Figure 4.5: 1 000 CPΓ(10, 1, 1) random variables were simulated and its empirical density is compared to the p.d.f. On the left, the dotted red line shows the 68% confidence interval of the expected frequency density.



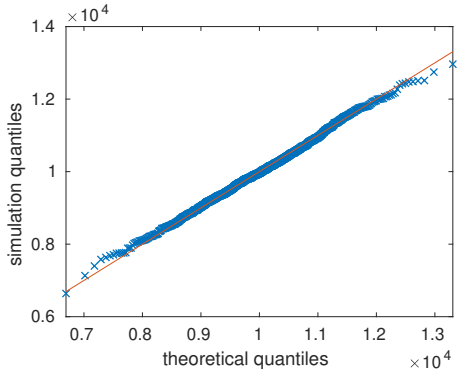
(a) Histogram - Exact method



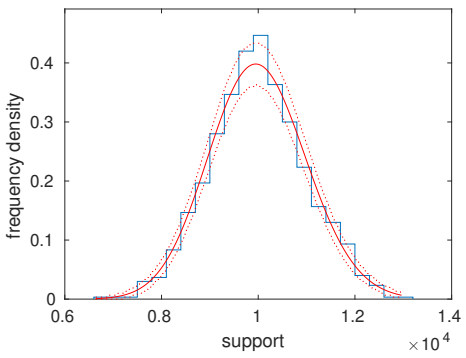
(b) Q-Q plot - Exact method



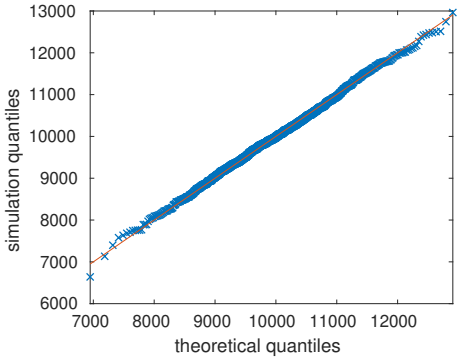
(c) Histogram - Normal approx.



(d) Q-Q plot - Normal approx.



(e) Histogram - Saddlepoint approx.



(f) Q-Q plot - Saddlepoint approx.

Figure 4.6: 1 000 CPΓ(100, 100, 1) random variables were simulated and its empirical density is compared to the p.d.f. On the left, the dotted red line shows the 68% confidence interval of the expected frequency density.

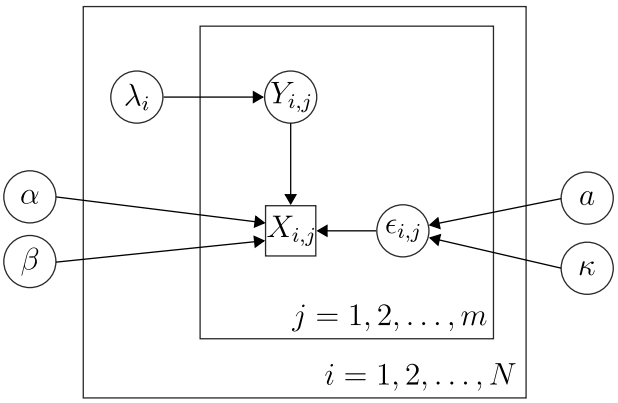


Figure 4.7: Graphical model of the grey value $X_{i,j}$ for each of the N pixels in the m replicate projections. $Y_{i,j} \sim \text{Poisson}(\lambda_i)$ is the photon count. The grey value has a compound Poisson gamma element $X_{i,j}|Y_{i,j} \sim \text{Gamma}(Y_{i,j}\alpha, \beta)$ and can be extended by adding electronic noise $\epsilon_{i,j} \sim \text{N}(a, \kappa)$.

with identical energy distributions, thus, α and β are the same for all pixels. Attenuation does affect the photon count, the more material a photon has to attenuate, the lower the number of detectable photons. The amount of attenuation depends on the specific path from the source to a pixel in a detector, so λ_i varies from pixel to pixel. Electronic Normal noise $\epsilon_{i,j} \sim \text{N}(a, \kappa)$ is added to model noise recorded by the x-ray detector when no x-rays are exposed. This was observed in the black images in Chapter 3.

The graphical model in Figure 4.7 illustrates how all of these variables are linked. a and κ are unknown parameters but can be estimated beforehand from replicate black images. Care must be taken to use the replicate black images for either shading correction or for estimating a and κ to avoid using the data twice. The energy of each photon $U \sim \text{Gamma}(\alpha, \beta)$ was omitted in the graphical model because the conditional distribution $X|Y \sim \text{Gamma}(Y\alpha, \beta)$ encapsulates each detected photon energy already.

The assumption of a gamma-distributed x-ray photon energy and Normal electronic noise was applied to XCT by Xu and Tsui (2009). The assumptions were verified by investigating the variance of radiographs for different x-ray exposures. More statistically sound methods could use goodness of fit tests or Q-Q plots. The use of the gamma distribution will not

be exactly correct because the sources of x-ray photon energy are characteristic and bremsstrahlung radiation which does not behave like a gamma random variable. However, characterisation of the density of the x-ray photon energy is difficult and usually involve more apparatuses and numerical methods (Fewell and Shuping, 1977; Stumbo et al., 2004).

The model may be simplified by omitting the electronic noise and this can be done by setting $a = 0$ and $\kappa = 0$. Further simplification can be done by removing spatial variation by setting $\lambda = \lambda_1 = \lambda_2 = \ldots = \lambda_N$. Parameter estimation can be done using gradient methods because the likelihood of the compound Poisson can be evaluated (Dunn and Smyth, 2005). However, the EM algorithm is faster and the model is well set up for it.

4.4 EM Algorithm

The EM algorithm (Dempster et al., 1977) was proposed to estimate the parameters of a $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$ random variable given a sample of measurements of it $\{x_1, x_2, x_3, \ldots, x_n\}$. The use of the EM algorithm for the compound Poisson distribution in XCT have been studied in Elbakri (2003); Xie (2008); Xu and Tsui (2009).

Let $\widehat{\lambda}$, $\widehat{\alpha}$ and $\widehat{\beta}$ be estimators of λ, α and β respectively. The log-likelihood is defined to be

$$\ln L(\lambda, \alpha, \beta; X) = \sum_{i=1}^{n} \left[\mathbb{I}(x_i = 0) \ln \mathbb{P}(X = x_i) + \mathbb{I}(x_i > 0) \ln p_X(x_i) \right]$$

and using the results in Appendix B.2

$$\ln L(\lambda, \alpha, \beta; X) = \sum_{i=1}^{n} \left[-\mathbb{I}(x_i = 0)\lambda \; + \mathbb{I}(x_i > 0) \left(-\beta x_i - \lambda - \ln x + \ln \sum_{y=1}^{\infty} W_y \right) \right] . \tag{4.44}$$

Maximum likelihood estimators are values of λ, α and β which jointly maximise the log-likelihood.

The EM algorithm (Dempster et al., 1977) treats $Y \sim \text{Poisson}(\lambda)$ as a latent variable. Let $\{Y_1, Y_2, Y_3, \ldots, Y_n\}$ be realisations of Y. Define the joint

log-likelihood to be

$$\ln L(\lambda,\alpha,\beta;X,Y) = \sum_{i=1}^{n}\big[\mathbb{I}(x_i=0)\ln\mathbb{P}(Y=0) \\ +\mathbb{I}(x_i>0)\ln\big[p_{X|Y}(x_i|Y_i)\,\mathbb{P}(Y=Y_i)\big]\big]$$

so that

$$\ln L(\lambda,\alpha,\beta;X,Y) = \sum_{i=1}^{n}\big[-\mathbb{I}(x_i=0)\lambda \\ +\mathbb{I}(x_i>0)\left(Y_i\alpha\ln\beta-\ln\Gamma(Y_i\alpha)+(Y_i\alpha-1)\ln x_i-\beta x_i\right. \\ \left.-\lambda+Y_i\ln\lambda-\ln(Y_i!)\right)\big]\ . \quad (4.45)$$

The estimators were found by optimising the joint log-likelihood. This was done iteratively by estimating the Y's given the X's and parameters (E step), followed by estimating the parameters given the Y's and X's (M step) until some convergence conditions were met.

It will be shown that using some approximations, the E step and M step can be implemented. However, simulations showed that these estimators struggle for high λ.

E Step

In the E step, the realisations of Y are estimated using

$$y_i = \mathbb{E}\left[Y|X=x_i\right] \quad (4.46)$$

given the parameters λ, α and β. The conditional expectation is calculated using

$$y_i = \begin{cases} 0 & \text{for } x=0 \\ \dfrac{\sum_{y=1}^{\infty} y\,\mathbb{P}(Y=y|X=x_i)}{\sum_{y=1}^{\infty}\mathbb{P}(Y=y|X=x_i)} & \text{for } x>0 \end{cases} \quad (4.47)$$

where

$$\mathbb{P}(Y=y|X=x) = \frac{p_{X|Y}(x|y)\,\mathbb{P}(Y=y)}{p_X(x)}\ .$$

Focussing on the $x>0$ case for now

$$\mathbb{P}(Y=y|X=x) = \frac{1}{p_X(x)}\frac{\beta^{y\alpha}}{\Gamma(y\alpha)}x^{y\alpha-1}\mathrm{e}^{-\beta x}\frac{\mathrm{e}^{-\lambda}\lambda^y}{y!}$$

which is simplified to

$$\mathbb{P}(Y = y|X = x) = W_y \frac{\mathrm{e}^{-\lambda-\beta x}}{x p_X(x)} \ . \tag{4.48}$$

As a result, the conditional expectation is

$$y_i = \frac{\sum_{y=1}^{\infty} y W_y}{\sum_{y=1}^{\infty} W_y} \ . \tag{4.49}$$

As discussed before, the sum in the denominator can be evaluated using the method by Dunn and Smyth (2005). A similar method for evaluating the numerator can be obtained by truncating the sum and summing over large terms.

Let

$$W_y^{(r)} = y^r W_y \quad \text{for } r = 1, 2, 3, \ldots \tag{4.50}$$

so that

$$y_i = \frac{\sum_{y=1}^{\infty} W_y^{(1)}}{\sum_{y=1}^{\infty} W_y} \ . \tag{4.51}$$

Similarly,

$$\zeta_i = \mathbb{V}\mathrm{ar}[Y|X = x_i] = \frac{\sum_{y=1}^{\infty} W_y^{(2)}}{\sum_{y=1}^{\infty} W_y} - (y_i)^2 \ . \tag{4.52}$$

The expectation terms, y_i and ζ_i, are evaluated here so that they can be used in the M step. The evaluation can be done by truncating the sum

$$\sum_{y=1}^{\infty} W_y^{(r)} \approx \sum_{y=y_{\mathrm{l}}}^{y_{\mathrm{u}}} W_y^{(r)} \tag{4.53}$$

where $y_{\mathrm{l}} < y_{\max} < y_{\mathrm{u}}$ and $y_{\max}$ is the value of y which maximises $W_y^{(r)}$. This term can be expressed as

$$\ln W_y^{(r)} = r \ln y + \ln W_y \tag{4.54}$$

and taking the derivative with respect to y obtains

$$\frac{\partial}{\partial y} \ln W_y^{(r)} = \frac{r}{y} + \frac{\partial}{\partial y} \ln W_y \ . \tag{4.55}$$

Keep in mind that $y = 1, 2, 3, \ldots$ so for large y, an approximation can be made $r/y \approx 0$ so that

$$\frac{\partial}{\partial y} \ln W_y^{(r)} \approx \frac{\partial}{\partial y} \ln W_y \ . \tag{4.56}$$

Therefore, the maximum of $W_y^{(r)}$ is located at $y_{\max} = \dfrac{x^{2-p}}{\phi(2-p)}$ for all $r = 0, 1, 2, \ldots$. As a result, the same method for evaluating W_y can be used to evaluate $W_y^{(r)}$. For a given $r = 0, 1, 2, \ldots$, the limit of the sum $\sum_{y=y_l}^{y_u} W_y^{(r)}$ were chosen such that $W_{y_l}^{(r)}$ and $W_{y_u}^{(r)}$ are less than $\epsilon W_{y_{\max}}^{(r)}$ where ϵ is some small constant. The limits will be different for different values of r.

M Step

In the M step, the conditional expected joint log-likelihood is maximised with respect to the parameters λ, α and β. The objective function is

$$\begin{aligned} T(\lambda, \alpha, \beta) = & \sum_{i=1}^{n} \mathbb{E}\left[-\mathbb{I}(x_i = 0)\lambda \right. \\ & + \mathbb{I}(x_i > 0) \left(Y_i \alpha \ln \beta - \ln \Gamma(Y_i \alpha) + (Y_i \alpha - 1) \ln x_i - \beta x_i \right. \\ & \left. \left. - \lambda + Y_i \ln \lambda - \ln(Y_i!) \right) | X_i = x_i \right] \\ = & -n\lambda \\ & + \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \left[\mathbb{E}[Y_i | X_i = x_i] \alpha \ln \beta - \mathbb{E}[\ln \Gamma(Y_i \alpha) | X_i = x_i] \right. \\ & \left. + \mathbb{E}[Y_i | X_i = x_i] \alpha \ln x_i - \beta x_i + \mathbb{E}[Y_i | X_i = x_i] \ln \lambda \right] + c \end{aligned} \tag{4.57}$$

where c is some constant not dependent on λ, α or β.

The conditional expectation $y_i = \mathbb{E}[Y_i | X_i = x_i]$ and $\zeta_i = \mathbb{V}\mathrm{ar}[Y_i | X_i = x_i]$ were calculated beforehand in the E step. The quantity $\mathbb{E}[\ln \Gamma(Y_i \alpha) | X_i = x_i]$ can be calculated using the approximation

$$\mathbb{E}[\ln \Gamma(Y_i \alpha) | X_i = x_i] \approx \ln \Gamma(\alpha y_i) + \frac{1}{2} \zeta_i \alpha^2 \psi'(y_i \alpha) \tag{4.58}$$

where $\psi(n)$ is the digamma function. The objective function is then

$$T(\lambda,\alpha,\beta) \approx -n\lambda + \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\Big[y_i\alpha\ln\beta - \ln\Gamma(\alpha y_i) - \frac{1}{2}\zeta_i\alpha^2\psi'(y_i\alpha) + y_i\alpha\ln x_i - \beta x_i + y_i\ln\lambda\Big] + c \ . \quad (4.59)$$

Taking the derivative with respect to λ

$$\frac{\partial T}{\partial\lambda} = -n + \frac{\sum_{i=1}^{n} y_i}{\lambda} \quad (4.60)$$

and setting it to zero

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{n} \quad (4.61)$$

obtains a M step estimator for λ. Taking the second-order derivative with respect to λ

$$\frac{\partial^2 T}{\partial\lambda^2} = -\frac{\sum_{i=1}^{n} y_i}{\lambda^2} < 0 \quad (4.62)$$

verifies that $\widehat{\lambda}$ maximises T. In addition,

$$\frac{\partial^2 T}{\partial\alpha\partial\lambda} = 0 \quad (4.63)$$

and

$$\frac{\partial^2 T}{\partial\beta\partial\lambda} = 0 \ . \quad (4.64)$$

Maximising T with respect to α and β can be done numerically using the Newton-Raphson method since derivatives up to the second-order can be obtained. For the first-order derivatives, these are

$$\frac{\partial T}{\partial\alpha} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\Big[y_i\ln\beta - \psi(\alpha y_i)y_i - \zeta_i\alpha\psi'(\alpha y_i) - \frac{1}{2}\zeta_i\alpha^2\psi''(\alpha y_i)y_i + y_i\ln x_i\Big] \quad (4.65)$$

and

$$\frac{\partial T}{\partial\beta} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[\frac{\alpha y_i}{\beta} - x_i\right] \ . \quad (4.66)$$

The second-order derivatives are

$$\frac{\partial^2 T}{\partial\alpha\partial\beta} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[\frac{y_i}{\beta}\right] \ , \quad (4.67)$$

$$\frac{\partial^2 T}{\partial \beta^2} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \left[-\frac{\alpha y_i}{\beta^2} \right] , \tag{4.68}$$

and

$$\frac{\partial^2 T}{\partial \alpha^2} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \Big[-y_i^2 \psi'(\alpha y_i) - \zeta_i \psi'(\alpha y_i) - \zeta_i \alpha y_i \psi''(\alpha y_i) \\ -\zeta_i \alpha \psi''(\alpha y_i) y_i - \frac{1}{2} \zeta_i \alpha^2 \psi'''(\alpha y_i) y_i^2 \Big]$$

simplifying to

$$\frac{\partial^2 T}{\partial \alpha^2} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \Big[-(y_i^2 + \zeta_i) \psi'(\alpha y_i) - 2\zeta_i \alpha y_i \psi''(\alpha y_i) \\ -\frac{1}{2} \zeta_i \alpha^2 y_i^2 \psi'''(\alpha y_i) \Big] . \tag{4.69}$$

All the derivatives can be used in the Newton-Raphson iterative update to update the estimators $\widehat{\alpha}$ and $\widehat{\beta}$. The update is

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} \leftarrow \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} - \left[\nabla_{\alpha,\beta} \nabla_{\alpha,\beta}^{\mathrm{T}} \, T \big|_{\alpha=\widehat{\alpha},\beta=\widehat{\beta}} \right]^{-1} \left[\nabla_{\alpha,\beta} \, T \big|_{\alpha=\widehat{\alpha},\beta=\widehat{\beta}} \right] \tag{4.70}$$

where

$$\nabla_{\alpha,\beta} = \begin{pmatrix} \partial/\partial\alpha \\ \partial/\partial\beta \end{pmatrix} . \tag{4.71}$$

Since increasing T is sufficient for the EM algorithm (Dempster et al., 1977), one step of the Newton-Raphson iterative update was chosen in the M step to avoid implementing a convergence condition for the Newton-Raphson method.

Simulations

An experiment was conducted to assess the performance on the EM algorithm. For a given set of parameters, 1 000 samples of a $\mathrm{CP\Gamma}(\lambda, \alpha, \beta)$ random variable were simulated. The EM algorithm was initialised with its parameters at the true values to investigate the convergence in that vicinity. The log-likelihood $\ln L(\lambda, \alpha, \beta; X)$ and the parameters were recorded at every EM step. The experiment was repeated 10 times using different simulated samples.

The results for $\text{CP}\Gamma(1,1,1)$, $\text{CP}\Gamma(1,100,1)$, $\text{CP}\Gamma(10,1,1)$ and $\text{CP}\Gamma(100,100,1)$ are shown in Figures 4.8, 4.9, 4.10 and 4.11 respectively. Good performance was observed in the $\lambda = 1$ case with convergence of all three parameters within a step or two. The Cramér-Rao lower bound captured the spread of the estimates well.

For $\lambda = 10$ and $\lambda = 100$, the estimates of α and β struggled to converge and increased/decreased without bounds without affecting the log-likelihood. It appeared that the EM algorithm failed for $\lambda > 10$ looking at these particular examples.

4.5 Failure Evaluation

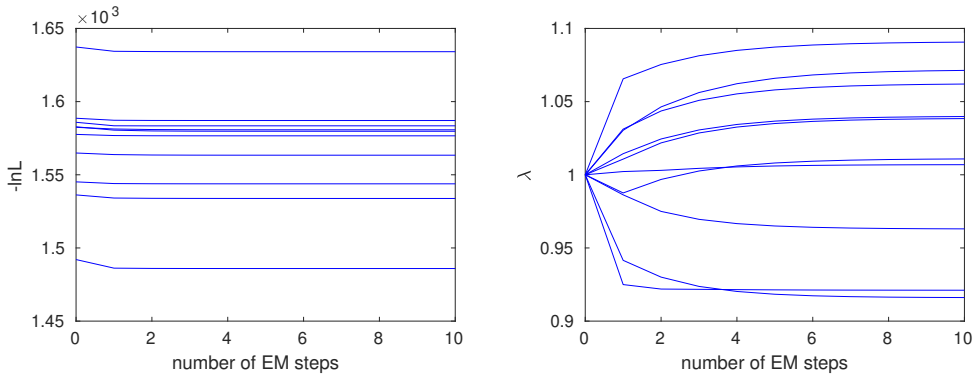
It should be convincing that the EM algorithm failed when the compound Poisson-gamma random variable starts behaving Normally. In particular, in Figure 4.11, estimates of λ were stable while estimates of α and β struggled to converge. This could be because as the compound Poisson-gamma distribution approaches the Normal distribution, the parameters (λ, α, β) becomes degenerate because there is more than one way to represent a two-parameter random variable $\text{N}(\mu, \sigma^2)$ using three parameters (λ, α, β).

Determinant of the Hessian

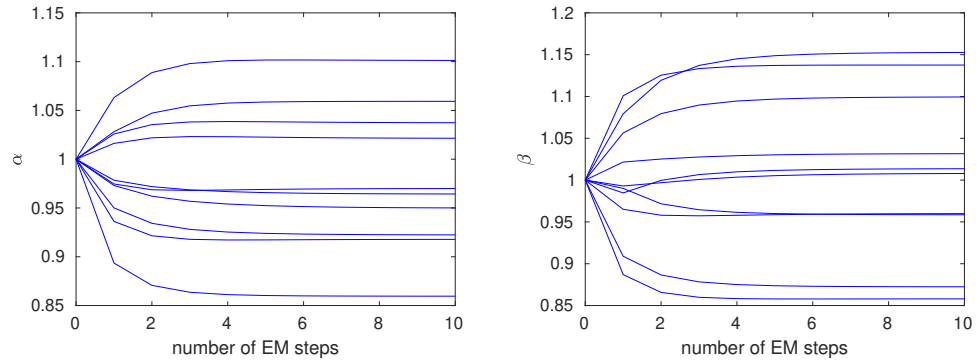
One investigation is to look at the Newton-Raphson step in the M step, in particular, the Hessian matrix $\nabla_{\alpha,\beta}\nabla_{\alpha,\beta}^{\text{T}}T$ for high λ. The elements of the Hessian matrix can be found in Equations (4.67), (4.68) and (4.69). Recall that

$$\frac{\partial^2 T}{\partial \alpha^2} = \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[-(y_i^2 + \zeta_i)\psi'(\alpha y_i) - 2\zeta_i \alpha y_i \psi''(\alpha y_i) - \frac{1}{2}\zeta_i \alpha^2 y_i^2 \psi'''(\alpha y_i)\right].$$

For high α and λ, and hence high y_i's, an approximation can be used for the polygamma functions $\psi^{(k)}(n)$. Using Stirling's approximation $\ln\Gamma(n) \approx$

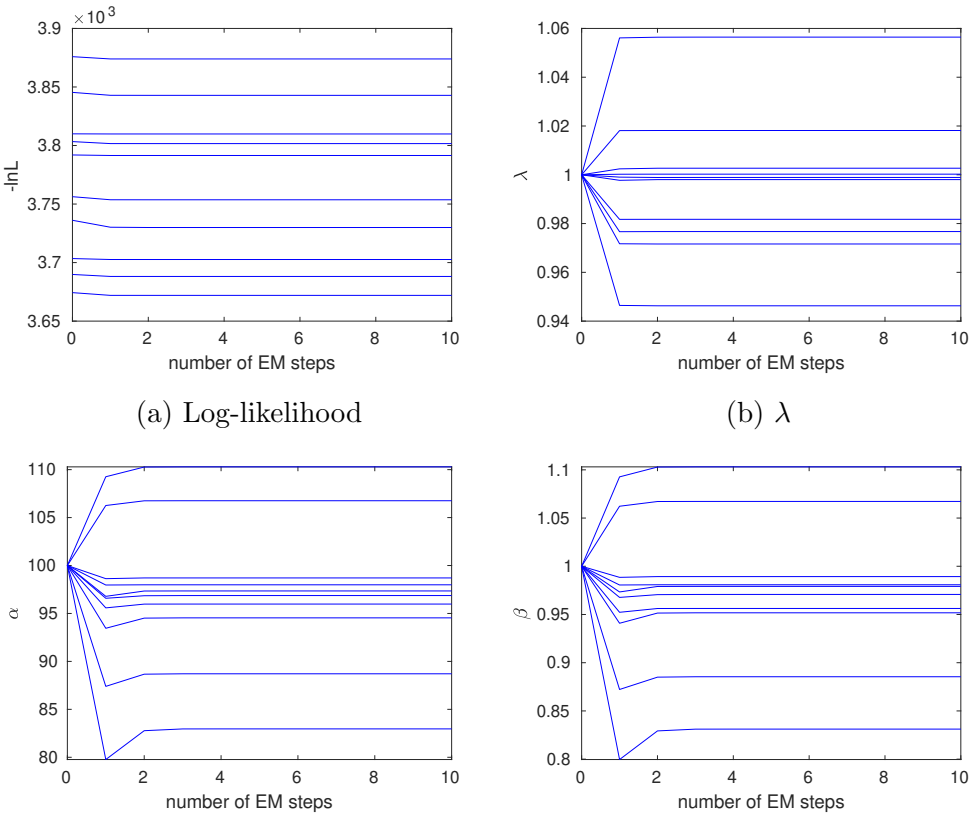


(a) Log-likelihood (b) λ



(c) α (d) β

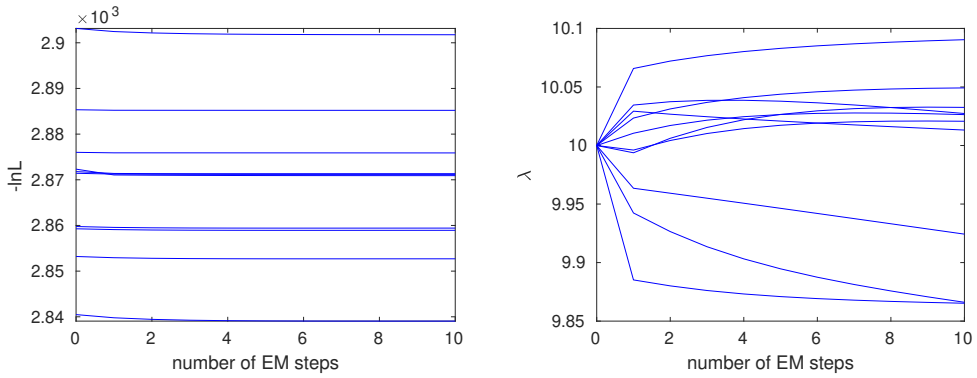
Figure 4.8: The EM algorithm was used to estimate the parameters of a $\mathrm{CP\Gamma}(1,1,1)$ random variable using 1 000 simulated samples. The graphs show the log-likelihood and the estimated parameters at each EM step for 10 different simulations.



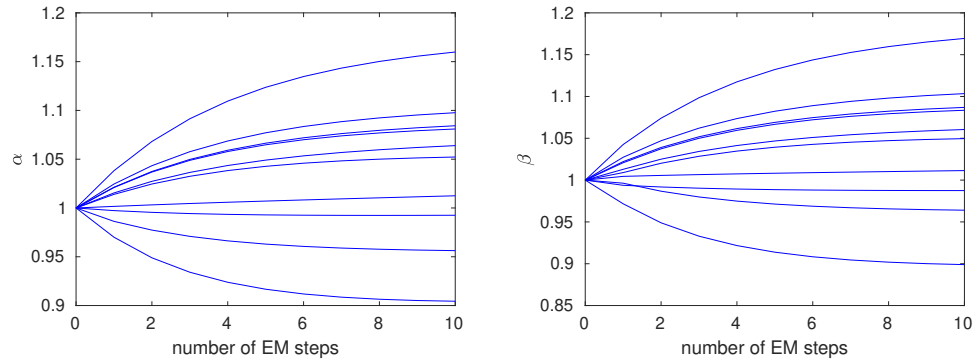
(a) Log-likelihood (b) λ

(c) α (d) β

Figure 4.9: The EM algorithm was used to estimate the parameters of a $\mathrm{CP\Gamma}(1, 100, 1)$ random variable using 1 000 simulated samples. The graphs show the log-likelihood and the estimated parameters at each EM step for 10 different simulations.

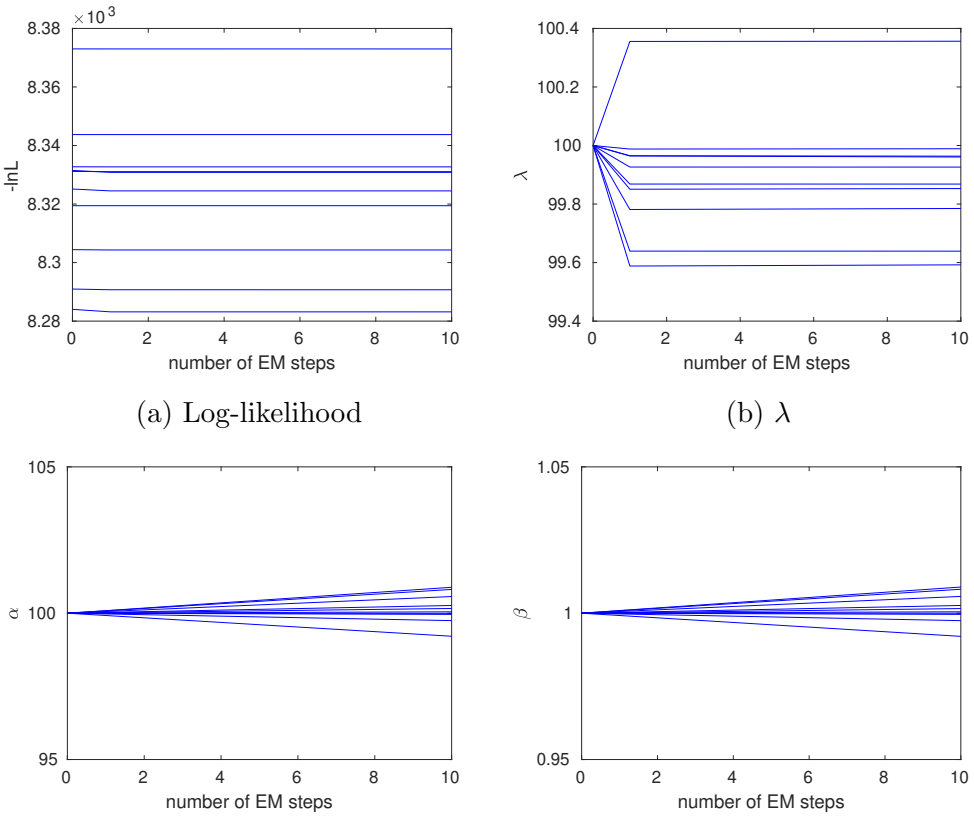


(a) Log-likelihood (b) λ



(c) α (d) β

Figure 4.10: The EM algorithm was used to estimate the parameters of a $\mathrm{CP\Gamma}(10, 1, 1)$ random variable using 1 000 simulated samples. The graphs show the log-likelihood and the estimated parameters at each EM step for 10 different simulations.



(a) Log-likelihood (b) λ

(c) α (d) β

Figure 4.11: The EM algorithm was used to estimate the parameters of a $\mathrm{CP\Gamma}(100, 100, 1)$ random variable using 1 000 simulated samples. The graphs show the log-likelihood and the estimated parameters at each EM step for 10 different simulations.

$\ln(n!) \approx n \ln n - n$, then

$$\psi(n) = \frac{\partial \ln \Gamma(n)}{\partial n} \approx \ln n \ . \tag{4.72}$$

Differentiating further

$$\psi'(n) \approx 1/n \tag{4.73}$$

$$\psi''(n) \approx -1/n^2 \tag{4.74}$$

$$\psi'''(n) \approx 2/n^3 \ . \tag{4.75}$$

$\dfrac{\partial^2 T}{\partial \alpha^2}$ can be approximated

$$\frac{\partial^2 T}{\partial \alpha^2} \approx \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \left[-\frac{y_i^2 + \zeta_i}{\alpha y_i} + 2\frac{\zeta_i \alpha y_i}{\alpha^2 y_i^2} - \frac{1}{2}\frac{2\zeta_i \alpha^2 y_i^2}{\alpha^3 y_i^3} \right]$$

to get

$$\frac{\partial^2 T}{\partial \alpha^2} \approx \sum_{i=1}^{n} \mathbb{I}(x_i > 0) \left[-\frac{y_i}{\alpha} \right] \ . \tag{4.76}$$

The Hessian matrix is, omitting the $\mathbb{I}(x_i > 0)$ term,

$$\nabla_{\alpha,\beta} \nabla_{\alpha,\beta}^{\mathrm{T}} T \approx \sum_{i=1}^{n} \begin{pmatrix} -y_i/\alpha & y_i/\beta \\ y_i/\beta & -\alpha y_i/\beta^2 \end{pmatrix} \ . \tag{4.77}$$

The determinant of the Hessian matrix is

$$\begin{aligned} \| \nabla_{\alpha,\beta} \nabla_{\alpha,\beta}^{\mathrm{T}} T \| &\approx \left(\sum_{i=1}^{n} -\frac{\alpha y_i}{\beta^2} \right) \left(\sum_{i=1}^{n} -\frac{y_i}{\alpha} \right) - \left(\sum_{i=1}^{n} \frac{y_i}{\beta} \right)^2 \\ &\approx \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{y_i y_j}{\beta^2} \right) - \left(\sum_{i=1}^{n} \frac{y_i}{\beta} \right) \left(\sum_{j=1}^{n} \frac{y_j}{\beta} \right) \\ &\approx \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{y_i y_j}{\beta^2} \right) - \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{y_i y_j}{\beta^2} \right) \end{aligned}$$

to obtain

$$\| \nabla_{\alpha,\beta} \nabla_{\alpha,\beta}^{\mathrm{T}} T \| \approx 0 \ . \tag{4.78}$$

This results in a few things. Firstly, $\nabla_{\alpha,\beta} \nabla_{\alpha,\beta}^{\mathrm{T}} T$ is singular thus its inverse cannot be evaluated which is needed for the Newton-Raphson method. Secondly, the sufficient conditions to classify a stationary point as a maximum are that the diagonal elements of the Hessian matrix are negative and the determinate is positive. For high λ and α, the second condition is not met.

Constrained Objective

The parameter β can be constrained for a given mean to be

$$\beta = \frac{\lambda\alpha}{\widehat{\mu}} \tag{4.79}$$

where

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.80}$$

and it was investigated whenever a unique solution to $\dfrac{\partial T}{\partial \alpha} = 0$ can be found for high λ and α. The constrained objective function is

$$T(\lambda,\alpha) = -n\lambda + \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[y_i\alpha \ln\left(\frac{\lambda\alpha}{\widehat{\mu}}\right) - \ln\Gamma(\alpha y_i) - \frac{1}{2}\zeta_i\alpha^2\psi'(y_i\alpha) + y_i\alpha\ln x_i - \frac{\lambda\alpha x_i}{\widehat{\mu}} + y_i\ln\lambda\right] + c \tag{4.81}$$

which can be approximated to

$$\begin{aligned} T(\lambda,\alpha) &\approx -n\lambda + \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[y_i\alpha\ln\left(\frac{\lambda\alpha}{\widehat{\mu}}\right) - \alpha y_i\ln(\alpha y_i) + \alpha y_i - \frac{1}{2}\frac{\zeta_i\alpha}{y_i} + y_i\alpha\ln x_i - \frac{\lambda\alpha x_i}{\widehat{\mu}} + y_i\ln\lambda\right] + c \\ &\approx -n\lambda + \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[y_i\alpha\ln\left(\frac{\lambda}{y_i\widehat{\mu}}\right) + \alpha y_i - \frac{1}{2}\frac{\zeta_i\alpha}{y_i} + y_i\alpha\ln x_i - \frac{\lambda\alpha x_i}{\widehat{\mu}} + y_i\ln\lambda\right] + c\ . \end{aligned} \tag{4.82}$$

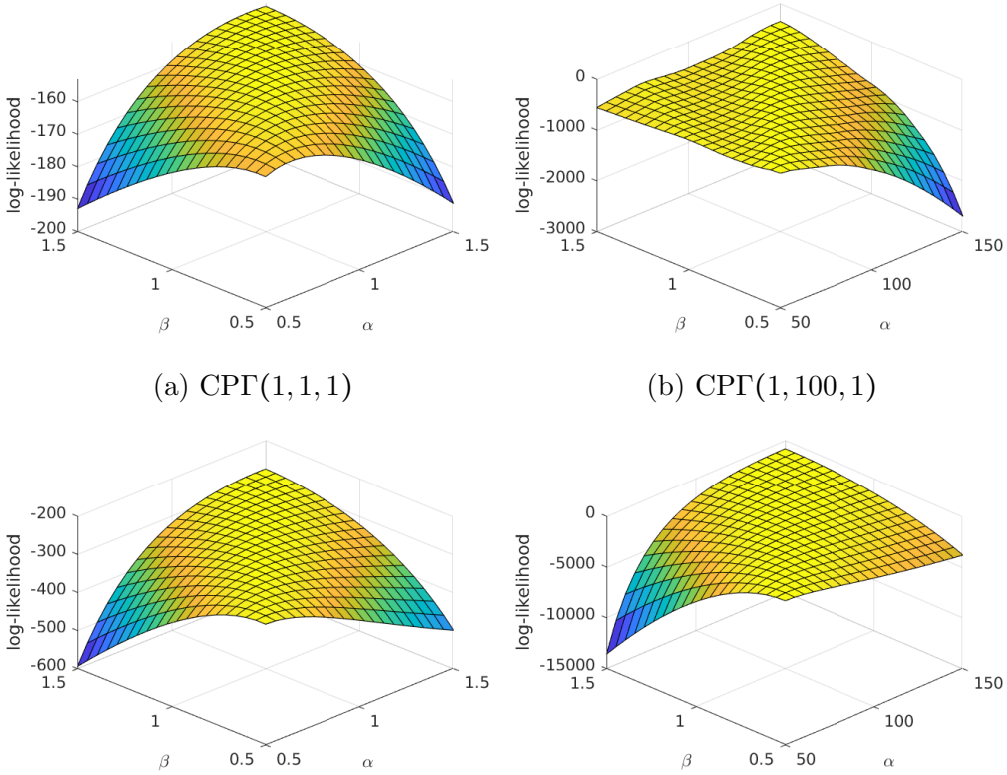
Taking the derivative with respect to α

$$\frac{\partial T}{\partial\alpha} \approx \sum_{i=1}^{n} \mathbb{I}(x_i > 0)\left[y_i\ln\left(\frac{\lambda}{y_i\widehat{\mu}}\right) + y_i - \frac{1}{2}\frac{\zeta_i}{y_i} + y_i\ln x_i - \frac{\lambda x_i}{\widehat{\mu}}\right] \tag{4.83}$$

and this is not α dependent, thus there is no solution for α.

Log-Likelihood Plot

The log-likelihood, using a simulation of 100 compound Poisson-gamma random variables, are plotted for fixed λ in Figure 4.12. These plots are concerning because the likelihood appeared not convex and there is no unique maximum for all λ investigated.



(a) CPΓ(1, 1, 1)

(b) CPΓ(1, 100, 1)

(c) CPΓ(10, 1, 1)

(d) CPΓ(100, 100, 1)

Figure 4.12: Log-likelihood from a simulation of 100 compound Poisson-gamma random variables. λ is fixed at the true value.

It was observed that there is a saddle point in the log-likelihood for the CPΓ(1, 100, 1) case. This is when the log-likelihood increased and decreased when varying α and β. This suggests the possibility of local maxima.

As a result, there is no unique maximum likelihood estimator and the EM algorithm would be very sensitive to the starting position, including for small λ.

4.6 Conclusion

The compound Poisson-gamma distribution was studied but it was found there were identifiability issues when fitting the model onto simulated data,

in particular for high λ. It was found that the log-likelihood was quite flat because, by using numerous approximations, it was found that the determinant of the Hessian is about zero. This cast doubts if a unique maximum likelihood estimator exists. Tighter bounds for the Hessian is beyond the scope of this thesis.

The problem with the EM algorithm is that it is sensitive to the initial value and does not guarantee convergence to a global maximum (Moon, 1996). Another disadvantage is that the EM algorithm does not produce uncertainty quantification for the estimators. Methods such as bootstrapping (Efron, 1979) and Louis' missing information principle (Louis, 1982) can be used to obtain such uncertainty quantification.

There are methods to deal with flat likelihoods by extending the likelihood to include penalising terms (Cole et al., 2014). Another approach is to use Bayesian methods (Bishop, 2006) where a prior distribution is defined for the parameters α, β, $\lambda_1, \lambda_2, \ldots, \lambda_N$, a and κ where N is the number of pixels and a and κ are the mean and variance terms for the Normal noise, defined in Section 4.3. The prior distribution can reflect prior knowledge of the model, for example, the statistical properties of the sample of black images can be embedded into the priors of a and κ. Back of the envelope calculations can be used to show that $\lambda_i \sim 10^7\,\text{photon}\,\text{px}^{-1}$ which can be used to aid in the creation of a prior distribution for λ_i. Poisson graphical models (Yang et al., 2013) and Gaussian processes (Williams and Rasmussen, 1996) can impose a spatial dependency on $Y_{i,j}$ and λ_i respectively as well.

Bayesian inference is done by studying the posterior distribution where

$$\text{posterior} \propto \text{likelhood} \times \text{prior} \ .$$

By multiplying the likelihood with the prior, the posterior should hopefully be convex and enable better convergence. Studying the posterior distribution typically requires Markov chains using Monte Carlo methods (Brooks et al., 2011) such as Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970), Gibbs sampling (Geman and Geman, 1984), Hamiltonian Monte Carlo (Neal, 2011; Hoffman and Gelman, 2014) and slice sampling (Neal, 2003; Murray et al., 2010).

To keep the frequentist theme throughout this thesis, the next chapter studies the linear relationship between the variance and mean of the grey value in a pixel. The uncertainty can be predicted by using such a relationship.

Chapter 5

Variance Prediction

In the previous chapter, the grey values were modelled using a compound Poisson distribution to reflect the science of x-rays. This model has a linear relationship between the variance and the expectation. As a result, it would be possible to predict the variance of the grey value, given the grey value. This opens up new ways to predict the uncertainty of each pixel in an x-ray projection.

In this chapter, generalised linear models (Nelder and Wedderburn, 1972; Nelder and Baker, 1972; McCullagh, 1984) with different link functions and polynomial features were selected using forward stepwise selection. These selected models were compared using cross-validation to find the best model and verified using residual analysis.

5.1 Generalised Linear Models

The sample variance-mean data were obtained from the replicate projections. Let $x_{i,j}$ be the grey value of the ith pixel, ignoring any spatial information, from the jth replicate projection for $i = 1, 2, 3, \ldots, N$ and $j = 1, 2, 3, \ldots, m$. The sample mean and sample variance grey value for the ith pixel are

$$x_i = \frac{1}{m} \sum_{j=1}^{m} x_{i,j} \tag{5.1}$$

and

$$y_i = \frac{1}{m-1}\sum_{j=1}^{m}\left(x_{i,j} - x_i\right)^2 \tag{5.2}$$

respectively. The symbol y should not be confused with the latent variable in the previous chapter. Only pixels in the region of interest (ROI) were considered here, that is, pixels which represent the test sample. The ROI was created by manually segmenting the test sample from the projection.

The aim is to model and predict the grey value variance of a pixel given its grey value by using a model fitted onto the sample variance-mean data. The variance-mean data is denoted as $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ where N is the area of the ROI or the size of the dataset. Let $Y(x)$ be a random variable and the sample variance given a grey value mean x. It was assumed that the standard error from estimating the mean was negligible so that the uncertainty is captured by the random variable Y. It was assumed that for a given pixel, the grey values are Normal and i.i.d. Let $\sigma^2(x)$ be the variance given a grey value x, then it can be shown that

$$\frac{(m-1)Y(x)}{\sigma^2(x)} \sim \chi^2_{m-1} \tag{5.3}$$

which results in

$$Y(x) \sim \text{Gamma}\left(\alpha, \frac{\alpha}{\sigma^2(x)}\right) \tag{5.4}$$

where $\alpha = (m-1)/2$ is the shape parameter. The expectation and variance are

$$\mathbb{E}\left[Y(x)\right] = \sigma^2(x) \tag{5.5}$$

and

$$\mathbb{V}\text{ar}\left[Y(x)\right] = \left(\frac{\sigma^2(x)}{\alpha}\right)^2 \tag{5.6}$$

respectively.

This framework allows the use of generalised linear models (GLM) (Nelder and Wedderburn, 1972; Nelder and Baker, 1972; McCullagh, 1984). In the gamma distribution case, a GLM can be used to model

$$Y(x) \sim \text{Gamma}\left(\alpha, \frac{\alpha}{g^{-1}(\eta(x))}\right) \tag{5.7}$$

where $g(y)$ is the link function and $\eta(x)$ is a linear function called the systematic component. It should be noted that

$$\mathbb{E}[Y(x)] = g^{-1}(\eta(x)) \tag{5.8}$$

which shows how the link function and systematic component work together. Examples of link functions are the identity link

$$g(y) = g^{-1}(y) = y \tag{5.9}$$

and, for the gamma distribution case, the canonical link

$$g(y) = g^{-1}(y) = 1/y \ . \tag{5.10}$$

An example of a systematic component are polynomial features $\eta(x) = \sum_{r=0}^{p} \beta_r x^r$ so that when used with the identity link, for example, $\mathbb{E}[Y(x)] = \sum_{r=0}^{p} \beta_r x^r$. Iterative reweighted least squares (Friedman et al., 2001) can be used to estimate the parameters $\beta_0, \beta_1, \ldots, \beta_p$ given data for the model to fit onto.

Once the parameters have been estimated, prediction of the variance given a grey value x is done by using $\widehat{y}(x) = g^{-1}(\widehat{\eta}(x))$ where $\widehat{\eta}(x)$ is the systematic component using the estimated parameters.

5.2 Model Selection

This section describes how forward stepwise selection (Efroymson, 1960; Friedman et al., 2001) was used to select which polynomial features to use in the systematic component.

In summary, forward stepwise selection fits a basic model to the data initially. A feature is added to make the model more complicated at each step to improve the fit onto the data. This is continued until the model cannot be improved subject to overfitting. The Akaike information criterion (AIC) (Akaike, 1974; Friedman et al., 2001) and the Bayesian information criterion (BIC) (Schwarz, 1978; Friedman et al., 2001) are criteria which can be used to assess the fit of the model at each step without overfitting to the data too much.

The AIC and BIC are given as

$$\text{AIC} = 2k - 2\ln L \tag{5.11}$$

and

$$\text{BIC} = k\ln N - 2\ln L \tag{5.12}$$

respectively where k is the number of parameters in the systematic component and $\ln L$ is the log-likelihood of the GLM. The model with the lowest AIC or BIC is preferred. GLM aims to maximise the log-likelihood but the additional terms in the criteria penalise models with too many terms. The log-likelihood is given as

$$\ln L = \sum_{i=1}^{N}\left[\alpha\ln\alpha - \ln\Gamma(\alpha) - \alpha\ln\widehat{y}_i + (\alpha-1)\ln y_i - \frac{\alpha y_i}{\widehat{y}_i}\right]. \tag{5.13}$$

where $\widehat{y}_i = \widehat{y}(x_i)$. $\alpha = (m-1)/2$ was assumed to be known so does not need to be estimated.

The procedure is as follows. A criterion and a link function were chosen beforehand. In the initial step, a GLM with systematic component $\eta(x) = \beta_0$ was fitted and the criterion was recorded. In the next step, a polynomial feature with one order higher was added to the systematic component $\eta(x) = \beta_0 + \beta_1 x$, fitted and the criterion recorded. Also, a polynomial feature with one order lower was added $\beta(x) = \beta_{-1}x^{-1} + \beta_0$ and fitted separately with the criterion recorded. The model which decreased the criterion the most was accepted. Adding higher and lower order polynomials to the systematic component was repeated, for example, after accepting $\eta(x) = \beta_0 + \beta_1 x$, the following systematic components $\eta(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ and $\eta(x) = \beta_{-1}x^{-1} + \beta_0 + \beta_1 x$ were fitted and assessed. The fitted parameters may change when adding more terms. This is repeated until the criterion cannot be decreased and the procedure is left with the final model.

Forward stepwise selection was conducted on the datasets `AbsNoFilter` and `AbsFilter`. The procedure was repeated 10 times by using a random permutation, with replacement, of the replicate projections to obtain a different sample variance-mean data which introduced some variation to the data (Efron, 1979). The procedure was also repeated using various

shading corrections to investigate the effects of shading correction on the variance-mean relationship.

The results are shown in Tables 5.1 and 5.2 for `AbsNoFilter` and Tables 5.3 and 5.4 for `AbsFilter`. The models selected are quite simple and all have two features. Different shading corrections or different criteria had no effect on the selected model. The method was quite robust to the variation introduced to the dataset when repeating the experiment because all 10 repeats consistently selected the same model.

There was some variation to the selected models between datasets. For example, when using the identity link, `AbsNoFilter` preferred $\eta(x) = \beta_0 + \beta_1 x$ whereas `AbsFilter` preferred $\eta(x) = \beta_{-1}x^{-1} + \beta_0$. Using the canonical link, both datasets selected $\eta(x) = \beta_{-1}x^{-1} + \beta_0$ which correspond to $\widehat{y}(x) = (\beta_{-1}x^{-1} + \beta_0)^{-1}$.

Figures 5.1 and 5.2 shows the GLM fits for the datasets `AbsNoFilter` and `AbsFilter` respectively. The prediction intervals were obtained using the distribution $\widehat{Y}(x) \sim \text{Gamma}\left(\alpha, \frac{\alpha}{g^{-1}(\widehat{\eta}(x))}\right)$. The fits all looked reasonable except for `AbsNoFilter` at 120° where the fit did not capture the inflection. The forward stepwise selection may not picked this up because there were a lot of low grey values, causing a leverage towards these low grey values. As a result, the inflection did not stand out to the method.

5.3 Cross-Validation

Instead of assessing the model fit using the AIC and BIC, the performance on predicting the variance given a grey value was assessed using cross-validation (Allen, 1974; Stone, 1974, 1977; Friedman et al., 2001). Cross-validation assesses the model to make predictions on data it has not seen before. This was done by randomly splitting the N data points into two disjoint sets, the training set and the test set τ. The model was fitted onto the training set. Afterwards, the fitted model predicts the variances given the grey values in the test set which were then compared to the actual variances.

Identity Link		null	bw	linear
AIC	order	0 to 1	0 to 1	0 to 1
	votes	10	10	10
	criterion	$(1.7 \pm 0.3) \times 10^6$	$(2.0 \pm 0.3) \times 10^6$	$(1.7 \pm 0.3) \times 10^6$
BIC	order	0 to 1	0 to 1	0 to 1
	votes	10	10	10
	criterion	$(1.5 \pm 0.2) \times 10^6$	$(1.9 \pm 0.3) \times 10^6$	$(1.7 \pm 0.2) \times 10^6$

Canonical Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(2.0 \pm 0.4) \times 10^6$	$(2.1 \pm 0.2) \times 10^6$	$(1.8 \pm 0.2) \times 10^6$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(1.8 \pm 0.3) \times 10^6$	$(1.9 \pm 0.3) \times 10^6$	$(2.0 \pm 0.3) \times 10^6$

Table 5.1: Forward stepwise selection was used to find suitable polynomial features when fitting a gamma GLM onto the sample variance-mean data from the projections in `AbsNoFilter` at 30°. The columns of the table represent different shading corrections. Forward stepwise selection was repeated 10 times by bootstrapping the replicate projections. 'Order' shows the most commonly selected polynomial orders and 'votes' shows the number of times it was selected out of the 10 repeats. The error bars are the standard deviation from the 10 repeats.

Identity Link		null	bw	linear
AIC	order	0 to 1	0 to 1	0 to 1
	votes	10	10	10
	criterion	$(-3 \pm 2) \times 10^5$	$(-2 \pm 2) \times 10^5$	$(-2 \pm 3) \times 10^5$
BIC	order	0 to 1	0 to 1	0 to 1
	votes	10	10	10
	criterion	$(-2 \pm 2) \times 10^5$	$(-1 \pm 3) \times 10^5$	$(-2 \pm 2) \times 10^5$

Canonical Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(-5 \pm 4) \times 10^5$	$(-6 \pm 2) \times 10^5$	$(-4 \pm 3) \times 10^5$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(-6 \pm 2) \times 10^5$	$(-7 \pm 2) \times 10^5$	$(-6 \pm 2) \times 10^5$

Table 5.2: Forward stepwise selection was used to find suitable polynomial features when fitting a gamma GLM onto the sample variance-mean data from the projections in `AbsNoFilter` at 120°. The columns of the table represent different shading corrections. Forward stepwise selection was repeated 10 times by bootstrapping the replicate projections. 'Order' shows the most commonly selected polynomial orders and 'votes' shows the number of times it was selected out of the 10 repeats. The error bars are the standard deviation from the 10 repeats.

Identity Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.6 \pm 0.4) \times 10^6$	$(6.2 \pm 0.3) \times 10^6$	$(6.4 \pm 0.3) \times 10^6$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.5 \pm 0.2) \times 10^6$	$(6.1 \pm 0.3) \times 10^6$	$(6.1 \pm 0.3) \times 10^6$

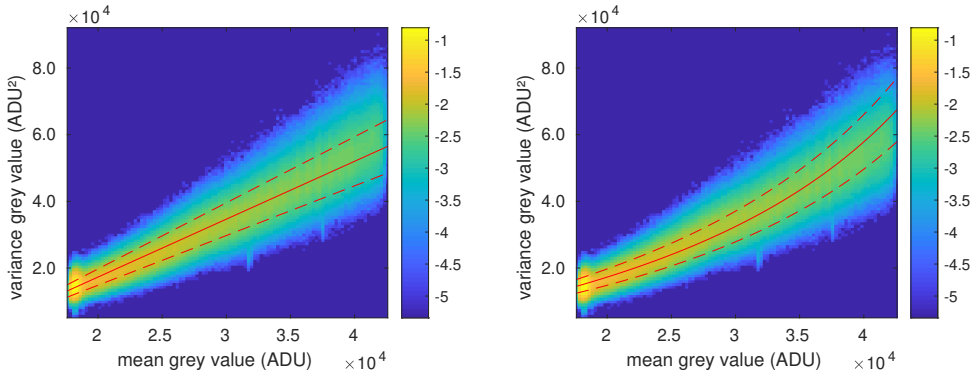
Canonical Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.5 \pm 0.2) \times 10^6$	$(6.4 \pm 0.3) \times 10^6$	$(6.3 \pm 0.2) \times 10^6$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.7 \pm 0.3) \times 10^6$	$(6.4 \pm 0.4) \times 10^6$	$(6.2 \pm 0.1) \times 10^6$

Table 5.3: Forward stepwise selection was used to find suitable polynomial features when fitting a gamma GLM onto the sample variance-mean data from the projections in `AbsFilter` at 30°. The columns of the table represent different shading corrections. Forward stepwise selection was repeated 10 times by bootstrapping the replicate projections. 'Order' shows the most commonly selected polynomial orders and 'votes' shows the number of times it was selected out of the 10 repeats. The error bars are the standard deviation from the 10 repeats.

Identity Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.7 \pm 0.3) \times 10^6$	$(6.4 \pm 0.2) \times 10^6$	$(6.3 \pm 0.2) \times 10^6$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.7 \pm 0.3) \times 10^6$	$(6.4 \pm 0.4) \times 10^6$	$(6.5 \pm 0.5) \times 10^6$

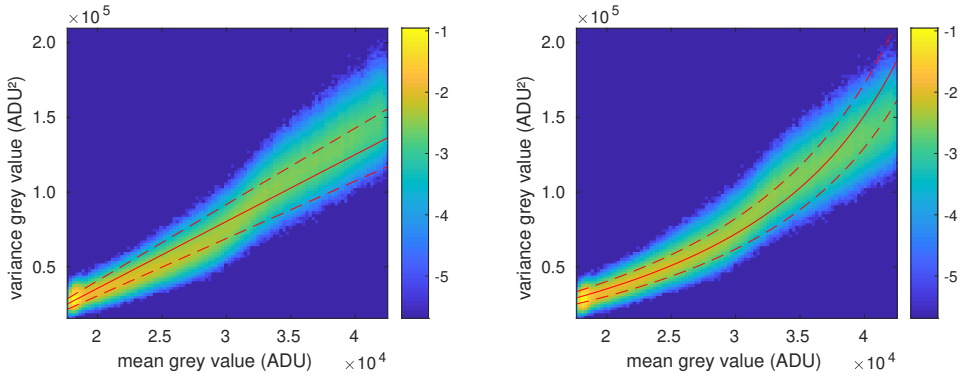
Canonical Link		null	bw	linear
AIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.5 \pm 0.2) \times 10^6$	$(6.5 \pm 0.2) \times 10^6$	$(6.5 \pm 0.3) \times 10^6$
BIC	order	-1 to 0	-1 to 0	-1 to 0
	votes	10	10	10
	criterion	$(6.7 \pm 0.2) \times 10^6$	$(6.4 \pm 0.2) \times 10^6$	$(6.3 \pm 0.2) \times 10^6$

Table 5.4: Forward stepwise selection was used to find suitable polynomial features when fitting a gamma GLM onto the sample variance-mean data from the projections in `AbsFilter` at 120°. The columns of the table represent different shading corrections. Forward stepwise selection was repeated 10 times by bootstrapping the replicate projections. 'Order' shows the most commonly selected polynomial orders and 'votes' shows the number of times it was selected out of the 10 repeats. The error bars are the standard deviation from the 10 repeats.



(a) $y(x) = \beta_0 + \beta_1 x$ at 30°

(b) $y(x) = (\beta_{-1} x^{-1} \beta_0)^{-1}$ at 30°



(c) $y(x) = \beta_0 + \beta_1 x$ at 120°

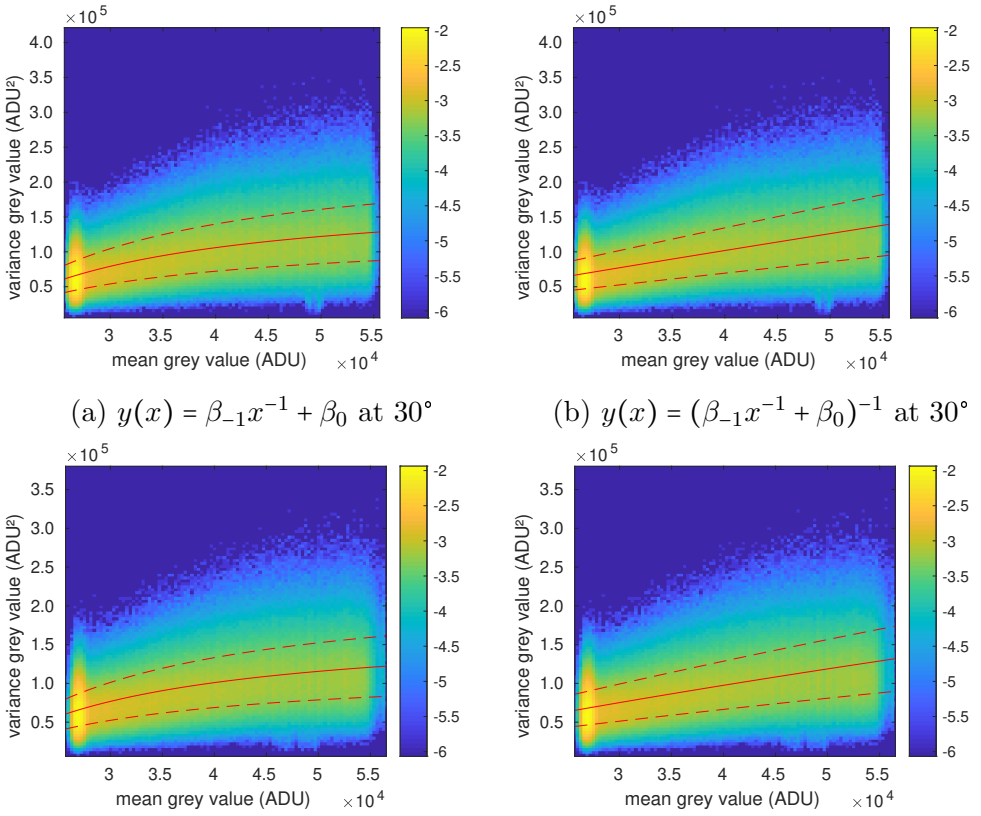
(d) $y(x) = (\beta_{-1} x^{-1} \beta_0)^{-1}$ at 120°

Figure 5.1: Log frequency density histogram of the sample variance-mean data from `AbsNoFilter` with linear shading correction. The solid red line shows the GLM fit along with the 68% prediction interval as dashed lines. The colour scales are in units of $\log \mathrm{ADU}^{-3}$.

It was chosen that the training and test set are of the same size, a 50:50 spilt. The sizes of the sets can be different, for example, a 75:25 spilt. A 50:50 spilt was chosen because a large training set is an overkill for a high data and low dimensional problem. Here, a model with a few parameters is fitted onto millions of pixels.

The mean scaled deviance was used to assess the performance of variance prediction of the test set. The deviance is defined to be

$$D = 2\left(\ln L_{\mathrm{s}} - \ln L_{\tau}\right) \tag{5.14}$$



(a) $y(x) = \beta_{-1}x^{-1} + \beta_0$ at 30°

(b) $y(x) = (\beta_{-1}x^{-1} + \beta_0)^{-1}$ at 30°

(c) $y(x) = \beta_{-1}x^{-1} + \beta_0$ at 120°

(d) $y(x) = (\beta_{-1}x^{-1} + \beta_0)^{-1}$ at 120°

Figure 5.2: Log frequency density histogram of the sample variance-mean data from `AbsFilter` with linear shading correction. The solid red line shows the GLM fit along with the 68% prediction interval as dashed lines. The colour scales are in units of $\log \mathrm{ADU}^{-3}$.

where

$$\ln L_\tau = \sum_{i \epsilon \tau} \left[\alpha \ln \alpha - \ln \Gamma(\alpha) - \alpha \ln \widehat{y}_i + (\alpha - 1) \ln y_i - \frac{\alpha y_i}{\widehat{y}_i} \right] \tag{5.15}$$

is the log-likelihood of the test set given the fitted model and $\ln L_\mathrm{s}$ is the saturated log-likelihood. This is obtained by replacing all $\widehat{y}_i$ with y_i in $\ln L_\tau$ so that

$$\ln L_\mathrm{s} = \sum_{i \epsilon \tau} \left[\alpha \ln \alpha - \ln \Gamma(\alpha) - \ln y_i - \alpha \right] \ . \tag{5.16}$$

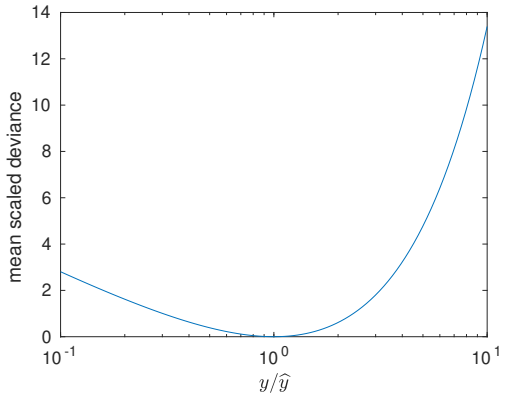


Figure 5.3: The scaled deviance loss function

Following from this, the deviance is

$$D = 2\alpha \sum_{i \in \tau} \left[\frac{y_i - \widehat{y}_i}{\widehat{y}_i} - \ln\left(\frac{y_i}{\widehat{y}_i}\right) \right] . \tag{5.17}$$

The mean scaled deviance is obtained by removing the factor of α and dividing by N_τ to get

$$D_s = \frac{2}{N_\tau} \sum_{i \in \tau} \left[\frac{y_i - \widehat{y}_i}{\widehat{y}_i} - \ln\left(\frac{y_i}{\widehat{y}_i}\right) \right] . \tag{5.18}$$

The mean scaled deviance is a loss function which increases as $y_i/\widehat{y}_i$ deviates from one, this is shown in Figure 5.3 and it should be noted that the x-axis is in log scale. Another way to show this is by letting $r_i = y_i/\widehat{y}_i$ and $d_i = 2\left[r_i - 1 - \ln r_i\right]$ be an element from the sum in the deviance. For $r_i \approx 1$, $\ln(r_i) \approx (r_i - 1) - (r_i - 1)^2/2$ which implies that $d_i \approx (r_i - 1)^2$ with minimum at one. For r_i deviate greatly from one, the loss function is asymmetric. For example a ratio of $r_i = 10^1$ has a greater penalty than $r_i = 10^{-1}$. This means that in extreme cases, overestimates are penalised less than underestimates relative to y_i.

Assuming the model is correct, it is given that $D \sim \chi^2_{N_\tau - k}$ which implies that for large N_τ,

$$\mathbb{E}\left[D_s\right] = \frac{1}{\alpha} \tag{5.19}$$

and

$$\mathbb{V}\mathrm{ar}\left[D_s\right] = \frac{2}{\alpha^2 N_\tau} . \tag{5.20}$$

Because $\alpha = (m-1)/2$, this shows that the number of replicated projections used to obtain the sample variance-mean data has an influence on the mean scaled deviance. This result can be used to estimate α if it is unknown.

Cross validation was performed on the datasets `AbsNoFilter` and `AbsFilter` with various shading corrections. The models selected from forward stepwise selection in the previous section were assessed. They are $y(x) = \beta_0 + \beta_1 x$ and $y(x) = \beta_{-1}x^{-1} + \beta_0$ using the identity link and $y(x) = (\beta_{-1}x^{-1} + \beta_0)^{-1}$ using the canonical link. The analysis was repeated 100 times by using a random permutation with replacement of the replicate projections to obtain a different sample variance-mean data which introduced some variation to the data (Efron, 1979).

The results from the cross-validation is shown in Figure 5.4. The performances of the three candidate models were very similar. One exception is the model $y(x) = \beta_{-1}x^{-1} + \beta_0$ fitted onto the `AbsNoFilter` dataset because the mean scaled deviance was significantly larger. This is expected from a model not favoured in the forward stepwise selection in the previous section.

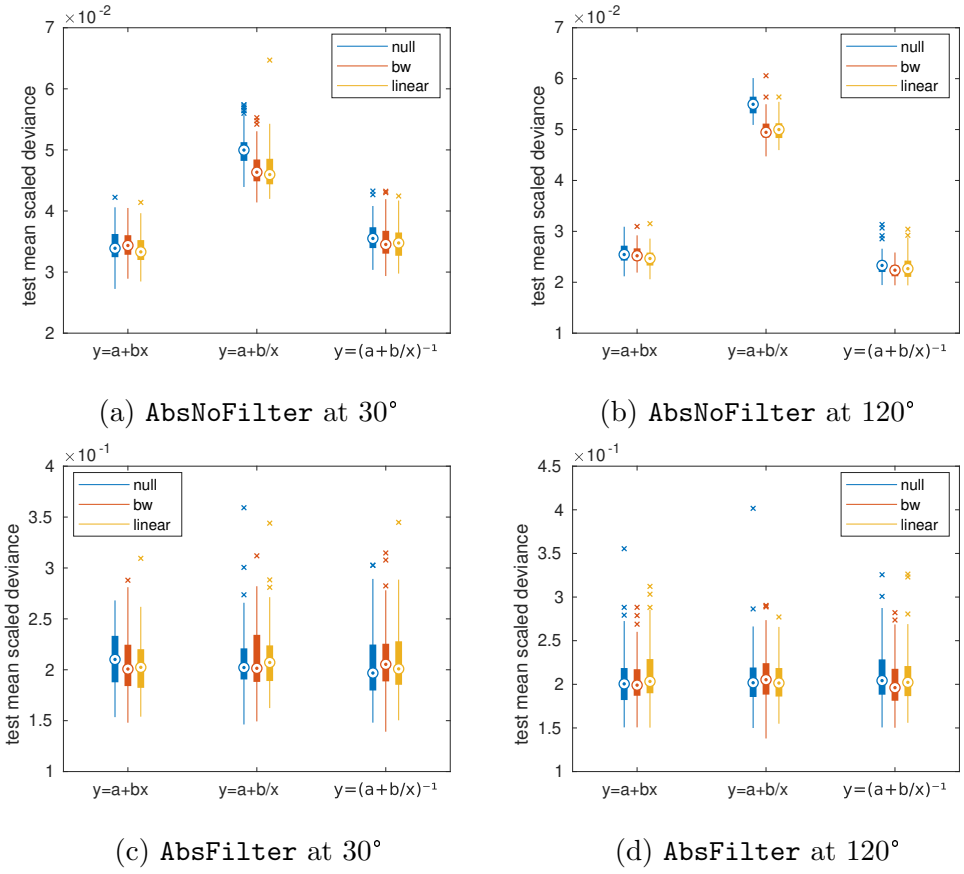
Shading correction did not have a significant effect on the analysis.

From these results, it is recommended that the relationship $y(x) = \beta_0 + \beta_1 x$ should be used for its simple form, similar performance to other candidate models and connections to the compound Poisson.

5.4 Residual Analysis

Residual analysis was conducted to study for anything overlooked. The model $y(x) = \beta_0 + \beta_1 x$ was fitted onto the entire sample variance-mean data. A residual plot plots $r_i = y_i - \widehat{y}(x_i)$ for all x_i in the dataset which was done using a 2D histogram. Due to the gamma GLM, higher grey values typically have bigger residuals in magnitude. A prediction residual interval is included in the histogram to aid judging the residuals. This interval was acquired by obtaining the prediction interval and subtracting it from the fit.

The residual plots are shown in Figure 5.5. The residuals are all sensible and increased in magnitude with the grey value which is captured by



(a) `AbsNoFilter` at 30°

(b) `AbsNoFilter` at 120°

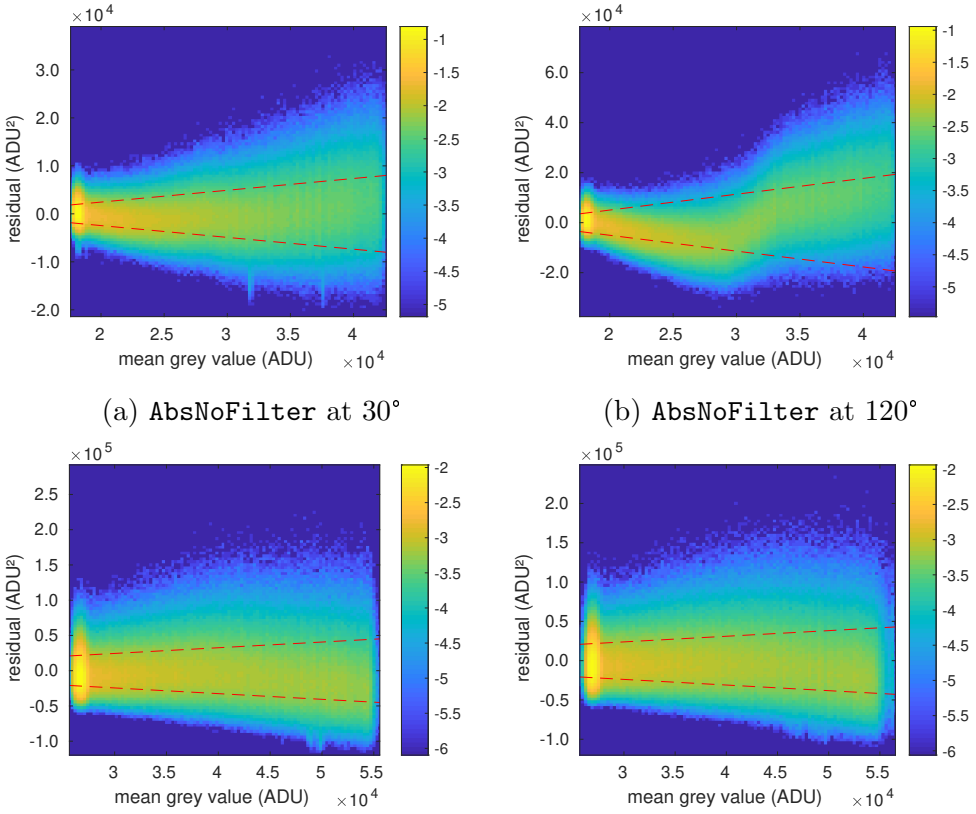
(c) `AbsFilter` at 30°

(d) `AbsFilter` at 120°

Figure 5.4: Test mean scaled deviance from predicting variances in a test set using a GLM fitted onto a training set. The different colours represent different shading corrections. The boxplots represent the 100 repeats of the analysis by bootstrapping the replicate projections.

the prediction residual interval. The inflection in `AbsNoFilter` at 120°, as discussed previously, can be seen more clearly here.

Because no spatial information was used in modelling the variance, the residuals were plotted on top of the projections to look for any spatial structure, this is shown in Figure 5.6. Because of the geometry of the test sample, lower grey values, thus smaller residuals, were found in the middle of the projection. Similarly, higher grey values, thus bigger residuals, were found on the edges of the test sample. Considering that, the residuals do not show any striking spatial structure. Again, the inflection in `AbsNoFilter`



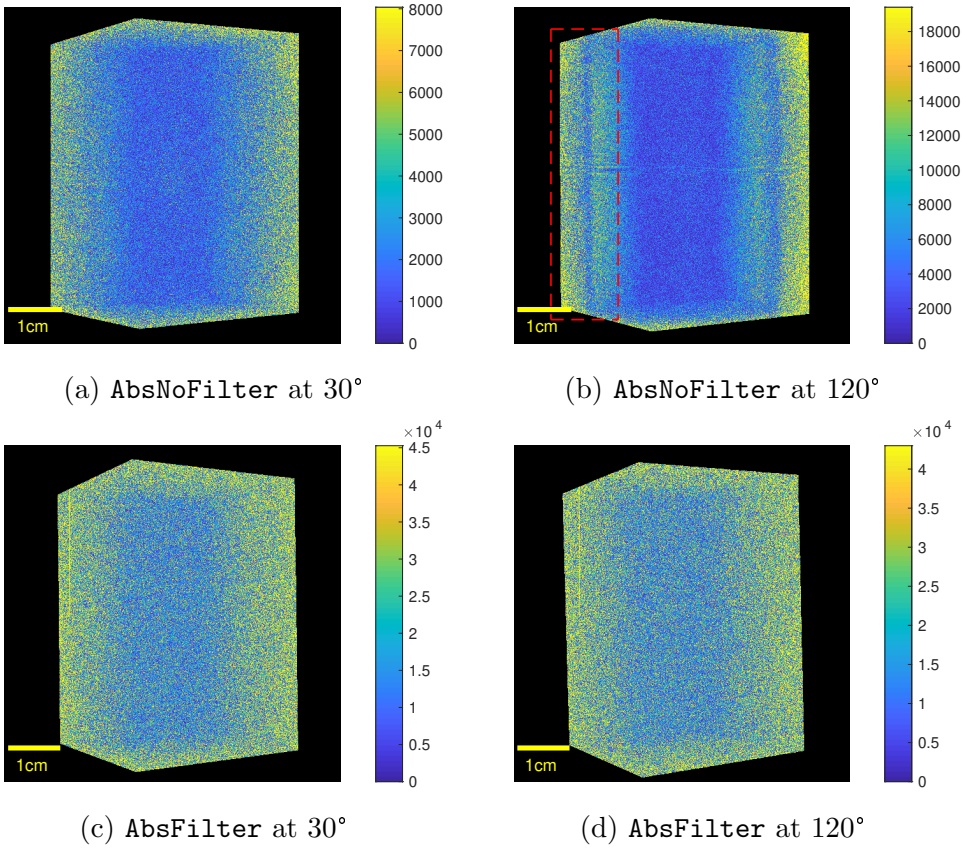
(a) `AbsNoFilter` at 30°

(b) `AbsNoFilter` at 120°

(c) `AbsFilter` at 30°

(d) `AbsFilter` at 120°

Figure 5.5: Log frequency density histogram of the residuals given the mean grey value. The residuals are from fitting a gamma GLM of the form $y(x) = \beta_0 + \beta_1 x$ onto the sample variance-mean data obtained from the replicate projections. The dashed red line shows the 68% prediction residual interval. The colour scales are in units of $\log \mathrm{ADU}^{-3}$.



(a) `AbsNoFilter` at 30° (b) `AbsNoFilter` at 120°

(c) `AbsFilter` at 30° (d) `AbsFilter` at 120°

Figure 5.6: Absolute value residuals from fitting a gamma GLM of the form $y(x) = \beta_0 + \beta_1 x$ onto the sample variance-mean data obtained from the replicate projections. Highlighted in a dashed red box is an example of an inflection. The colour scales are in units of ADU^2.

at 120° can be seen more clearly here.

5.5 Conclusion

Various parametric models were investigated. A gamma GLM with identity and canonical link functions were looked at with different polynomial features. In terms of cross-validation, most of the models looked at have similar prediction performance when predicting the variance given a grey value. Thus, the relationship $\widehat{y}(x) = \beta_0 + \beta_1 x$ using the identity link is

attractive for its simplicity and links with the compound Poisson model.

It was found that shading correction had little effect on the variance-mean relationship.

As discussed in the compound Poisson chapter, a linear relationship can be derived by assuming no beam hardening, then it should be expected that a linear relationship is observed in the dataset `AbsFilter`. In the forward stepwise selection, `AbsNoFilter` consistently selected the linear model, given the identity link function. `AbsFilter` did not select the linear model but in terms of cross-validation, the linear model was just as good as the other models.

It was unusual to see an inflection in the sample variance-mean data in `AbsNoFilter` at 120° which was not captured by the GLM. A good explanation was not found but perhaps the absence of an x-ray filter may have contributed to this. As a result, the `AbsNoFilter` dataset was discarded in favour of the `AbsFilter` dataset from the next chapter.

There exist model selection methods such as lasso and elastic net (Tibshirani, 1996; Zou and Hastie, 2005; Friedman et al., 2010) but they are catered for high dimensional problems involving hundreds of parameters. In this problem, only a few parameters were needed to find a good model thus forward stepwise selection was sufficient. More flexible models such as non-parametric and machine learning methods may be used, however, they are slow and are unnecessary in a low dimensional problem with a large number of data points.

With the ability to predict the variance given a grey value of a pixel, the uncertainty can be quantified. In the next chapter, the projection is compared to a simulated projection in the face of that uncertainty.

Chapter 6

Inference

Disagreements between the obtained and the *aRTist* simulated projections can be found by subtracting one projection from the other. Any values too big in magnitude can be considered as a defect. However, in the previous chapters, it was found that x-ray photons behave randomly and differences in the comparison can be due to chance. Thus, the comparisons should be done in the face of uncertainty.

A pixel by pixel hypothesis test was proposed to do defect detection. To illustrate this method of inference, the projections from the dataset `AbsFilter` at 120° was used. To recap, 20 replicate projections of a test sample, with purposefully manufactured voids, were obtained. *aRTist* was used to create a simulation of that projection but as if the voids were not there. Thus, the method should pick these voids up.

Beforehand, linear shading correction was applied to the projections. The *aRTist* projections were shading corrected using the simulated greyscale projections. The 20 replicate projections were split into two. 19 randomly selected projections were used for the variance-mean model to fit onto. The remaining projection was compared with the *aRTist* projection. This remaining projection and the *aRTist* projections are shown in Figures 6.1a and 6.1b respectively.

A gamma GLM, with a basic linear relationship, was used for the variance-mean model, as described in Chapter 5. The variance was pre-

dicted using the grey value in the *aRTist* simulation as the predictor variable.

The test statistic, for the pixel located at (x, y), is

$$Z_{x,y} = \frac{\text{projection}_{x,y} - aRTist_{x,y}}{\sqrt{\widehat{y}[aRTist_{x,y}]}} \tag{6.1}$$

where $\widehat{y}[aRTist_{x,y}]$ is the predicted grey value variance. The test statistic was calculated for each pixel in the ROI, that is, pixels which represent the test sample. The ROI was created by manually segmenting the test sample from the projection.

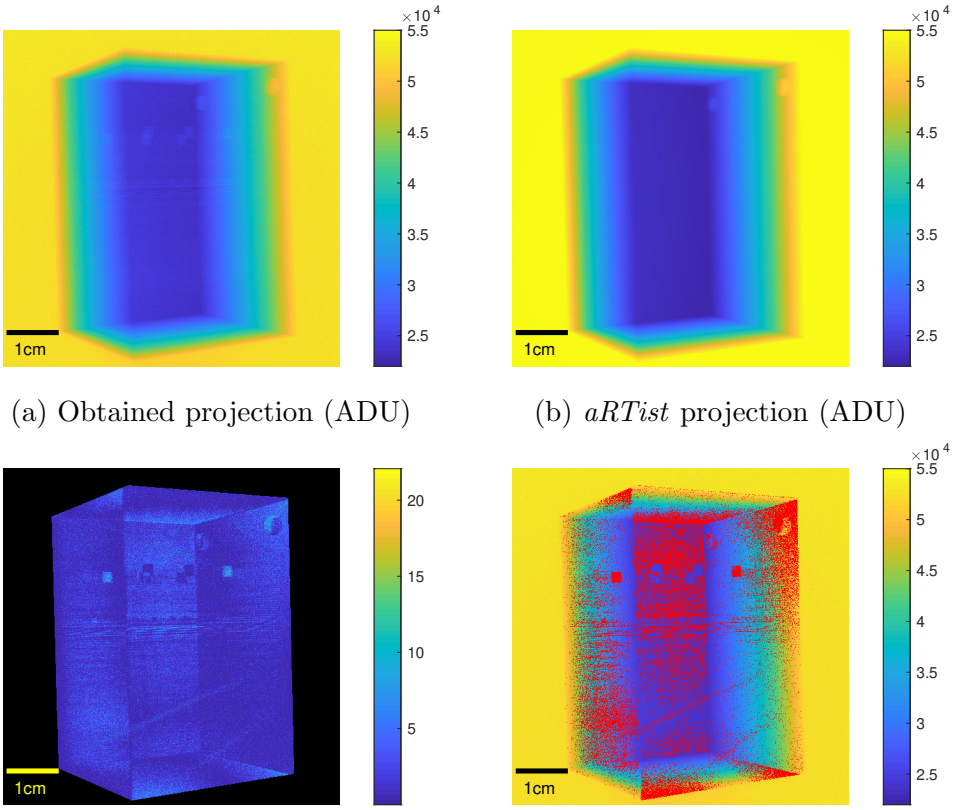
It is important to identify which quantities are random and which are not. For high photon rates, it was shown in the previous chapters that the grey values in the projection can be modelled using a Normal distribution. Thus, $\text{projection}_{x,y}$ is a random quantity. The simulation $aRTist_{x,y}$ is not random because this was obtained through computer software. Given the 19 selected projections used for training the variance-mean model, the variance prediction $\widehat{y}[aRTist_{x,y}]$ is not random because the variance-mean model was fitted before encountering the remaining projection to be compared with *aRTist*. The variance-mean model can be made random if a different set of 19 projections was used to train the model each time this inference was conducted, but this shall not be considered here.

If there are no significant differences between the obtained and *aRTist* projections, then the randomness of the test statistics can be approximately quantified as

$$Z_{x,y} \sim \text{N}(0, 1) \ . \tag{6.2}$$

As with usual statistical convention, an upper case Z denotes a random variable. A lower case z denotes a realisation or an observation of that random variable.

For each pixel, a test statistic was calculated, which forms a z image. A test statistic too large in magnitude, relative to the anticipated variance, will classify that pixel as a positive result. In two-tailed hypothesis testing (Pearson, 1900; Neyman and Pearson, 1933; Fisher, 1970), p-values can be



(a) Obtained projection (ADU)

(b) *aRTist* projection (ADU)

(c) $-\log p$-values

(d) Positive pixels overlaid on a)

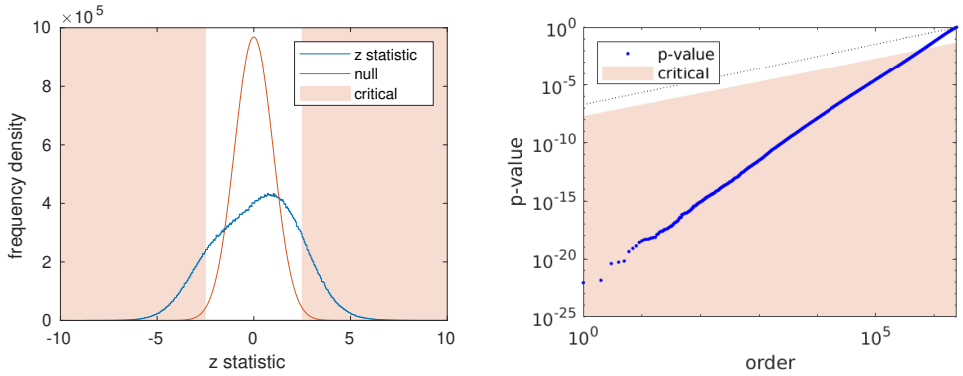
Figure 6.1: The obtained projection of the test sample (a), from the `AbsFilter` at 120° dataset, was compared to the *aRTist* projection (b) to detect purposefully manufactured voids. The p-values (c) obtained were used for hypotheses testing. Pixels detected as positive are shown in red in d) using the Benjamini and Hochberg (1995) procedure at the 5% false discovery rate level.

used to represent the test statistics in a different way

$$p_{x,y} = 2(1 - \Phi(|z_{x,y}|)) \tag{6.3}$$

which can takes values $0 \leqslant p_{x,y} \leqslant 1$. A p-value too small is considered a positive result. The p-values are shown in Figure 6.1c.

The resulting p-values are concerning. This is because the p-values are not very smooth on the surfaces of the sample. It should be expected that



(a) Histogram (b) p-values

Figure 6.2: a) The histogram of the test statistics, from the `AbsFilter` projection at 120°, is compared with the standard Normal distribution. The standard Normal distribution is known as the null distribution in this scenario. b) The p-values are ordered and plotted. The critical region corresponds to controlling the false discovery rate at the 5% level using the Benjamini and Hochberg (1995) procedure. The dotted line shows the result if the p-values were uniformly distributed.

small p-values are in areas of the defects. Pixels were tested positive or considered to be evidence of a defect, when $|z_{x,y}| > 2.48$ to 2 decimal places. This value was chosen by controlling the false discovery rate at 5% using the Benjamini and Hochberg (1995) (BH) procedure. The positive pixels are shown in Figure 6.1d.

This proposed method for defect detection failed because too many false positives were detected. These false positives appeared to have some structure, for example, clustering in the corners or on surfaces. Also, false negatives were detected because not all of the defects were detected.

Model misspecification appeared to be the main source of error. The test statistics and p-values were inspected in Figure 6.2. It can be seen that the test statistics were not compatible with the standard Normal distribution. Also, the majority of the p-values did not look uniformly distributed. This seems to suggest that the assumption of $Z_{x,y} \sim \mathrm{N}(0,1)$ is incorrect.

This chapter recaps hypothesis testing, for a single test and then for

multiple tests, treating each pixel as a test. Assumptions used in the hypotheses testing can be relaxed by using the empirical null (Efron, 2004), which is reviewed here. The empirical null is then extended to an image filter, called the empirical null filter. This filter adjusts each test statistic according to its neighbours, ironing out false positive results. Simulations and results are shown towards the end of the chapter.

6.1 Literature Review

Consider a test statistic Z from a pixel. If there are no defects, then the statistic is null and has the null distribution $Z|H_0 \sim \mathrm{N}(0,1)$. This can be described by specifying the random variable as $Z \sim \mathrm{N}(\mu,1)$ and the null hypothesis as $H_0 : \mu = 0$. A hypothesis test can judge how much Z deviates from 0 by defining the alternative hypothesis $H_1 : \mu \neq 0$. A statistic which does not has the null distribution is known as non-null.

In two-tailed hypothesis testing (Pearson, 1900; Neyman and Pearson, 1933; Fisher, 1970), the p-value can be compared with the user-defined size of the test α, also known as the significance level in this specific example. A positive result is declared when $p < \alpha$, else it is a negative result. Correct and incorrect testing of the statistics can occur. For example, testing a null statistic as positive is known as a false positive. It can be shown that α controls the false positive rate. $\alpha = 5.0\%$ is a typical choice (Wasserstein et al., 2019) and is used throughout this thesis.

Multiple hypotheses testing occurs when there are more than one hypotheses to test. For example, N pixels can be tested where the test statistics are $Z_1, Z_2, \ldots, Z_N$ and $Z_i \sim \mathrm{N}(\mu_i, 1)$ for $i = 1, 2, \ldots, N$. The null hypotheses are $H_{0,i} : \mu_i = 0$ and are tested against the alternative $H_{1,i} : \mu_i \neq 0$ for $i = 1, 2, \ldots, N$. Let the corresponding p-values be $p_1, p_2, \ldots, p_N$.

The uncorrected test classify any $p_i < \alpha$ as positive. This is flawed because the possibility of obtaining at least one false positive increases as the number of tests increases (Shaffer, 1995). This method does control the per-comparison error rate (PCER) (Benjamini and Hochberg, 1995). PCER is the proportion of false positives out of all tests. The notation for

	Negative	Positive	Total
Null	U	V	N_0
Non-null	T	S	$N - N_0$
	$N - R$	R	N

Table 6.1: Random variable definitions for the number of true/false positives/negatives made in multiple hypotheses testing

the number of true/false positive/negatives obtained are defined in Table 6.1. Using the notation, the PCER is defined as

$$\text{PCER} = \frac{1}{N}\mathbb{E}[V] \ . \tag{6.4}$$

It can be seen that if the uncorrected test controls the false positive rate such that $\mathbb{E}[V]/N_0 = \alpha$, then it controls the PCER such that

$$\text{PCER} \leqslant \alpha \ . \tag{6.5}$$

The Bonferroni correction (Shaffer, 1995; Bland and Altman, 1995; Perneger, 1998) controls the family-wise error rate (FWER) (Shaffer, 1995) where

$$\text{FWER} = \mathbb{P}(V \geqslant 1) \ . \tag{6.6}$$

This is done by adjusting the size of the test to be α/N. By using the adjusted size, then $\text{FWER} = 1 - [(1 - \alpha/N)^N]$. Using the approximation $(1 - \alpha/N)^N \approx 1 - \alpha$ then $\text{FWER} \approx \alpha$. This shows that the Bonferroni correction controls the family-wise error rate such that

$$\text{FWER} \leqslant \alpha \ . \tag{6.7}$$

In practice, the Bonferroni correction is not very powerful (Perneger, 1998), meaning it gives too many false negatives. This is because the correction traded too many false positives for false negatives.

The Benjamini and Hochberg (1995) (BH) procedure controls the false discovery rate (FDR) (Benjamini, 2010) rather than the PCER or FWER.

The FDR is the proportion of false positives out of all positive results, that is

$$\text{FDR} = \mathbb{E}\left[\frac{V}{R}\right] . \tag{6.8}$$

It is defined that $V/R = 0$ when $R = 0$.

The BH procedure adjusts the size of the test between the Bonferroni correction and the uncorrected test. It adapts to the data and chooses different sizes for different data. The procedure is as follows, the p-values are ordered such that $p_{(1)} \leqslant p_{(2)} \leqslant \ldots \leqslant p_{(N)}$. Suppose a size α is provided beforehand. The size is adjusted to

$$\alpha_{\text{BH}} = \frac{\alpha k}{N} \tag{6.9}$$

where

$$k \text{ is the largest } i \text{ for which } p_{(i)} \leqslant \frac{i}{N}\alpha . \tag{6.10}$$

This test the statistics with p-values $p_{(1)}, p_{(2)}, \ldots, p_{(k)}$ as positive. For the case where $p_{(1)} > \alpha/N$ then $k = 1$ so that the Bonferroni correction is used when there are no positive results, this is only for illustration purposes. For example in Figure 6.2b, the p-values were plotted against their order so that the critical boundary was shown by a linear curve with gradient α/N.

The BH procedure comes from the fact that if all the statistics are null and independent, then the p-values are uniformly distributed (Simes, 1986). However, it can be shown that the BH procedure works for many scenarios of dependencies (Benjamini and Yekutieli, 2001). It can be shown that the BH procedure controls the FDR such that

$$\text{FDR} \leqslant \alpha \tag{6.11}$$

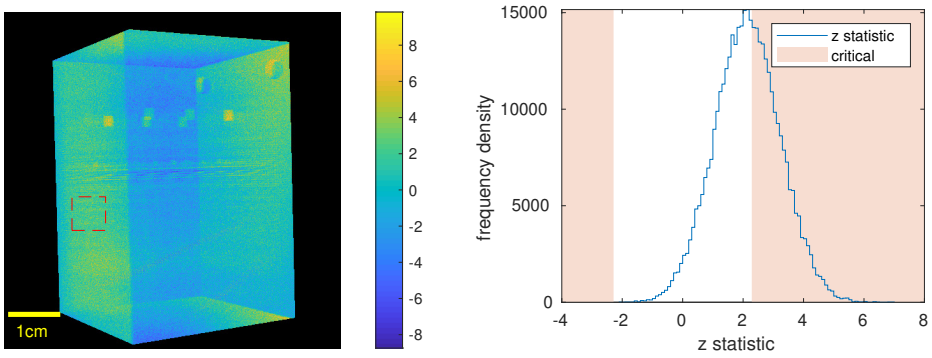
(Benjamini and Hochberg, 1995).

In summary, the uncorrected, Bonferroni and BH correction controls for different error rates according to the threshold α, summarised in Table 6.2.

Consider a small example where a $200\,\text{px} \times 200\,\text{px}$ section of the z image was investigated, as shown Figure 6.3. The distribution of the test statistics appeared Normal but not centred at zero. By using the BH procedure to obtain a critical region of $|Z| > 2.29$ to 2 decimal places, almost half of

Correction	Controls for
No correction	PCER $\leqslant \alpha$
Bonferroni	FWER $\leqslant \alpha$
BH	FDR $\leqslant \alpha$

Table 6.2: Different types of corrections for multiple hypotheses testing are listed here, along with what they control for.



(a) Test statistics (b) Histogram

Figure 6.3: The resulting test statistics for the `AbsFilter` projection at 120° are shown in a). A 200 px × 200 px sub-image was taken shown by the dashed lines. A histogram of the test statistics in the sub-image is shown in b). The critical region corresponds to the 5% FDR level.

the pixels were tested positive at the 5% FDR level. It is questionable whether such many positive results are sensible, in particular, in an area where defects were not expected.

This problem commonly occurs in large scale multiple hypotheses testing (Efron, 2004) such as in microarrays (Hedenfalk et al., 2001; Efron and Tibshirani, 2002; Efron et al., 2003). It appeared that the null distribution was misspecified as $Z_i|H_{0,i} \sim \mathrm{N}(0,1)$ when in reality they are distributed as $Z_i|H_{0,i} \sim \mathrm{N}(\mu_0, \sigma_0^2)$, where μ_0 and σ_0 are the null mean and null standard deviation respectively.

In practice, μ_0 and σ_0 are unknown. The empirical null (Efron, 2004) replaces the parameters of the null distribution with its estimate. Sup-

pose $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ are the estimated null mean and null standard deviation respectively. The statistics are specified by $Z_i \sim \mathrm{N}(\mu_i, \widehat{\sigma}_0^2)$ so that the following null hypotheses $H_{0,i} : \mu_i = \widehat{\mu}_0$ are tested against $H_{1,i} : \mu_i \neq \widehat{\mu}_0$ for $i = 1, 2, \ldots, N$.

In Efron (2004), a few assumptions were made in order to obtain the estimators $\widehat{\mu}_0$ and $\widehat{\sigma}_0$. Let the test statistic Z have the p.d.f.

$$p_Z(z) = \pi_0 p_{Z|H_0}(z) + \pi_1 p_{Z|H_1}(z) \tag{6.12}$$

where $0 \leqslant \pi_0 \leqslant 1$ and $\pi_1 = 1 - \pi_0$. In addition, the null distribution is

$$p_{Z|H_0}(z) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{z-\mu_0}{\sigma_0}\right)^2\right] \tag{6.13}$$

as it was assumed to be Normal. The non-null distribution $p_{Z|H_1}(z)$ does not need to be specified. Assume that the majority of the data are null and non-null test statistics are rare, say $\pi_0 > 0.9$ (Efron, 2004), then around the mode, the probability density function would be dominated by the null distribution. This implies that

$$p_Z(z) \approx \pi_0 p_{Z|H_0}(z) \tag{6.14}$$

for values of z around the mode. Finding the mode for $p_Z(z)$ and $p_{Z|H_0}(z)$ should yield the same solution. This justify the use of the mode for the empirical null mean (Efron, 2004)

$$\widehat{\mu}_0 = \operatorname{argmax} \widehat{p}_Z(z) \tag{6.15}$$

where $\widehat{p}_Z(z)$ is the density estimation of $p_Z(z)$, for example it could be a smoothing spline fitted onto the histogram (Efron, 2004).

The null standard deviation is estimated from the log density (Efron, 2004). For values of z at and around the mode

$$\ln p_Z(z) = \ln\left[\frac{\pi_0}{\sqrt{2\pi}\sigma_0}\right] - \frac{1}{2}\left(\frac{z-\mu_0}{\sigma_0}\right)^2 . \tag{6.16}$$

Taking derivatives

$$\frac{\partial}{\partial z} \ln p_Z(z) = -\left(\frac{z-\mu_0}{\sigma_0^2}\right) \tag{6.17}$$

$$\frac{\partial^2}{\partial z^2} \ln p_Z(z) = -\frac{1}{\sigma_0^2} \tag{6.18}$$

which motivates the estimator (Efron, 2004)

$$\widehat{\sigma}_0 = \left[-\frac{\partial^2}{\partial z^2} \ln \widehat{p}(z) \Big|_{z=\widehat{\mu}_0} \right]^{-1/2} . \tag{6.19}$$

The evaluation of $z = \widehat{\mu}_0$ is used because at the mode, it was assumed that the null test statistics would dominate and non-null test statistics would not contribute much to the density estimate.

This method does not need estimations of π_0 which makes it quite convenient. Estimations of π_0 are discussed in literature such as Benjamini and Hochberg (2000); Pounds and Morris (2003); Storey and Tibshirani (2003); Pounds and Cheng (2004); Langaas et al. (2005); Durnez et al. (2014). Efron (2004) motivated the use of the empirical null by investigating various types of false discovery rates (Storey, 2002, 2003; Efron and Tibshirani, 2002; Efron, 2007) which will not be discussed here.

The empirical null is used widely, including in neuroimaging (Schwartzman et al., 2008a, 2009) and has been extended to include non-Normal null distributions (Schwartzman et al., 2008a,b). There exist various methods for the empirical null, for example, Schwartzman et al. (2008b) used Poisson regression on the histogram counts in regions where the null statistics dominate. The empirical characteristic function was used in Jin and Cai (2007).

6.2 Empirical Null

Returning to the sub-image example in Figure 6.3. The histogram was smoothed using a kernel density estimate (Parzen, 1962; Friedman et al., 2001) as shown in Figure 6.4. The kernel density estimate (Parzen, 1962) is

$$\widehat{p}_Z(z) = \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{z_i - z}{h} \right) \tag{6.20}$$

where h is the bandwidth, $\phi(x)$ is the standard Normal density and $n = N$ is number of terms in the summation. The bandwidth was chosen such that

$$h = (0.9n^{-1/5} + 0.16) \times \min(s_z, \mathrm{IQR}_z/1.34) \tag{6.21}$$

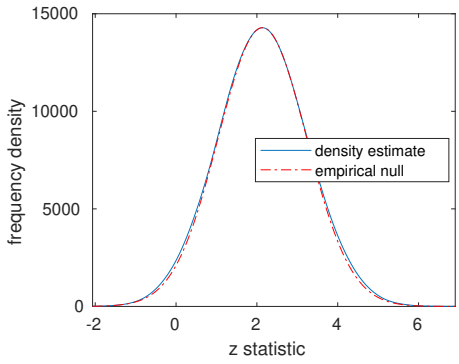


Figure 6.4: The density of the test statistics, from Figure 6.3, was estimated using a kernel density estimate. The dot-dashed line shows the empirical null distribution multiplied by some constant, this is to illustrate that the two densities have the same curvature at the mode.

where s_z and IQR_z are the sample standard deviation and sample interquartile range. This bandwidth is justified later on in this section.

The null mean was estimated by numerically finding the mode of the density estimate. The null standard deviation was estimated using Equation (6.19). This resulted in the empirical null density to have the same curvature as the density estimate at the mode, this is illustrated in Figure 6.4. In this particular example, it was found that $\widehat{\mu}_0 = 2.14$ and $\widehat{\sigma}_0 = 1.09$ to 2 decimal places.

The use of the kernel density estimate has its advantages. The only tuning parameter is the bandwidth h and the density estimate is simple enough to do calculus on it. Speed may be an issue as an evaluation of the density estimate requires the sum over the n test statistics.

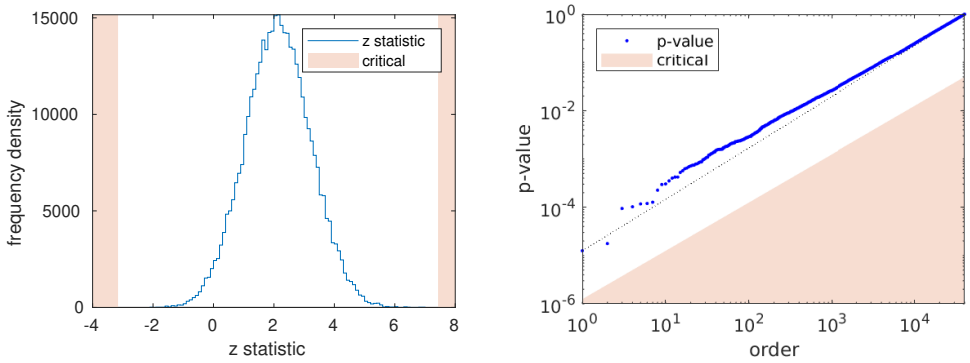
The test statistics were normalised to $T_1, T_2, \ldots, T_N$ by using

$$T_i = \frac{Z_i - \widehat{\mu}_0}{\widehat{\sigma}_0} \tag{6.22}$$

for $i = 1, 2, \ldots, N$. The normalised p-values were obtained by using

$$p_i = 2(1 - \Phi(|t_i|)) \tag{6.23}$$

and were used in the BH procedure. The critical region was found to be $|t| \geqslant 4.85$ at the 5% FDR level. In terms of the original units, the critical



(a) Histogram (b) p-values

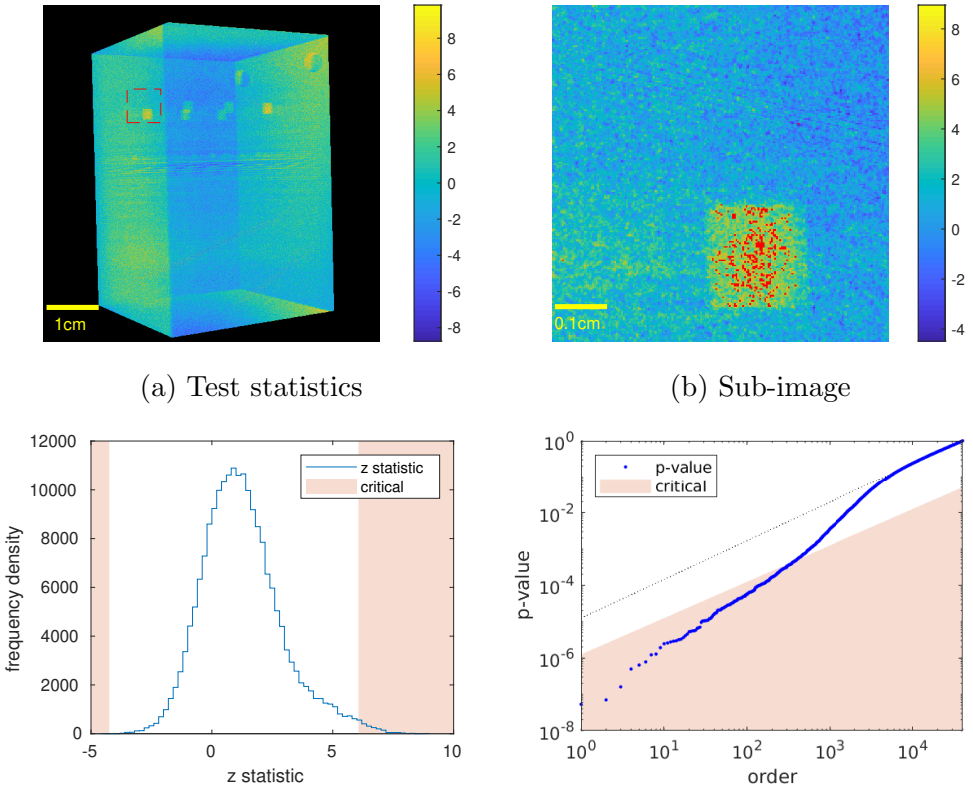
Figure 6.5: Histogram and p-values of the test statistics from the sub-image in Figure 6.3. The critical regions were adjusted using the empirical null and correspond to the 5% FDR level.

region was $z \leqslant -3.16$ and $7.43 \leqslant z$, all to 2 decimal places. The critical region and the normalised p-values are shown in Figure 6.5.

No positive results were found and the p-values were sensible as they resemble the uniform distribution. This demonstrated that the empirical null adjusted the parameters of the null distribution to fit onto the majority of the data to make a sensible inference.

Another example is shown in Figure 6.6 where a $200\,\text{px} \times 200\,\text{px}$ sub-image containing a defect was used. It can be seen that the null distribution was not centred at zero which can be taken into account by using the empirical null. The estimation of the empirical null parameters was robust as it depends on the density estimate at the mode only, it should not be affected by non-null test statistics.

The BH procedure was conducted using the normalised p-values. 389 pixels were tested positive at the 5% FDR level. The majority of the positive pixels were found to be clustered together which highlighted the defect. The entire area of the defects was not tested positive but a good portion of them are. This should be enough to raise suspicion in that particular area. Only a few pixels were falsely tested as positive, however, they are typically isolated single pixels. Isolated positive pixels should be discarded as they are more



(a) Test statistics

(b) Sub-image

(c) Histogram

(d) p-values

Figure 6.6: The resulting test statistics for the `AbsFilter` projection at 120° are shown in a). A 200 px × 200 px sub-image was taken shown by the dashed lines. Positive pixels are shown in red in b). The histogram of the test statistics in the sub-image are shown in c) along with the critical region. The p-values are shown in d). The critical regions were adjusted using the empirical null and correspond to the 5% FDR level.

than likely to be tested positive by random chance.

This section describes the empirical null implementation in detail, in particular, the numerical and computational aspects. The empirical properties of the sampling distribution of the null parameter estimators were studied.

Mode Finding

The mode was found by solving $\widehat{\mu}_0 = \operatorname{argmax} \widehat{p}_Z(z)$ numerically. This was done by using the Newton-Raphson method to solve

$$\frac{\partial}{\partial z} \ln \widehat{p}_Z(z) = 0 \tag{6.24}$$

for z which finds stationary points. The method is an iterative algorithm and requires an initial value $z^{(0)}$. The iterative step is

$$z^{(r+1)} = z^{(r)} - \frac{\left.\frac{\partial}{\partial z} \ln \widehat{p}_Z(z)\right|_{z=z^{(r)}}}{\left.\frac{\partial^2}{\partial z^2} \ln \widehat{p}_Z(z)\right|_{z=z^{(r)}}} \tag{6.25}$$

for $r = 0, 1, 2, 3, \ldots$ until some convergence condition is met. The derivatives for the log density were obtained via the following. Recall the kernel density estimate in Equation (6.20), the log density is

$$\ln \widehat{p}_Z(z) = \ln\left(\frac{1}{nh}\right) + \ln\left[\sum_{i=1}^{n} \phi\left(\frac{z_i - z}{h}\right)\right] . \tag{6.26}$$

Taking the first-order derivative

$$\frac{\partial}{\partial z} \ln \widehat{p}_Z(z) = \frac{1}{\sum_{i=1}^{n} \phi\left(\frac{z_i - z}{h}\right)} \times \sum_{i=1}^{n} \phi'\left(\frac{z_i - z}{h}\right)\left(-\frac{1}{h}\right) .$$

Using the fact that $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$, then $\phi'(z) = -z\phi(z)$. This is used to simplify the equation to be

$$\frac{\partial}{\partial z} \ln \widehat{p}_Z(z) = \frac{\sum_{i=1}^{n} \phi\left(\frac{z_i - z}{h}\right)\left(\frac{z_i - z}{h}\right)}{h \sum_{i=1}^{n} \phi\left(\frac{z_i - z}{h}\right)} . \tag{6.27}$$

Taking the derivative again

$$\frac{\partial^2}{\partial z^2}\ln\widehat{p}_Z(z) = \left[h\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\right]^{-2}\times\left\{h\left[\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\right]\right.$$
$$\times\sum_{i=1}^{n}\left[\phi'\left(\frac{z_i-z}{h}\right)\left(-\frac{1}{h}\right)\left(\frac{z_i-z}{h}\right)+\phi\left(\frac{z_i-z}{h}\right)\left(-\frac{1}{h}\right)\right]$$
$$\left.-\left[\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\left(\frac{z_i-z}{h}\right)\right]\left[h\sum_{i=1}^{n}\phi'\left(\frac{z_i-z}{h}\right)\left(-\frac{1}{h}\right)\right]\right\}\ .$$

Using the fact that $\phi'(z) = -z\phi(z)$, then it is simplified to

$$\frac{\partial^2}{\partial z^2}\ln\widehat{p}_Z(z) = \left[h\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\right]^{-2}\times\left\{\left[\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\right]\right.$$
$$\times\left[\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\left(\left(\frac{z_i-z}{h}\right)^2-1\right)\right]-\left[\sum_{i=1}^{n}\phi\left(\frac{z_i-z}{h}\right)\right.$$
$$\left.\left.\left(\frac{z_i-z}{h}\right)\right]^2\right\}\ . \quad (6.28)$$

The convergence criteria were met when either 10 update steps were taken or when

$$\log\left[\left|\frac{\partial}{\partial z}\ln\widehat{p}_Z(z)\Big|_{z=z^{(r)}}\right|\right] < -5 \quad (6.29)$$

at the current step. This was chosen arbitrary to speed up the algorithm without losing too much accuracy. At the end of the algorithm, for a successful convergence it was required, in addition, that

$$\frac{\partial^2}{\partial z^2}\ln\widehat{p}_Z(z)\Big|_{z=z^{(r)}} < 0\ . \quad (6.30)$$

Following a successful convergence, the estimator $\widehat{\sigma}_0$ was calculated straight away.

The algorithm does depend on the initial value so using different initial values were used. Further initial values were generated by sampling from $\mathrm{N}(z^{(0)}, s_z^2)$. This was done multiple times until three valid solutions were obtained. The best solution, the one with the largest $\ln\widehat{p}_Z(\widehat{\mu}_0)$, out of all the different initial values was used as the final answer. The Newton-Raphson method may converge to a local maximum.

It can be noted that this mode finding method can be applied to other distributions in general.

Bandwidth Tuning

The bandwidth h controls how smooth the kernel density estimator is, higher values produce smoother curves (Friedman et al., 2001). Cross-validation methods to select h do exist (Bowman, 1984; Sheather, 2004) but they can be computationally expensive. Rules of thumb (Silverman, 1986; Sheather, 2004) can be used instead and are usually of the form

$$h = bn^{-1/5} \times \min\left(s_z, \mathrm{IQR}_z/1.34\right) \tag{6.31}$$

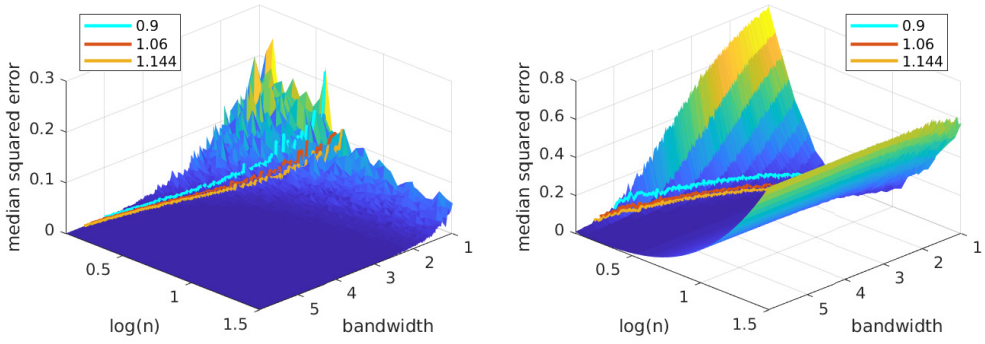
where $b = 0.9$ (Silverman, 1986). This rule of thumb was developed with consideration of bimodal Normal distributions (Silverman, 1986). Other options include $b = 1.06$ and $b = 1.144$ to produce smoother curves (Silverman, 1986; Sheather, 2004).

The kernel density estimator was only used to obtain values of $\widehat{\mu}_0$ and $\widehat{\sigma}_0$. Thus, a good bandwidth, in the context of the empirical null, is one which has good properties of $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ rather than the density estimate. Exact properties of these estimators based on the kernel density estimators can be rather complicated. For example, numerous approximations were needed to show that $\widehat{\sigma}_0^2$ is an unbiased estimator of σ_0^2 to the first-order and higher-order terms are functions of $(\widehat{\mu}_0 - \mu_0)^2$ and h. This is shown in Appendix C.

An experiment was conducted to investigate how the estimators, $\widehat{\mu}_0$ and $\widehat{\sigma}_0$, varied with h and n. This was done on a simulated dataset of n standard Normal random variables. For a given h and n, 100 values of $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ were obtained by repeating the simulation of the dataset. The median squared error was obtained by taking the median over all the squared errors from the 100 estimates. 300 values of n from 10 to 10^6 and 30 values of h from 0.09 to 1.5 were investigated. Figure 6.7 shows the resulting median squared errors with the rules of thumb.

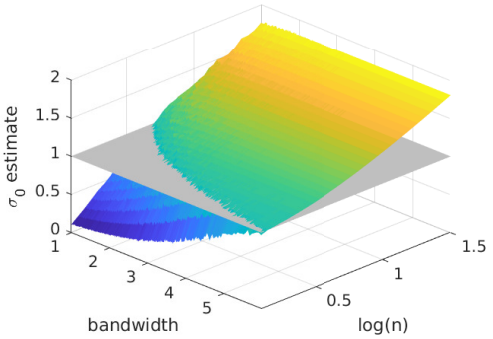
$\widehat{\mu}_0$ has a low median squared error for large bandwidths. For low n, large bandwidths are particularly useful because smoother curves prevent any false bimodal features appearing, making it easier to find the mode.

A smooth valley can be seen for the median squared error for $\widehat{\sigma}_0$. It appeared for a given n, there exist a bandwidth which minimises the median



(a) μ_0 (b) σ_0

Figure 6.7: The median squared error of the estimates of the empirical null parameters over 100 repeats of n simulated standard Normal data. Lines represent the rule of thumb for various values of b.



(a) Heatmap

Figure 6.8: Median value of $\widehat{\sigma}_0$ over 100 repeats of n simulated standard Normal data. The true value is $\sigma_0 = 1$ which is represented by a horizontal plane.

squared error. The rules of thumb did not optimise for $\widehat{\sigma}_0$ and undershot it. Figure 6.8 shows the median of the estimates of σ_0. The true value in this simulation is $\sigma_0 = 1$ which is represented as a horizontal plane. Bandwidths too small underestimated σ_0, bandwidths too big overestimated σ_0.

Overestimates and underestimates are both dangerous and it can affect the false positive and false negative rate. This is because $\widehat{\sigma}_0$ was used to

	Estimate
Intercept	$(1.586 \pm 0.027) \times 10^{-1}$
Gradient	$(9.68 \pm 0.11) \times 10^{-1}$

Table 6.3: The estimated and standard error of the gradient and intercept from the linear relationship in Figure 6.10.

rescale the test statistics accordingly. Recall that the normalised test statistics are $t_i = \dfrac{z_i - \widehat{\mu}_0}{\widehat{\sigma}_0}$. Low values of $\widehat{\sigma}_0$ produce large values, in magnitude, of t_i, potentially testing more of these test statistics as positive. High values of $\widehat{\sigma}_0$ do the opposite. A large number of positives can lead to more false positives and vice versa.

The optimal bandwidth for a given n was found numerically. Figure 6.9 shows the log squared error against the bandwidth for various h given n. A smoothing spline (Friedman et al., 2001) was fitted and this was optimised to find the bandwidth which minimises the log squared error for a given n. *MATLAB*'s `fit(x,y,'smoothingspline')` function was used to fit the smoothing spline which has inbuilt cross validation procedures. The optimisation was done using *MATLAB*'s gradient free `fminsearch` function, using $h = 0.9n^{-1/5}$ as the initial value.

A relationship between the optimal bandwidth and n was attempted to be found. It was assumed such a relationship has the linear form

$$h_{\text{optimal}} = bn^{-1/5} + a \tag{6.32}$$

where b and a are parameters to be fitted. This was chosen to establish a simple relationship with the rules of thumb. Figure 6.10 show the fitting of an identity link gamma GLM in order to estimate b and a. The estimated parameters are shown in Table 6.3.

The rules of thumb can be improved if a small bias is added to it. This is because an intercept was found when fitting a linear relationship between the optimal bandwidth and $n^{-1/5}$. Adding this bias should improve the performance of the estimator $\widehat{\sigma}_0$. The gradient is similar to the rules of thumb in the literature (Sheather, 2004).

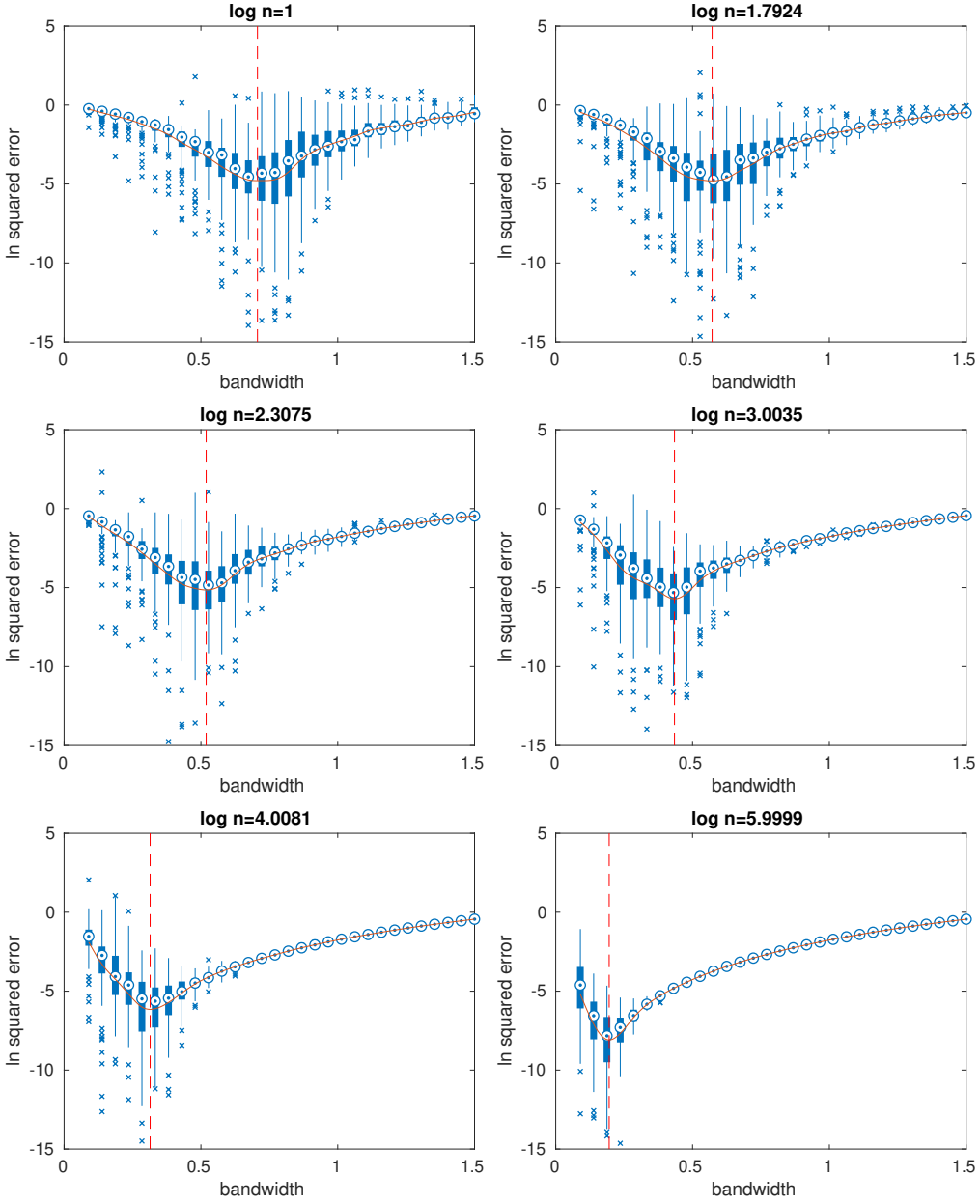


Figure 6.9: A smoothing spline was fitted on the log squared error, for estimating σ_0 on n simulated standard Normal random variables, versus bandwidth. The experiment was repeated by simulating the standard Normal random variables again. The boxplots represent the 100 repeats. The red dashed line shows the bandwidth which minimised the fitted spline.

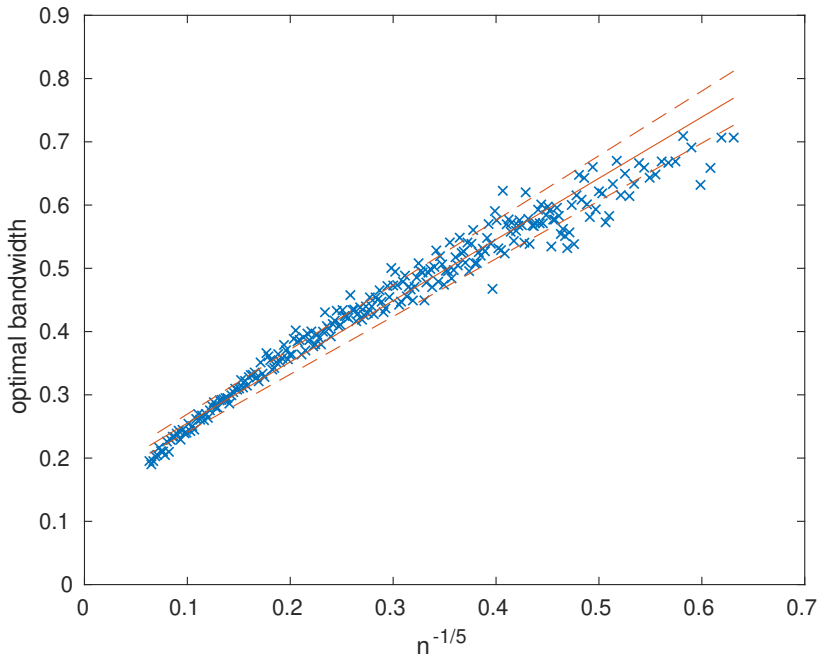


Figure 6.10: An identity link gamma GLM was fitted onto the relationship between the optimal bandwidth and $n^{-1/5}$ from the experiment. The dashed lines show the 68% prediction interval.

The GLM fit in Figure 6.10 appeared to underfit, which suggest there may be a more complicated relationship between the optimal bandwidth and n, for example, a quadratic term can be added to the GLM fit. The method used in fitting the smoothing spline could contribute to the underfitting. For example, if the smoothing spline was not smooth enough so that multiple local minima appeared, it would add variance to the optimal bandwidth if the optimisation algorithm converges to a local minimum. A basic linear relationship was chosen to avoid overfitting and to keep the rule of thumb simple. It would also choose a large bandwidth for small n which would improve the performance $\widehat{\mu}_0$ by working on a smoother density estimate for the price of a less optimal $\widehat{\sigma}_0$.

In conclusion, a small bias added to the rule of thumb improved the performance of the estimator $\widehat{\sigma}_0$. Silverman (1986) pointed out that a smaller bandwidth should be used when, in particular, the distribution is bimodal. This is why Silverman (1986) suggested the use of $b = 0.9$. Following from

this, it was chosen that the intercept was added to Silverman's rule of thumb for the use in the empirical null $h = (0.9n^{-1/5} + 0.16) \times \min(s_z, \text{IQR}_z/1.34)$.

Comparison with Other Robust Estimators

The empirical null mean is effectively the mode and the empirical null standard deviation is a measure of dispersion. It was investigated how they compare with various estimates of central tendency and dispersion for varying sample sizes n. The empirical sampling distribution of these estimators was observed to look for properties such as bias, variance and robustness.

The central tendency estimates looked at were the median, the mean and Tukey's biweighted mean (Beaton and Tukey, 1974). In summary, the biweighted mean weighs each data point using a smooth function of its residual and ignores any data points with residuals greater than a certain value. *MATLAB*'s `robustfit` function was used for the biweighted mean and selects any tuning parameters.

The dispersion estimates looked at were the interquartile range (IQR) ÷1.349, the standard deviation and the median absolute deviation around the mode (MADA-mode) ×1.483. The normalisation constants were used so that MADA-mode and IQR can be compared directly to the standard deviation. MADA-mode was considered to investigate how estimators, which depends on the mode, perform.

In the next section, these estimators are used on a circular kernel with radius r containing πr^2 pixels. To make results comparable, the sample size n may be represented using $r = \sqrt{n/\pi}$.

A sample of standard Normal random variables was simulated to obtain estimates of central tendency and dispersion. This was repeated 100 times to obtain a sampling distribution. The results are shown in Figures 6.11 and 6.12. The mode and empirical null standard deviation both suffered from high variance, much higher than typical robust estimators.

The empirical null standard deviation had some positive bias and it was found that this bias was quite sensitive to the bandwidth. Further fine-tuning on the bandwidth would be required for the empirical null standard

deviation to be unbiased. For example, it appeared there was a consistent positive bias for all r investigated, perhaps suggesting the bandwidth parameter a should be decreased ever so slightly. The source of bias could come from the fact that the bandwidth was tuned to minimise the log squared error rather than the squared bias.

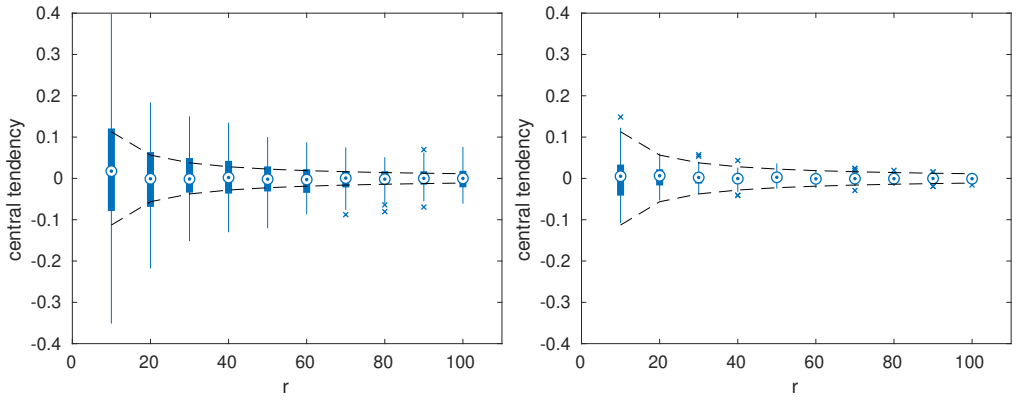
It is interesting to compare the empirical null standard deviation with MADA-mode, in Figure 6.12, because they both depend on the mode. The mode suffered from high variance which may affect dispersion estimators using the mode. The figure shows that MADA-mode had less variance than the empirical null standard deviation. This suggests that using the curvature of the density estimate at the mode was the main source of variance for the empirical null standard deviation.

To test for robustness, a sample of mixture of Gaussian random variables were simulated with the following distribution: $\mathrm{N}(0,1)$ with probability 90%, $\mathrm{N}(3,1)$ with probability 10%. The $\mathrm{N}(3,1)$ component acts as non-null statistics and robust estimators should be unaffected by it. The results are shown in Figures 6.13 and 6.14. The empirical null estimators are robust to the non-null component but still suffers from high variance. The rest of the estimators were affected by the non-null statistics, making them unsuitable to estimate parameters of the null parameters.

6.3 Empirical Null Filter

Returning to the example at the start of the chapter, the empirical null could be used on the histogram shown in Figure 6.2. The resulting empirical null distribution is shown in Figure 6.15. The critical boundary, adjusted using the empirical null, which corresponds to the 5% FDR level was found to be $z \leqslant -11.34$ and $13.39 \geqslant z$ to 2 decimal places. This would test all pixels as negative which is incorrect.

The problem was that the empirical distribution of all of the test statistics looked like it did not come from a Normal distribution. It appeared that the test statistics varied spatially, suggesting that the empirical null varied spatially as well. To account for the spatial variability of the null



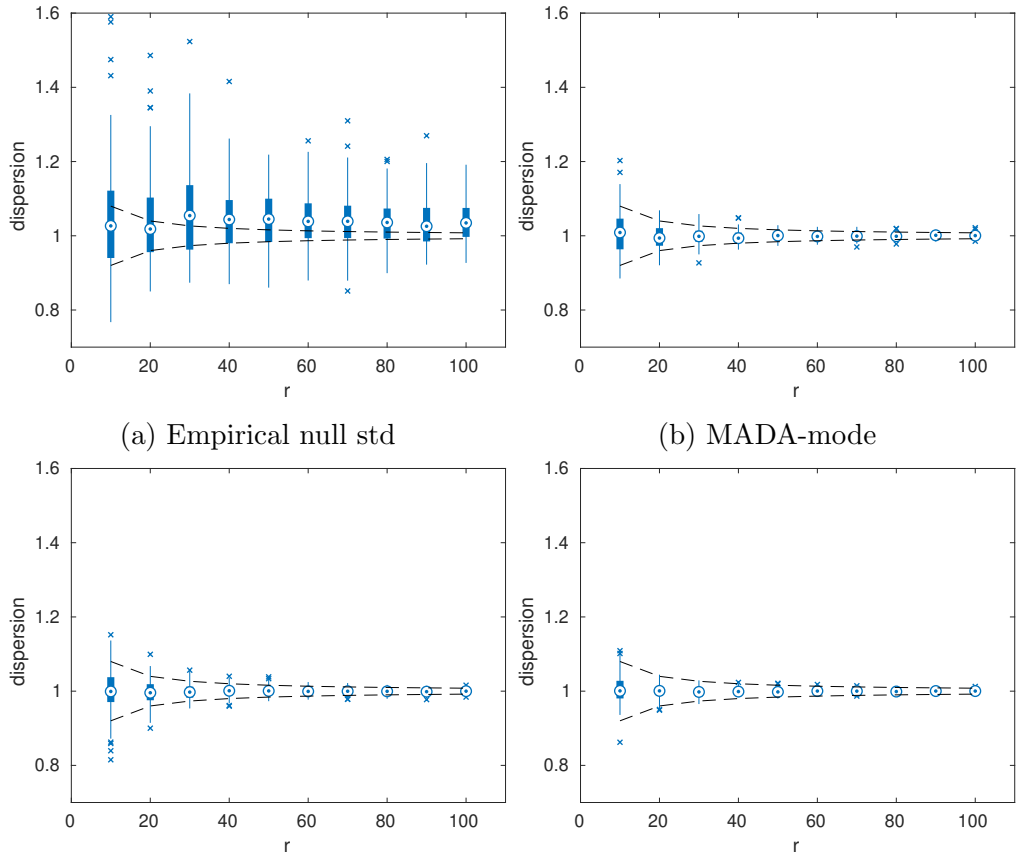
(a) Mode

(b) Bisquare weighted mean

(c) Median

(d) Mean

Figure 6.11: Different types of central tendency estimators were used on πr^2 standard Normal random variables. The boxplots represent the sampling distributions over 100 estimates. The dashed lines show the 95% confidence interval for the sample mean.



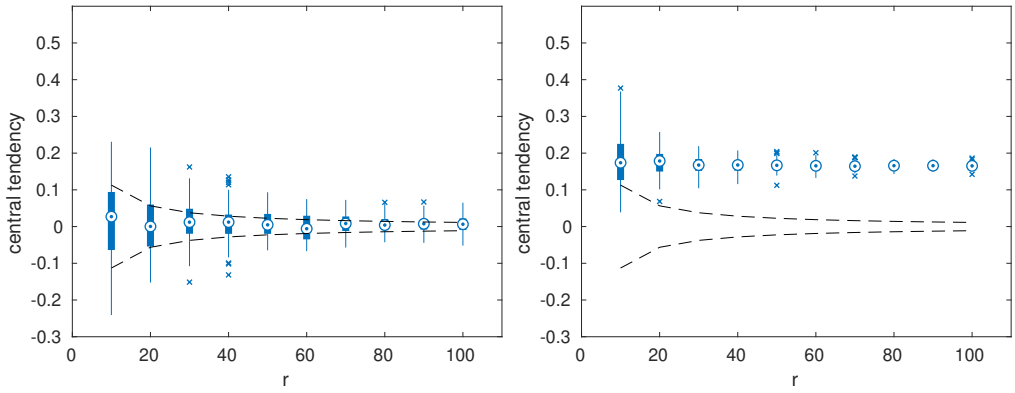
(a) Empirical null std

(b) MADA-mode

(c) IQR

(d) Standard deviation

Figure 6.12: Different types of statistical dispersion estimators were used on πr^2 standard Normal random variables. The boxplots represent the sampling distributions over 100 estimates. The dashed lines show the 95% confidence interval for the sample variance.



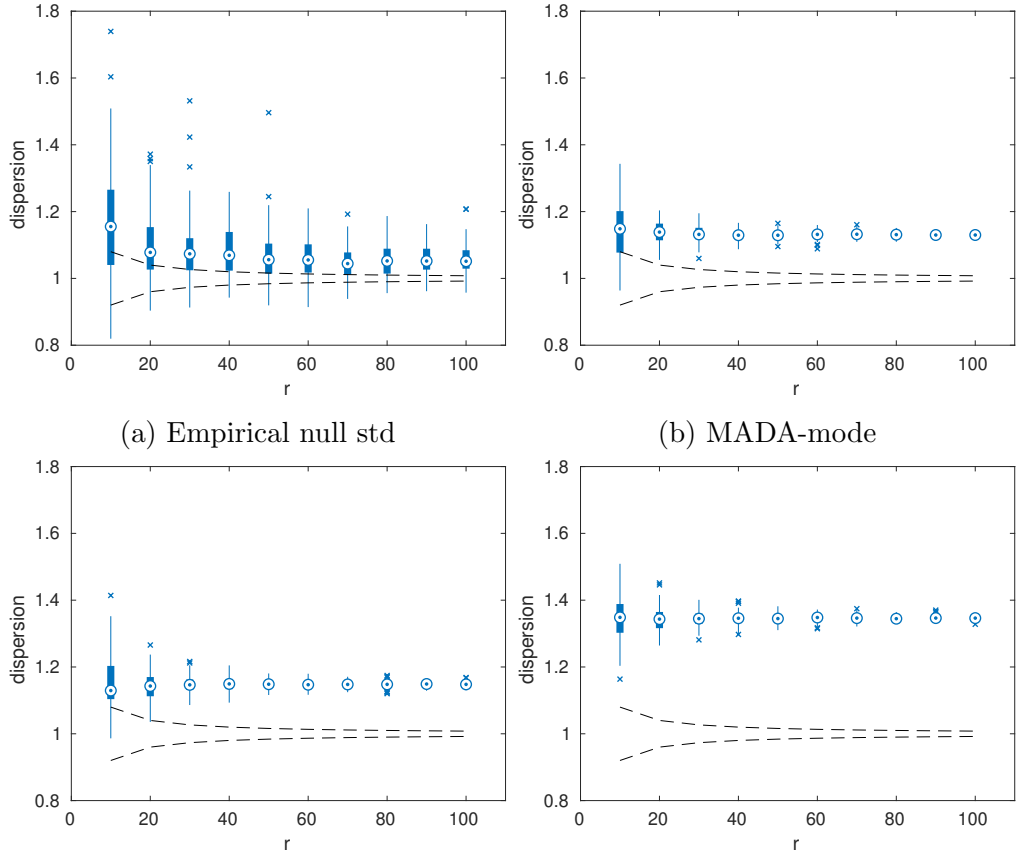
(a) Mode

(b) Bisquared weighted mean

(c) Median

(d) Mean

Figure 6.13: Different types of central tendency estimators were used on πr^2 mixture of Gaussian random variables. The random variable has distribution $\mathrm{N}(0,1)$ with probability 0.9 and $\mathrm{N}(3,1)$ with probability 0.1. The boxplots represent the sampling distributions over 100 estimates. The dashed lines show the 95% confidence interval for the sample mean.



(a) Empirical null std

(b) MADA-mode

(c) IQR

(d) Standard deviation

Figure 6.14: Different types of statistical dispersion estimators were used on πr^2 mixture of Gaussian random variables. The random variable has distribution $\mathrm{N}(0,1)$ with probability 0.9 and $\mathrm{N}(3,1)$ with probability 0.1. The boxplots represent the sampling distributions over 100 estimates. The dashed lines show the 95% confidence interval for the sample variance.

parameters, one possible extension to the empirical null is to spilt the z image into a grid and conduct inference in each section separately. There are problems with this though. The first problem is that how the grid is overlaid can be arbitrary, for example, the grid can be translated to produce different sections. Secondly, it is not clear how to combine the results from each section together (Efron, 2008).

The empirical null filter extends the empirical null to images. The parameters of the null distribution varies spatially, slowly and smoothly. Let the test statistic of the pixel at position (x, y) be

$$Z_{x,y} \sim \mathrm{N}(\mu_{0,x,y}, \widehat{\sigma}^2_{0,x,y}) \tag{6.33}$$

for $x = 1, 2, \ldots, W$ and $y = 1, 2, \ldots, H$. $\mu_{0,x,y}$ and $\sigma_{0,x,y}$ are the null mean and null standard deviation at position (x, y) respectively. Define the null hypotheses to be

$$H_{0,x,y} : \mu_{0,x,y} = \widehat{\mu}_{0,x,y} \tag{6.34}$$

which are tested against

$$H_{1,x,y} : \mu_{0,x,y} \neq \widehat{\mu}_{0,x,y} \ . \tag{6.35}$$

The empirical null filter aims to estimate the parameters $\mu_{0,x,y}$ and $\sigma_{0,x,y}$ for all x and y.

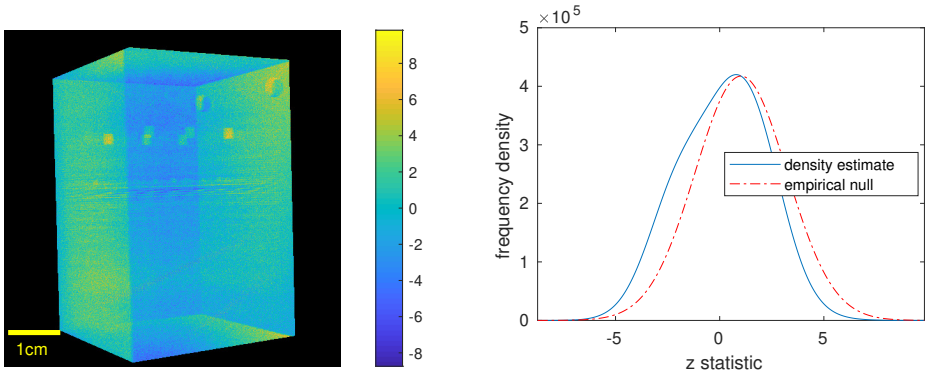
The empirical null filter uses the empirical null on neighbouring pixels to estimate the null parameters. To estimate $\mu_{0,x,y}$ and $\sigma_{0,x,y}$, a circular kernel $C_r(x, y)$ of radius r was centred at (x, y). All the pixels captured by the circular kernel and the ROI were used for the empirical null to obtain $\widehat{\mu}_{0,x,y}$ and $\widehat{\sigma}_{0,x,y}$, in other words

$$\widehat{\mu}_{0,x,y} = \operatorname{argmax} \widehat{p}_{Z_{x,y}}(z) \tag{6.36}$$

$$\widehat{\sigma}_{0,x,y} = \left[-\frac{\partial^2}{\partial z^2} \ln \widehat{p}_{Z_{x,y}}(z) \Big|_{z=\widehat{\mu}_{0,x,y}} \right]^{-1/2} \tag{6.37}$$

where

$$\widehat{p}_{Z_{x,y}}(z) = \frac{1}{nh} \sum_{i,j \in K_{x,y}} \phi\left(\frac{z_{i,j} - z}{h} \right) \ , \tag{6.38}$$



(a) Test statistics (b) Distribution of test statistics

Figure 6.15: The resulting test statistics for the `AbsFilter` projection at 120° are shown in a). The density estimate of the statistics is shown in b). The empirical null distribution multiplied by some constant is also plotted in b); this is to illustrate the curvature is the same as the density estimate at the mode.

$K_{x,y} = C_r(x,y) \cap \text{ROI}$ and $n = |K_{x,y}|$, not to be confused with N. A bandwidth of $h = (0.9n^{-1/5} + 0.16) \times \min\left(s_{x,y}, \text{IQR}_{x,y}/1.34\right)$ was used where $s_{x,y}$ and $\text{IQR}_{x,y}$ are the sample standard deviation and sample interquartile range of the test statistics in $K_{x,y}$.

After obtaining the estimates of the empirical null parameters, the test statistics were normalised using

$$T_{x,y} = \frac{Z_{x,y} - \widehat{\mu}_{0,x,y}}{\widehat{\sigma}_{0,x,y}} \tag{6.39}$$

and only the normalised test statistics were used in hypotheses testing. It should be noted that because the kernels can overlap, correlation between $T_{x,y}$ for different (x,y) is introduced.

The parameter of choice is the kernel radius r. A radius too small will treat defects as the null and will not detect them, too big can also cause problems as well. The empirical null filter assumes that the null parameters vary smoothly and slowly so that the null test statistics captured by the kernel is somewhat Normal. If the radius is too big, the kernel will capture distanced test statistics with very different null distributions. If all of these

test statistics are pulled together into a histogram, the null distribution may not look Normal at all which can cause problems when fitting a Normal null distribution. Furthermore, if $r \to \infty$, then this is just using the empirical null on all pixels.

The radius controls the proportion of pixels which are null captured by the kernel, denoted as π_0. It should be selected so that π_0 is of sensible value, for example, $> 90\%$. In practice, the defects are unknown so various kernel radiuses should be tried out. The kernel radius should be much larger than the size of the defect.

There exist similar filters, for example the local normalisation filter (Sage and Unser, 2003; Sage, 2018) uses a Gaussian filter to obtain $\widehat{\mu}_{0,x,y}$. $\widehat{\sigma}^2_{0,x,y}$ was obtained by applying another Gaussian filter on the $(z_{x,y} - \widehat{\mu}_{0,x,y})^2$ image. The purpose of this filter was to normalise images with uneven illumination. The disadvantage of using this filter, for estimating the null parameters, is that it uses a Gaussian filter. This is a type of weighted mean which is not robust to non-null statistics. Also, the two Gaussian filters can have different radiuses, it is not clear how to set them.

Design and Implementation

The empirical null filter was implemented using open-source software called *ImageJ* (Abràmoff et al., 2004; Schneider et al., 2012; Mateos-Pérez and Pascau, 2013) by modifying the existing class `RankFilters`, available on *GitHub* (ImageJ, 2018). This class implemented filters, such as the mean filter and median filter, using a circular kernel, making this a suitable framework for the empirical null filter. Figure 6.16 shows the graphical user interface of the empirical null filter in *Fiji* (Schindelin et al., 2012), a distribution of *ImageJ*.

As a test, the empirical null filter with $r = 5\,\text{px}$ was used on a *ImageJ* sample image `bridge.gif`, as shown in Figure 6.17. It is meaningless to use the empirical null filter on an arbitrary image, however, it did verify that the filter coped with it with some computational considerations. The empirical null images $\widehat{\mu}_{0,x,y}$ and $\widehat{\sigma}_{0,x,y}$ could be of interest, in particular, the empirical

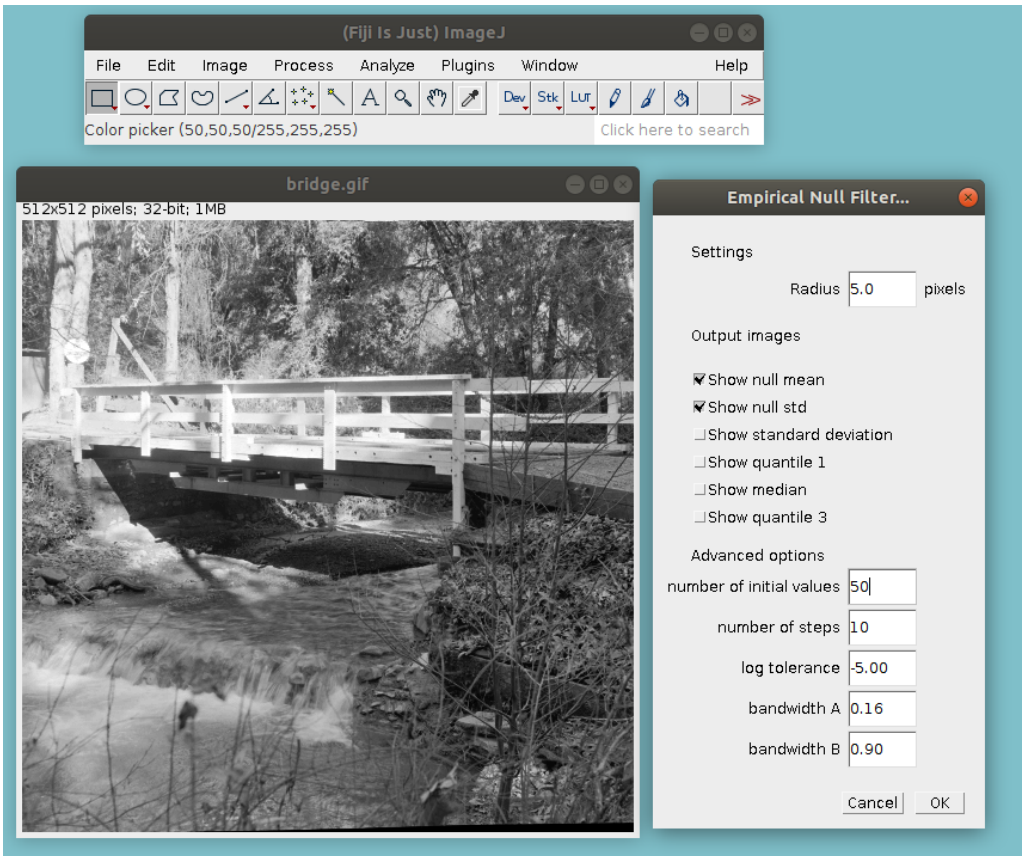


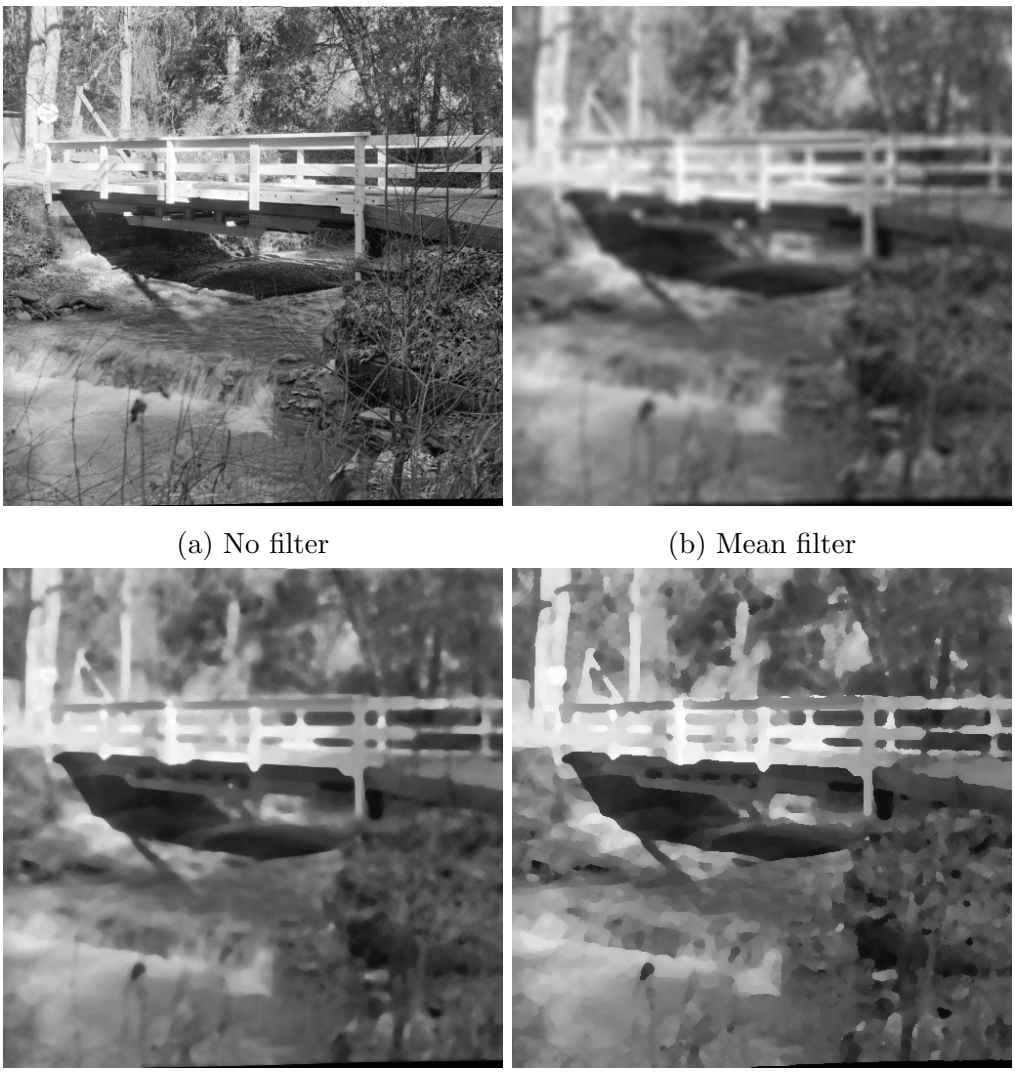
Figure 6.16: The graphical user interface of the empirical null filter in *Fiji*. The user can adjust the kernel radius as well as other advanced options relating to the Newton-Raphson method and the kernel density estimate. By-products such as the empirical null parameters can be shown after the filtering as well.

(a) Before filtering (b) After filtering

Figure 6.17: The empirical null filter, with kernel radius $r = 5\,\text{px}$, was used on the sample image `bridge.gif`.

null mean image could be interpreted as the result of a mode filter (Griffin, 2000). Figure 6.18 compares the empirical null mean with other averaging filters. The empirical null mean has an impasto effect and preserved edges which are similar to the result in Griffin (2000). Dispersion filters such as the standard deviation filter can be used to detect edges. Figure 6.19 compares the standard deviation filter with the empirical null standard deviation. The resulting images were similar but it was notable that the edges in the empirical null standard deviation were sharper, suggesting that the empirical null standard deviation is a more robust measure of dispersion compared to the standard deviation.

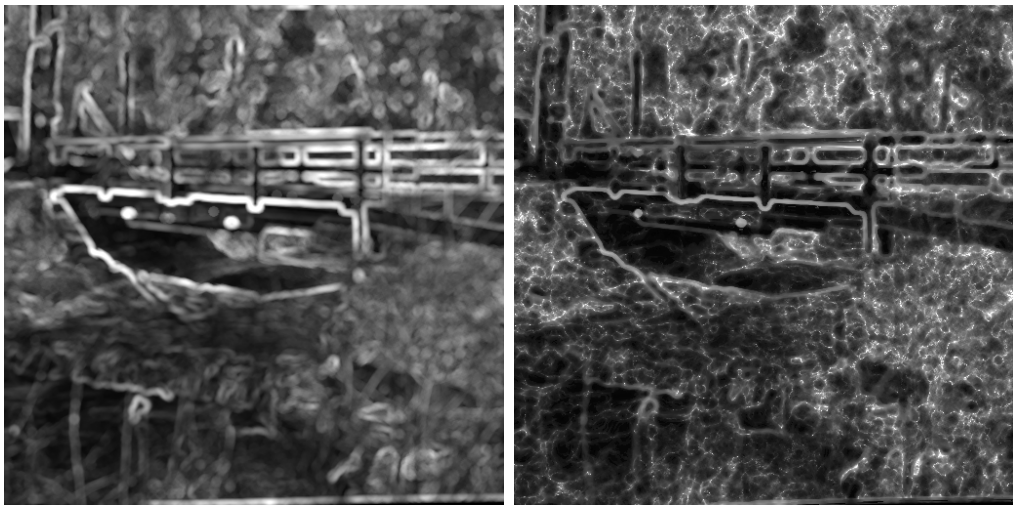
When using the empirical null filter in *ImageJ* or *Fiji*, the user is presented with a menu, as shown in Figure 6.16. The user can input the kernel radius, the filter then modifies the currently selected image $z_{x,y}$ to $t_{x,y}$ for all x and y using a circular kernel with the specified radius. The menu also has options for by-product images to be shown: the empirical null mean $\widehat{\mu}_{0,x,y}$, the empirical null standard deviation $\widehat{\sigma}_{0,x,y}$, the standard deviation filter, quartile filters. Advanced options are available to the user, for ex-



(a) No filter (b) Mean filter

(c) Median filter (d) Empirical null mean

Figure 6.18: Various averaging filters, with kernel radius $r = 5\,\text{px}$, were used on the sample image `bridge.gif`.



(a) Standard deviation filter (b) Empirical null standard deviation

Figure 6.19: Various dispersion filters, with kernel radius $r = 5\,\text{px}$, were used on the sample image `bridge.gif`.

ample, the number of initial values set the number of valid solutions to be found when using the Newton-Raphson method to find the mode of the density estimate. The user can also set the maximum number of steps and tolerance for the Newton-Raphson method. The bandwidth parameters for the kernel density estimate can also be adjusted here.

Line filtering was done from left to right. At the start on the far left, an initial value of the median over the pixels in the kernel was used. Also, various initial values were randomly tried out until a requested number of valid solutions were found. For the following pixel to the right, the empirical null mean of the neighbouring left pixel was used as the initial value. This was chosen as it was assumed the empirical null mean would vary slowly and smoothly spatially. The implementation uses multiple threads, each thread filters a row in parallel.

When the kernel captures pixels outside the boundary of the image or ROI, `RankFilters` uses nearest pixel padding which fills in pixels outside the ROI with values to the nearest pixels. Figure 6.20a shows an example of nearest pixel padding on the top left corner of a rectangle ROI. This is

$$\begin{array}{cccccc}
\ddots & \vdots & \vdots & \vdots & \vdots & \iddots \\
\cdots & z_{1,1} & z_{1,1} & z_{1,1} & z_{1,2} & \cdots \\
\cdots & z_{1,1} & z_{1,1} & z_{1,1} & z_{1,2} & \cdots \\
\cdots & z_{1,1} & z_{1,1} & z_{1,1} & z_{1,2} & \cdots \\
\cdots & z_{2,1} & z_{2,1} & z_{2,1} & z_{2,2} & \cdots \\
\iddots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

(a) Nearest pixel padding

$$\begin{array}{cccccc}
\ddots & \vdots & \vdots & \vdots & \vdots & \iddots \\
\cdots & \texttt{NaN} & \texttt{NaN} & \texttt{NaN} & \texttt{NaN} & \cdots \\
\cdots & \texttt{NaN} & \texttt{NaN} & \texttt{NaN} & \texttt{NaN} & \cdots \\
\cdots & \texttt{NaN} & \texttt{NaN} & z_{1,1} & z_{1,2} & \cdots \\
\cdots & \texttt{NaN} & \texttt{NaN} & z_{2,1} & z_{2,2} & \cdots \\
\iddots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

(b) `NaN` padding

Figure 6.20: When a kernel contains pixels outside the ROI, as shown by the solid line, the pixels can either be extrapolated using the nearest pixel or completely ignored by filling in the missing pixels with `NaN`.

not suitable for the empirical null filter as this will cause bias in the density estimation. The empirical null filter uses `NaN` padding which fills pixels outside the ROI with `NaN`. This indicates that these pixels are outside the ROI and are ignored. The number of non-`NaN` captured by the kernel was kept track for calculations such as the bandwidth.

When filtering arbitrary images such as `bridge.gif`, a few computational considerations were needed. `bridge.gif` is an 8-bit image so values were represented as integers. This opened up opportunities for the interquartile range and the standard deviation to be zero. This caused problems as this would set the bandwidth to zero. When the standard deviation is zero, it was set to 0.289 which corresponds to the standard deviation of a uniform random variable. When the inter-quartile range is zero, the inter-quartile range was set to the standard deviation $\times 1.34$.

When the number of initial values is too few, poor solutions to $\widehat{\mu}_0$ are possible which then can propagate to the pixel to the right and used as its initial value. Poor initial values could cause the Newton-Raphson method to never converge. If the Newton-Raphson method failed to converge too many times, the median, over pixels in the kernel, is used as the new initial value. Should it fail too, the program gives up and puts a `NaN` in that pixel.

Filtering an Image with No Defects

An experiment was conducted to demonstrate the effects of the empirical null filter on a simulated 256 px×256 px Gaussian image, where all the pixels have value distributed as standard Normal. The purpose is to simulate test statistics which are null and few positive results, subject to the FDR, should be detected. The test statistics after filtering should preserve its original distribution, or at least close to it. The statistical moments before and after filtering should be the same and were investigated. Because the empirical null mean and empirical null standard deviation are random variables, the filtered pixels will never be Normal and any normality tests would be too strict for this experiment.

Contamination is defined as some smooth and slowly varying function added and/or multiplied to an image. Conducting hypotheses testing on a contaminated Gaussian image may cause errors if the null distributions were incorrectly specified as standard Normal. In this experiment, the contamination was a gradient such that the pixel at position (x, y) has a value distributed as $Z_{x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma^2_{0,x,y})$ where

$$\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0) \ , \tag{6.40}$$

(x_0, y_0) is the centre of the image and $\sigma_{0,x,y} = 2$ for all (x, y).

A caveat is that some bias to the null standard deviation estimation would be introduced because of the contamination. This is because sources of variance captured by the kernel are from $\sigma_{0,x,y}$ and also from the variability of $\mu_{0,x,y}$. This can be shown with an example. Let $Z_{x,y}$ be the value of the pixels captured by the circular kernel, centred at the origin, for integer values of (x, y) such that $x^2 + y^2 \leqslant r^2$. Suppose that $Z_{x,y} \sim \mathrm{N}(ax + by, \sigma^2)$. The quantities of interest are the expectation and variance of all the values contained in the kernel because this is the quantity the estimators are estimating.

The calculations of the expectation and variance of $Z_{x,y}$ can be approximated by treating x and y as uniformly distributed within a circle centred

at the origin with radius r. The expectation is

$$\mathbb{E}[Z_{X,Y}] = \mathbb{E}\,\mathbb{E}[Z_{X,Y}|X,Y] = 0 \ . \tag{6.41}$$

The variance is given as $\mathbb{V}\mathrm{ar}[Z_{X,Y}] = \mathbb{E}\,\mathbb{V}\mathrm{ar}[Z_{X,Y}|X,Y] + \mathbb{V}\mathrm{ar}\,\mathbb{E}[Z_{X,Y}|X,Y]$. From the distribution of $Z_{X,Y}$, $\mathbb{E}\,\mathbb{V}\mathrm{ar}[Z_{X,Y}|X,Y] = \sigma^2$ and $\mathbb{V}\mathrm{ar}\,\mathbb{E}[Z_{X,Y}|X,Y] = \mathbb{V}\mathrm{ar}[aX+bY]$. This is worked out to be

$$\begin{aligned}\mathbb{V}\mathrm{ar}[aX+bY] &= \frac{1}{\pi r^2}\int_{\rho=0}^{\rho=r}\int_{\theta=0}^{\theta=2\pi}\rho^3(a\cos\theta+b\sin\theta)^2\mathrm{d}\theta\mathrm{d}\rho \\ &= \frac{1}{4}(a^2+b^2)r^2 \end{aligned} \tag{6.42}$$

thus

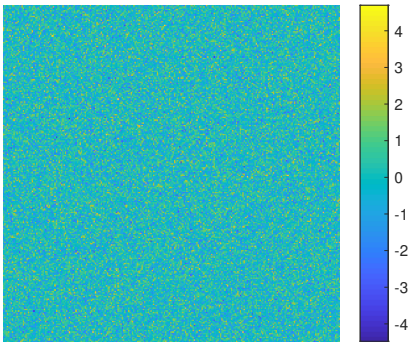
$$\mathbb{V}\mathrm{ar}[Z_{X,Y}] = \frac{1}{4}(a^2+b^2)r^2+\sigma^2 \ . \tag{6.43}$$

This shows that when estimating the null variance on a contaminated Gaussian image, there may be a $(a^2+b^2)r^2/4$ bias. If the contamination is slowly varying, that is a and b are small, then the bias should be small.

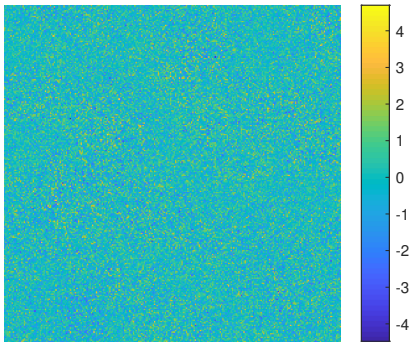
An example of a Gaussian image with/without contamination before/after filtering are shown in Figures 6.21 and 6.22. In the contaminated example, the filter managed to estimate the null mean, picking up the gradient. This enabled the filtered image to look flat and removed the contamination. The p-values after filtering are shown in Figure 6.23. A quick inspection suggests that the filtered pixels appeared reasonably Normal for this particular example.

For a given kernel radius r, 100 different Gaussian images were filtered to investigate the within image mean, standard deviation and kurtosis of the filtered statistics. Various other filters for normalisation were used as well such as the MADA-mode null filter, median IQR null filter and the mean variance null filter. They use different estimators to estimate the null parameters as their name suggests, Table 6.4 clarifies them.

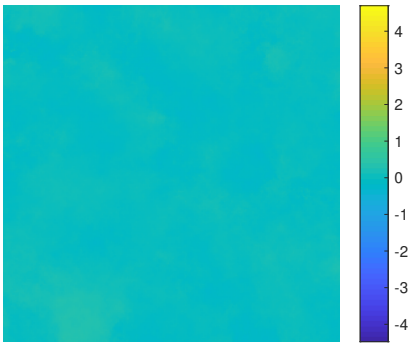
The results for the filtered images are shown in Figures 6.24, 6.25 and 6.26. For the filtered contaminated images, they are in Figures 6.27, 6.28 and 6.29. The filtered test statistic means did agree with zero as expected, showing that the filters centred the statistics when normalising. When using



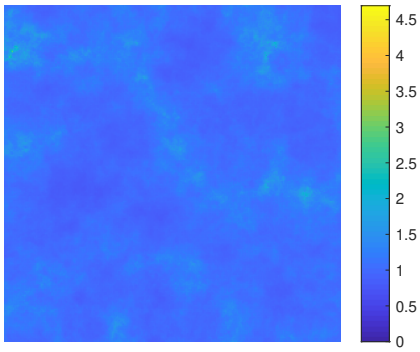
(a) Unfiltered



(b) Filtered



(c) Empirical null mean

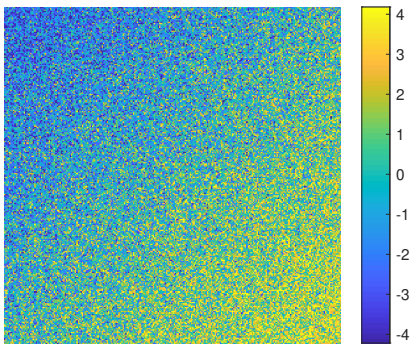


(d) Empirical null std

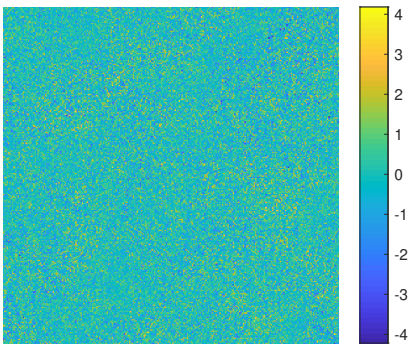
Figure 6.21: A 256 px×256 px Gaussian image before and after filtering with kernel radius 20 px. Also shown are the empirical null mean and empirical null standard deviation.

Filters	Null mean	Null standard deviation
MADA-mode null	Mode	MADA-mode ×1.483
Median IQR null	Median	IQR ÷1.349
Mean variance null	Mean	Standard deviation

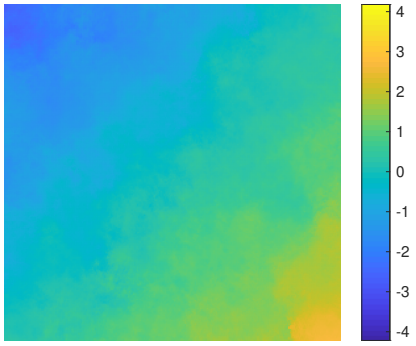
Table 6.4: Various filters using different estimators to estimate the null parameters are described here.



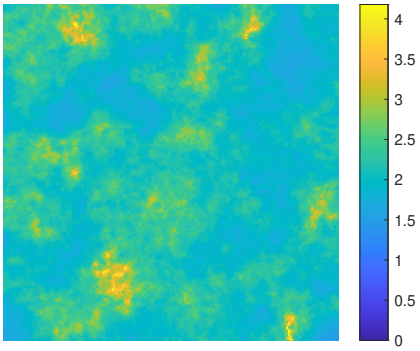
(a) Unfiltered



(b) Filtered

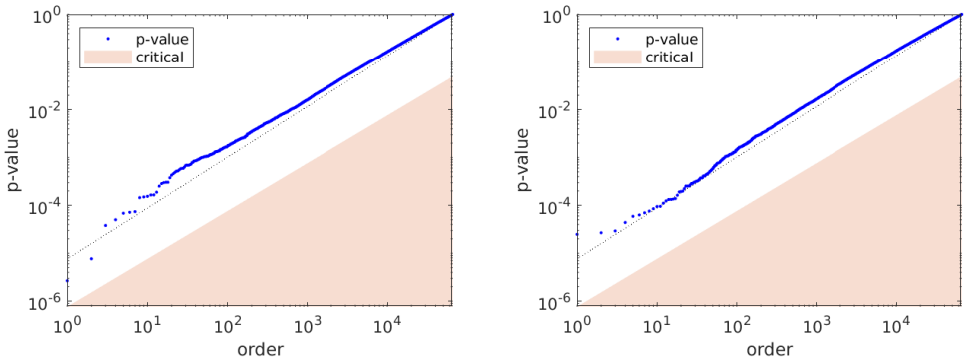


(c) Empirical null mean



(d) Empirical null std

Figure 6.22: A 256 px × 256 px contaminated Gaussian image before and after filtering with kernel radius 20 px. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. Also shown are the empirical null mean and empirical null standard deviation.



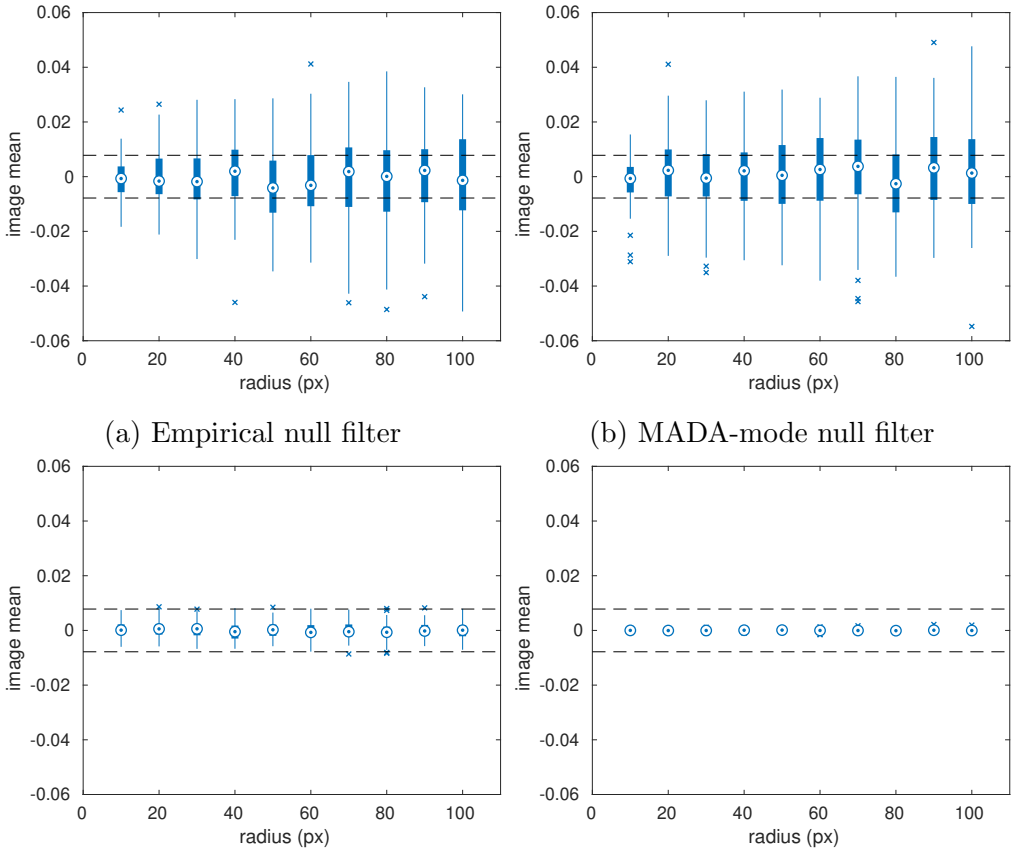
(a) No contamination (b) With contamination

Figure 6.23: p-values from a filtered 256 px × 256 px Gaussian image with kernel radius 20 px. In b), the Gaussian image was contaminated before filtering. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The p-values were obtained from the images in Figures 6.21b and 6.22b. The critical region corresponds to the 5% FDR level. The dotted lines shows the p-values if they were uniformly distributed.

the empirical null filter, there was some negative bias with the standard deviation of the filtered test statistics. The negative bias comes from a positive bias when estimating the null standard deviation.

When filtering the contaminated image, all estimators suffered from negative bias in the standard deviation of the filtered statistics, in particular when the kernel radius increased. This is shown in Figure 6.28. As discussed before, the kernel captured the variation due to the contamination and this added bias to the estimation of the null standard deviation.

The kurtosis in Figures 6.26 and 6.29 showed that the empirical null filter caused the filtered statistics to have heavy tails, in particular, for small kernel radiuses. This could cause problems in hypotheses testing because heavy tails could be misinterpreted as a contribution from non-null statistics. The source of the kurtosis inflation is from the estimation of the null standard deviation. This can be seen by comparing the empirical null filter with the MADA-mode null filter because the only difference between



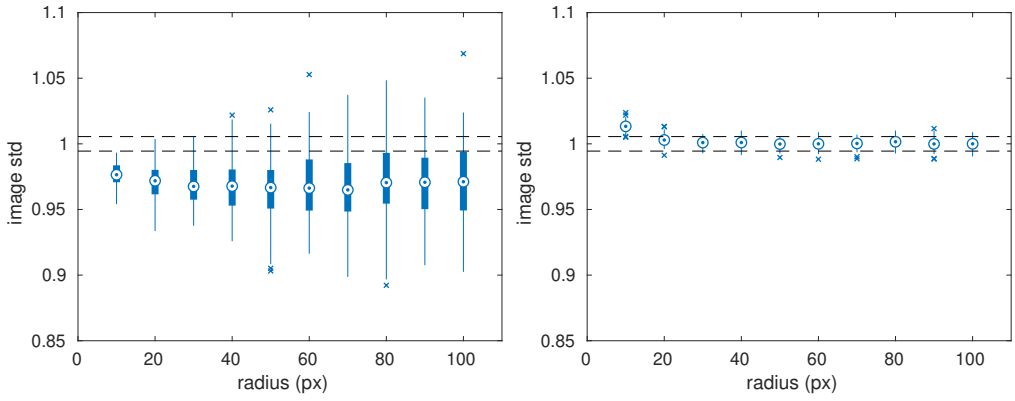
(a) Empirical null filter

(b) MADA-mode null filter

(c) Median IQR null filter

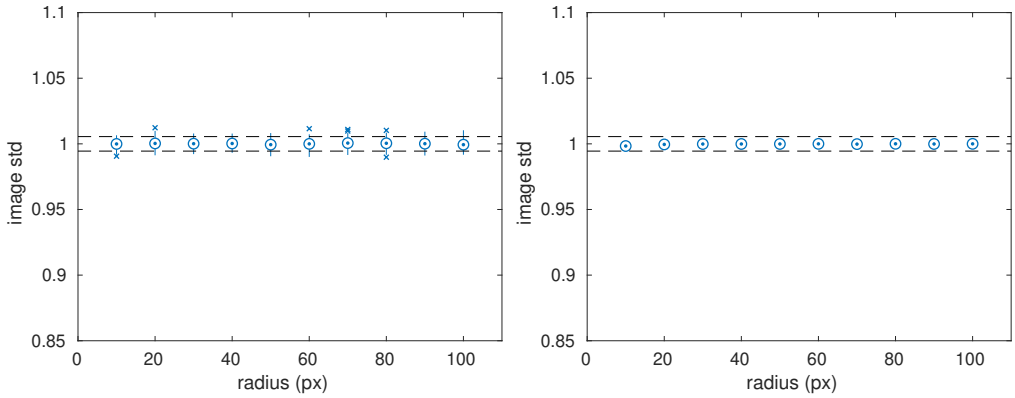
(d) Mean var null filter

Figure 6.24: The within image sample mean of a filtered Gaussian image of size 256 px × 256 px. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.



(a) Empirical null filter

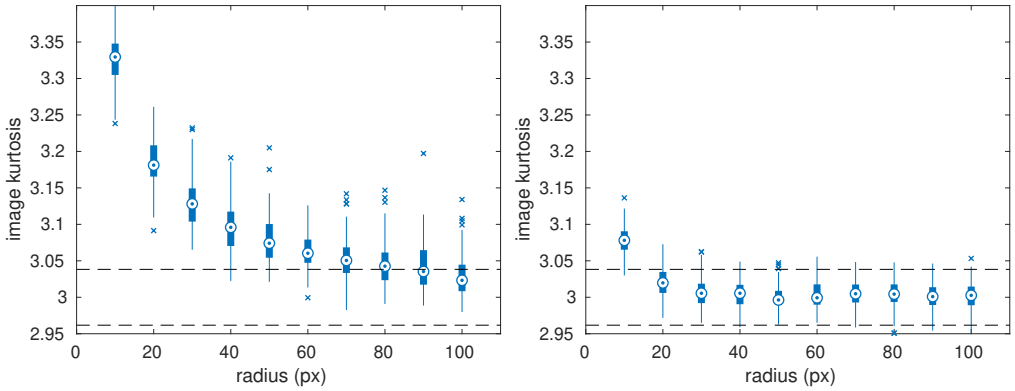
(b) MADA-mode null filter



(c) Median IQR null filter

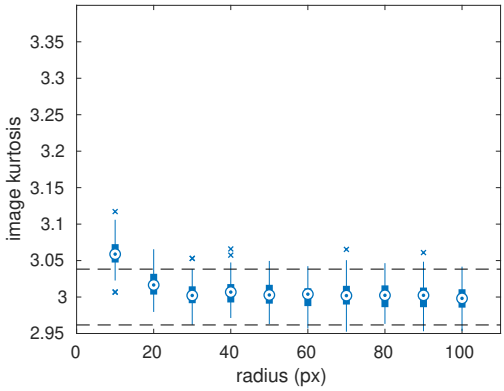
(d) Mean var null filter

Figure 6.25: The within image sample standard deviation of a filtered Gaussian image of size 256 px × 256 px. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.

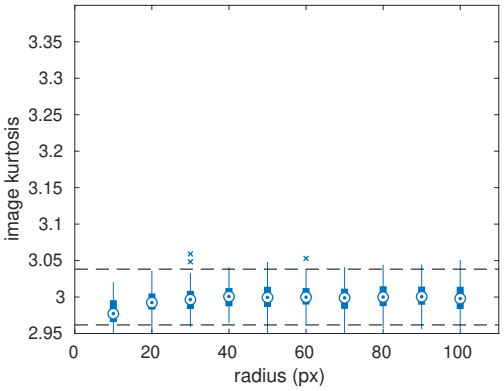


(a) Empirical null filter

(b) MADA-mode null filter

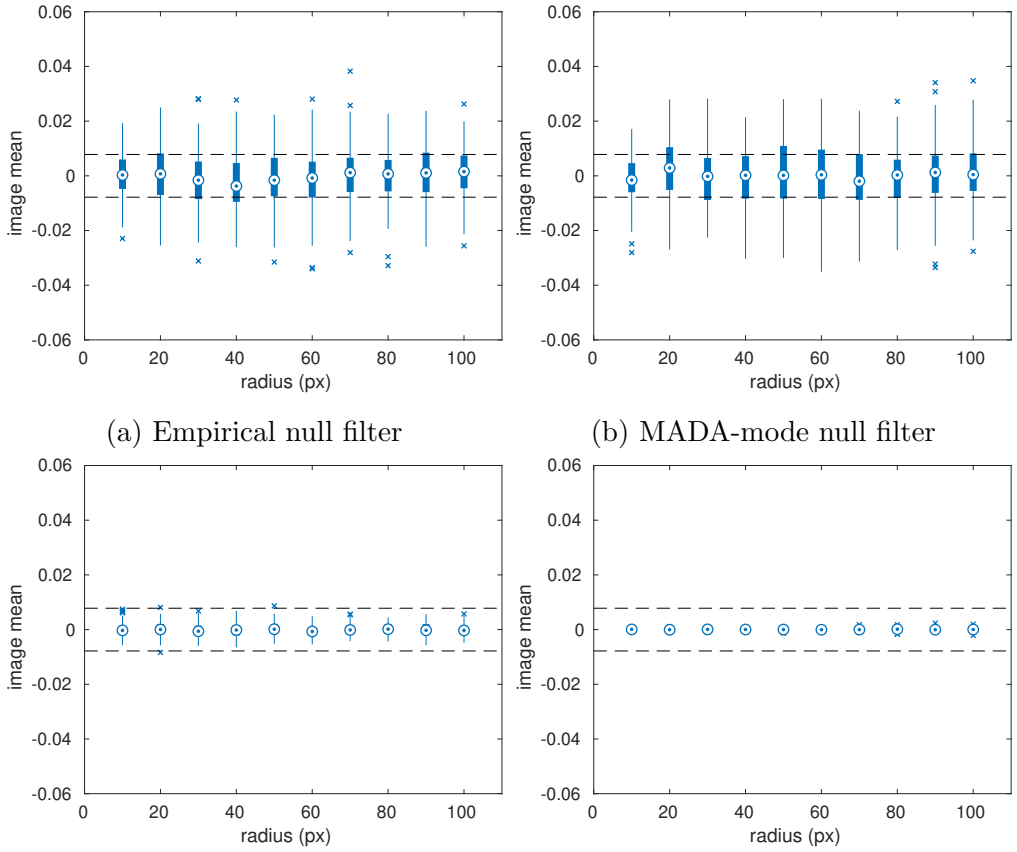


(c) Median IQR null filter



(d) Mean var null filter

Figure 6.26: The within image sample kurtosis of a filtered Gaussian image of size 256 px×256 px. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.



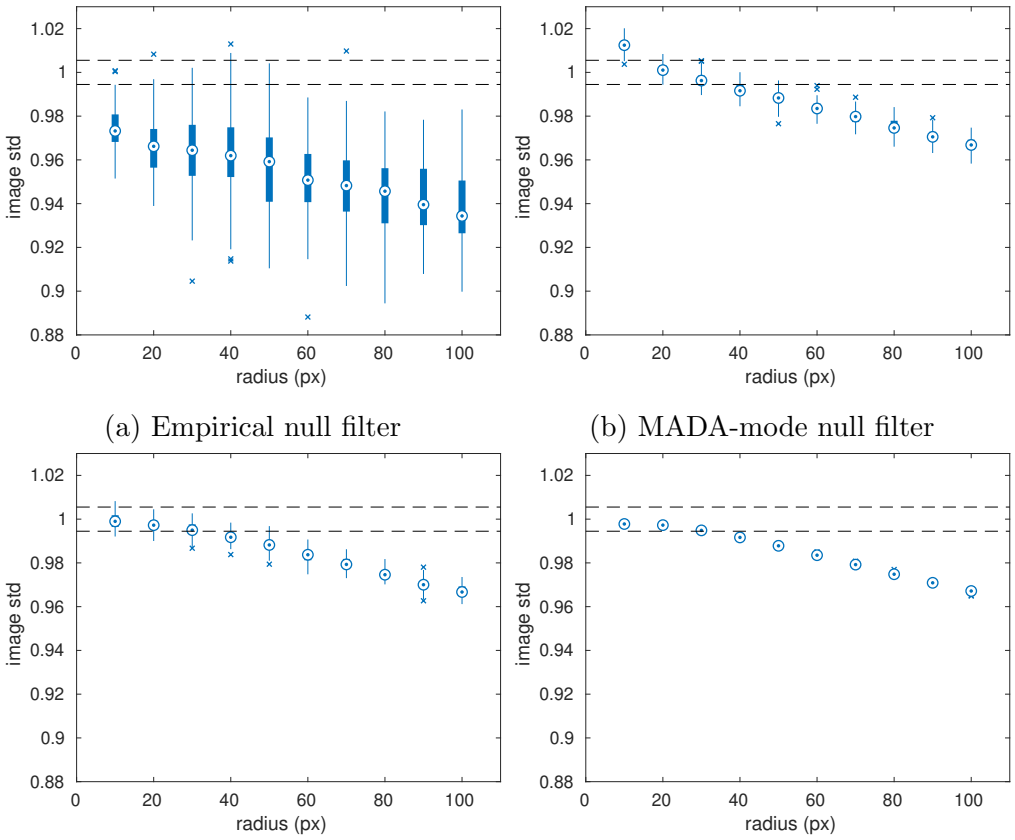
(a) Empirical null filter

(b) MADA-mode null filter

(c) Median IQR null filter

(d) Mean var null filter

Figure 6.27: The within image sample mean of a filtered contaminated Gaussian image of size 256 px × 256 px. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.



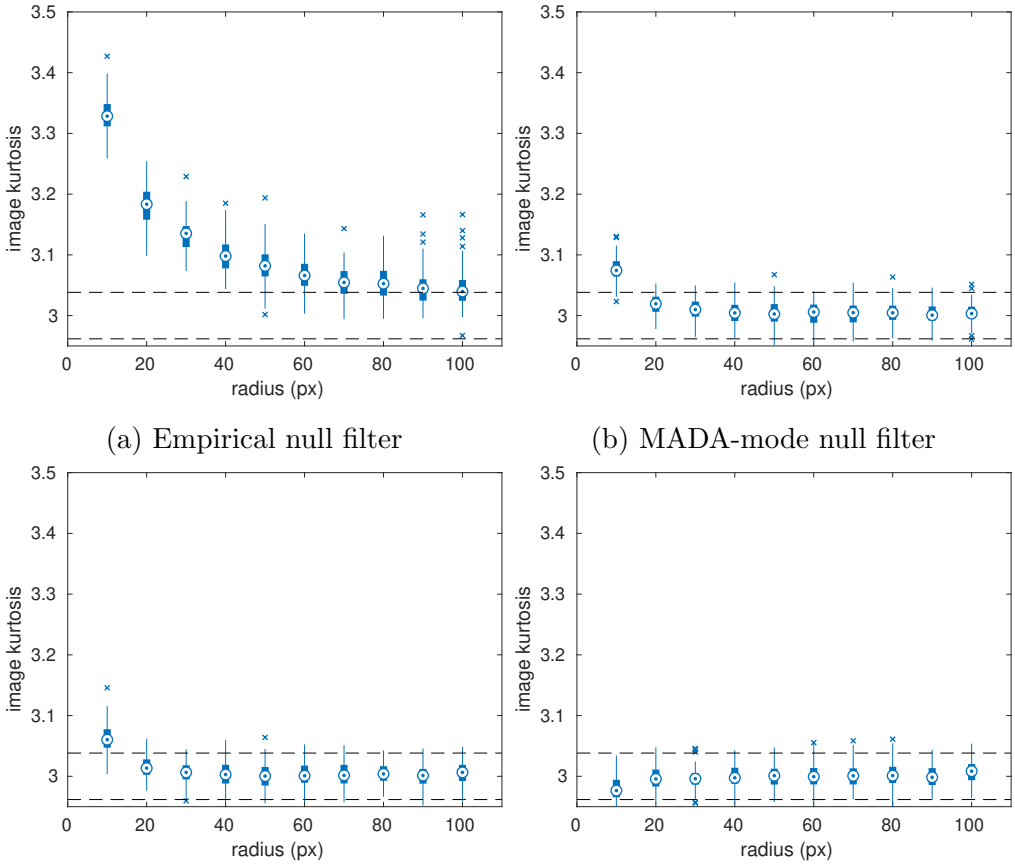
(a) Empirical null filter

(b) MADA-mode null filter

(c) Median IQR null filter

(d) Mean var null filter

Figure 6.28: The within image sample standard deviation of a filtered contaminated Gaussian image of size 256 px × 256 px. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.



(a) Empirical null filter

(b) MADA-mode null filter

(c) Median IQR null filter

(d) Mean var null filter

Figure 6.29: The within image sample kurtosis of a filtered contaminated Gaussian image of size 256 px × 256 px. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The boxplots summarise the 100 repeated simulations of the image. The dashed lines show the 95% confidence interval using standard tests and assuming independence.

the two filters is in the estimation of the null standard deviation.

Detection of Simulated Defects

The empirical null filter was tested to see if it can assist in detecting simulated defects from an image with/without contamination. A defect assigns pixels to have a value not distributed under the null distribution, but instead, under an alternative, or non-null, distribution. For example, suppose $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(0,1)$, then a defect assign certain pixels to have a value distributed as $Z_{x,y}|H_{1,x,y} \sim \mathrm{N}(\mu_1, 1)$ where $\mu_1 \neq 0$.

To recap, contamination is the result of a linear transform of the test statistics $Z_{x,y}$. For example, in this experiment, the image was multiplied by 2 and a gradient was added to it. The resulting null and alternative distributions are

$$Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, 2^2) \tag{6.44}$$

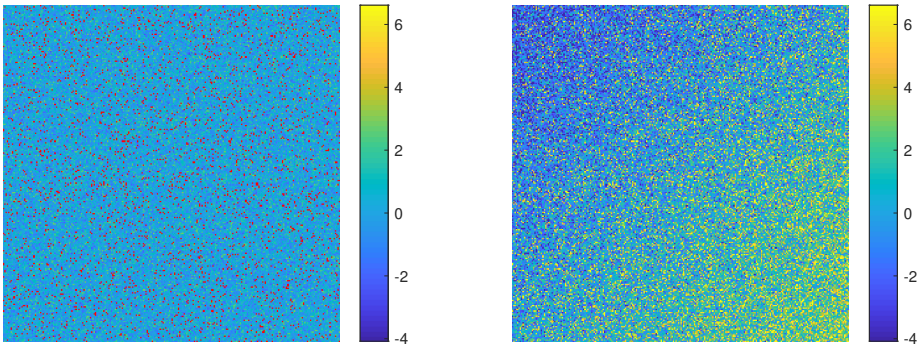
$$Z_{x,y}|H_{1,x,y} \sim \mathrm{N}(2\mu_1 + \mu_{0,x,y}, 2^2) \tag{6.45}$$

respectively where $\mu_{0,x,y} = 0.01(x-x_0)+0.01(y-y_0)$ and (x_0, y_0) is the centre of the image. This was used in this experiment to simulate a contaminated image with defects.

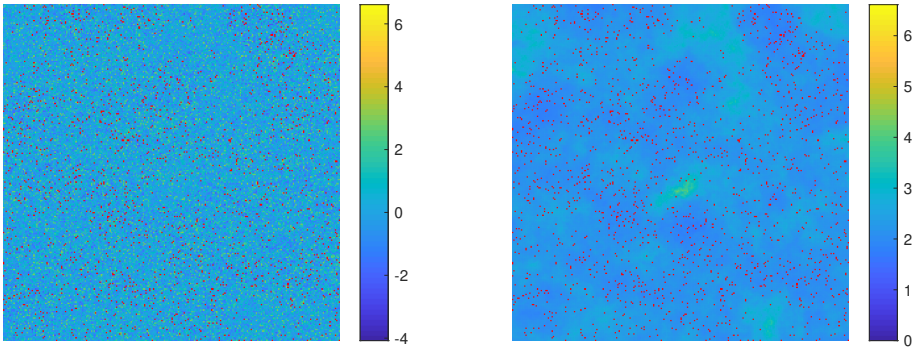
The empirical null filter aims to estimate the null distribution parameters from $Z_{x,y}$ to normalise it to form $T_{x,y}$. By normalising it, $T_{x,y}|H_{0,x,y}$ should be approximately standard Normal and hypotheses testing can be used to detect defects.

Various defects were investigated. Speckle defect with density π_1 assign all test statistics to be non-null with probability π_1 and are null otherwise. This was chosen because the proportion of null statistics captured by the kernel should be the same for all kernel radiuses. A line defect assigns columns of pixels to be non-null, a kernel would only capture a section of the defect. A square defect was also investigated and a kernel can capture the entire defect only if its radius is large enough.

An example of a speckle defected image is shown in Figure 6.30, the figure also shows the image contaminated and then filtered. Without filter-

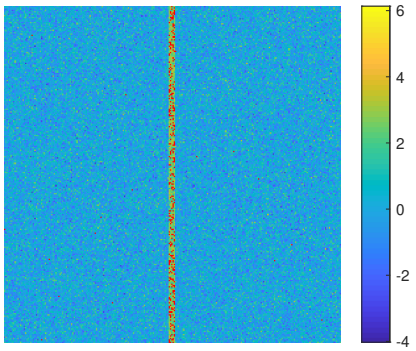


(a) Uncontaminated (b) Contaminated

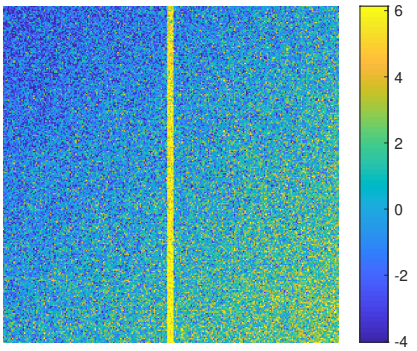


(c) Filtered (d) Empirical null std

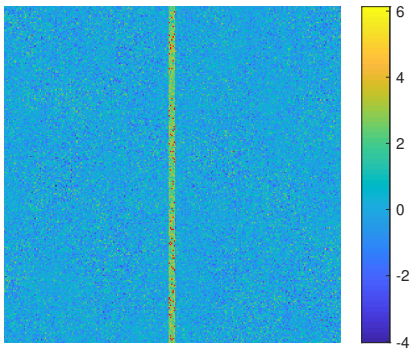
Figure 6.30: A 256 px × 256 px Gaussian image with speckle defect is shown in a). The image was contaminated in b) and then filtered with kernel radius 20 px in c). Non-null pixels have the distribution $\mathrm{N}(3, 1)$. The speckle defect has density $\pi_1 = 0.1$. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. In a), c) and d), highlighted in red are pixels tested as positive at the 5% FDR level.



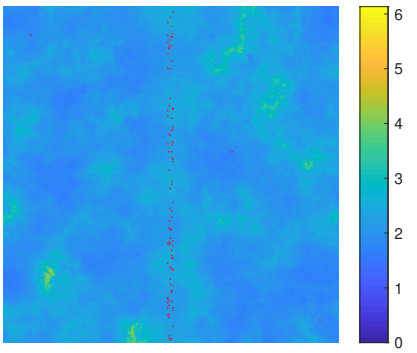
(a) Uncontaminated



(b) Contaminated

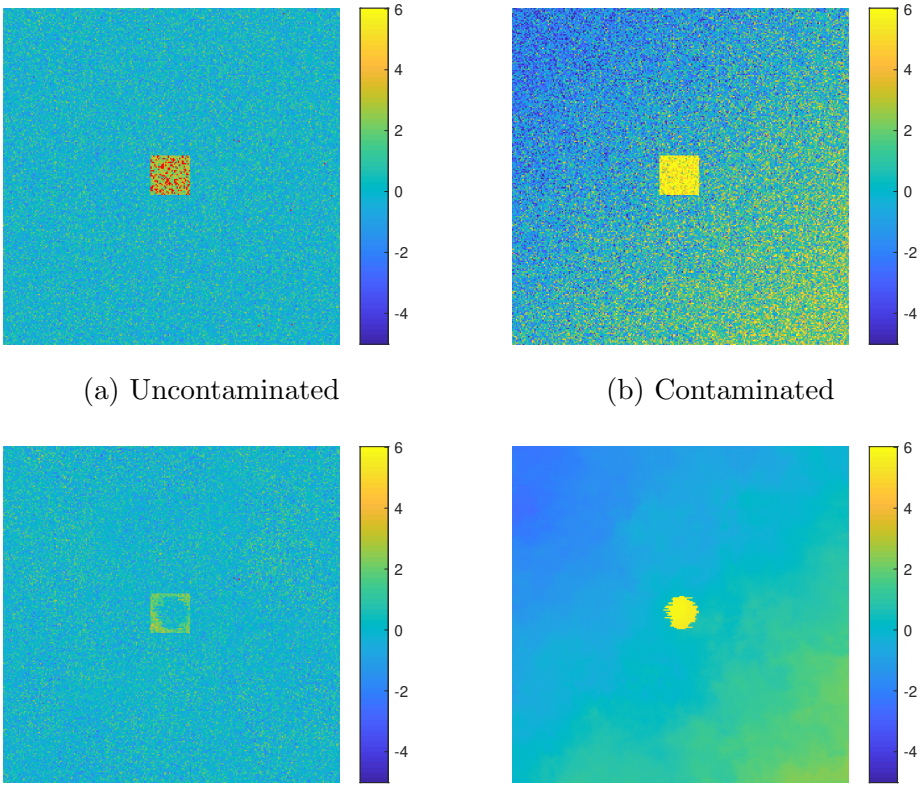


(c) Filtered



(d) Empirical null std

Figure 6.31: A 256 px × 256 px Gaussian image with a line defect is shown in a). The image was contaminated in b) and then filtered with kernel radius 20 px in c). Non-null pixels have the distribution $\mathrm{N}(3, 1)$. The line defect is 5 px thick. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. In a), c) and d), highlighted in red are pixels tested as positive at the 5% FDR level.



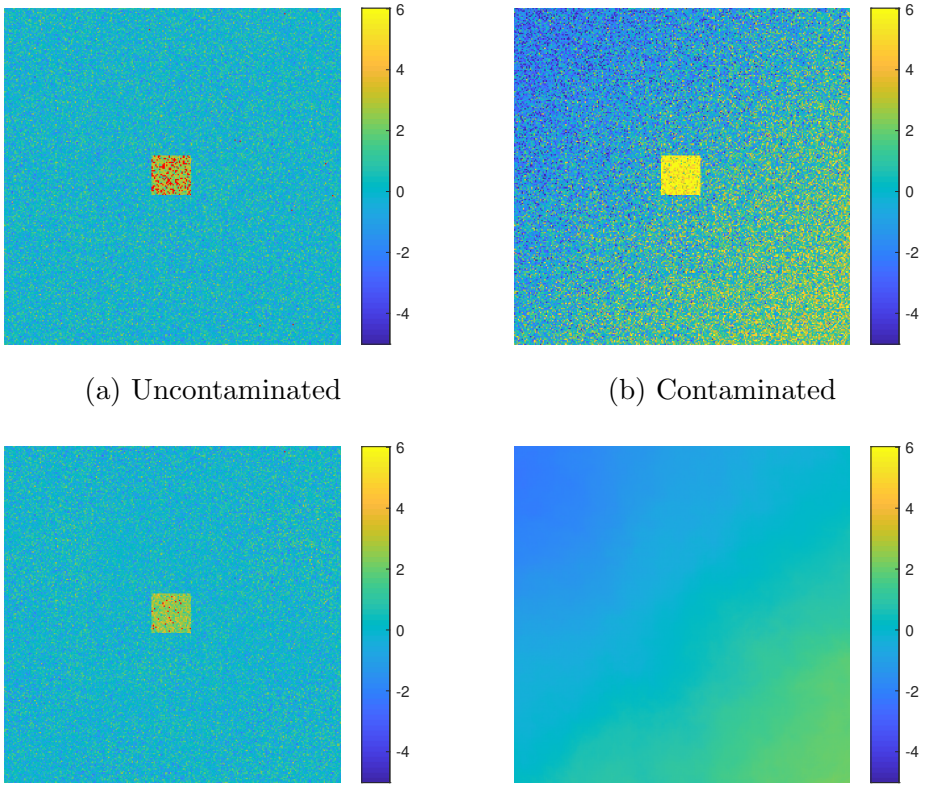
(a) Uncontaminated

(b) Contaminated

(c) Filtered

(d) Empirical null mean

Figure 6.32: A 256 px × 256 px Gaussian image with a square defect is shown in a). The image was contaminated in b) and then filtered with kernel radius 20 px in c). Non-null pixels have the distribution $\mathrm{N}(3, 1)$. The square defect is 30 px×30 px in size. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. In a) and c), highlighted in red are pixels tested as positive at the 5% FDR level.



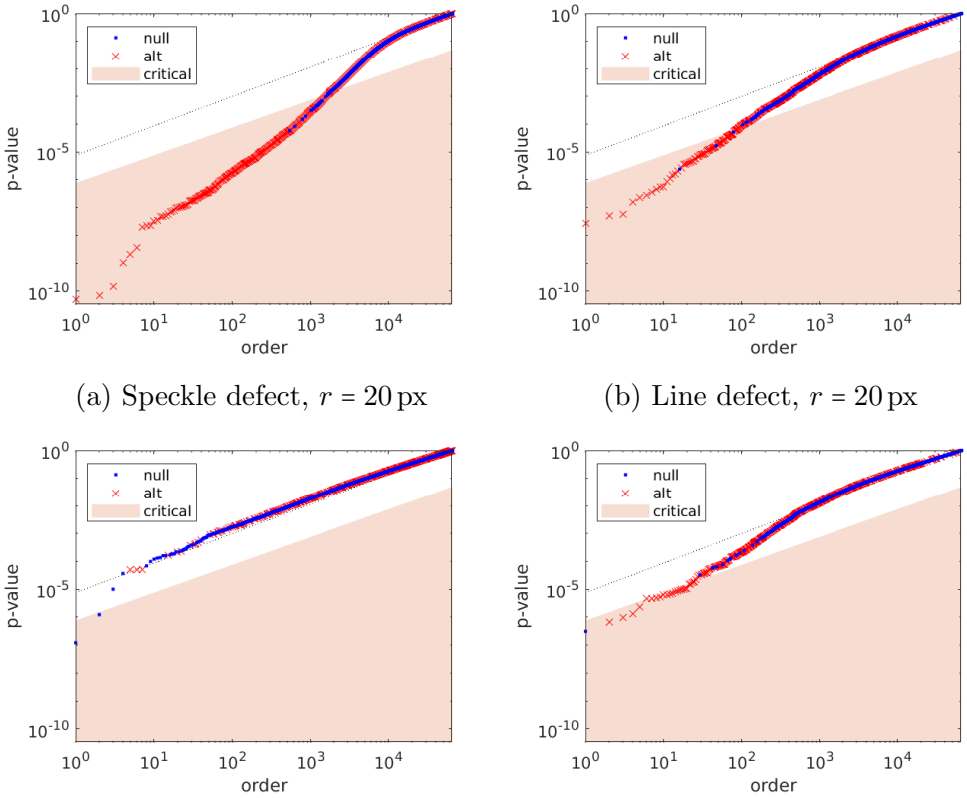
(a) Uncontaminated

(b) Contaminated

(c) Filtered

(d) Empirical null mean

Figure 6.33: A 256 px × 256 px Gaussian image with a square defect is shown in a). The image was contaminated in b) and then filtered with kernel radius 40 px in c). Non-null pixels have the distribution $\mathrm{N}(3, 1)$. The square defect is 30 px×30 px in size. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. In a) and c), highlighted in red are pixels tested as positive at the 5% FDR level.



(a) Speckle defect, $r = 20\,\text{px}$

(b) Line defect, $r = 20\,\text{px}$

(c) Square defect, $r = 20\,\text{px}$

(d) Square defect, $r = 40\,\text{px}$

Figure 6.34: A $256\,\text{px} \times 256\,\text{px}$ contamined defected Gaussian image was filtered using the empirical null filter with kernel radius r. Shown are the p-values converted from the filtered images. Non-null pixels have the distribution $\mathrm{N}(3, 1)$. The speckle defect has density $\pi_1 = 0.1$. The line defect is $5\,\text{px}$ thick. The square defect is $30\,\text{px} \times 30\,\text{px}$ in size. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The critical region corresponds to the 5% FDR level. The dotted lines shows the p-values if they were uniformly distributed.

ing, the top-left and bottom-right of the contaminated image have large, in magnitude, test statistics. In these areas, a lot of pixels would be tested as falsely positive. By using the empirical null filter, the resulting filtered image reassemble the uncontaminated image and appropriate inference can be done on the filtered image. However, some statistical power was lost when conducting hypotheses testing on the filtered image because, in areas of high empirical null standard deviation, the normalised statistics became too small which then decreased the detection power. This cannot be avoided as areas of high empirical null standard deviation was due to the randomness in sampling. Similar comments can be made for the line defect as shown in Figure 6.31.

The square defected images, including the contaminated ones, are shown in Figures 6.32 and 6.33. Figure 6.32 filtered the contaminated image using a kernel radius of $r = 20\,\text{px}$, while Figure 6.33 used a kernel radius of $r = 40\,\text{px}$. With $r = 20\,\text{px}$, the kernel is smaller than the square defect, therefore, the defect was treated as the null. This is evident because the empirical null mean captured the defect. This resulted in difficulty detecting the defect. With $r = 40\,\text{px}$, the kernel is bigger than the defect and treated the defect as non-null. In this scenario, the empirical null filter recovered the gradient in the empirical null which was then used for normalisation.

The empirical null mean in Figure 6.32 demonstrated the multi-thread nature of the implementation of the empirical null filter. It can be seen that there were horizontal streaks where the defect is. This is the result of each thread filtering a row and jumping from one mode, the null, to the other, the non-null, at different times. The horizontal streaks can be removed by using more initial values so that the Newton-Raphson method can pinpoint which mode is greater. However, this would be at a computational cost.

Figure 6.34 shows the p-values after filtering a contaminated defected image. The FDR can be estimated in these figures by dividing the number of null statistics in the critical region by the number of statistics in the critical region. With a sensible kernel radius, some of the defects can be detected. False negatives are common but this is inevitable to control for the FDR. The figure shows that using a kernel radius of $r = 20\,\text{px}$, for

the square defect, failed to capture the defect. This, again, emphasise the importance of a good kernel radius.

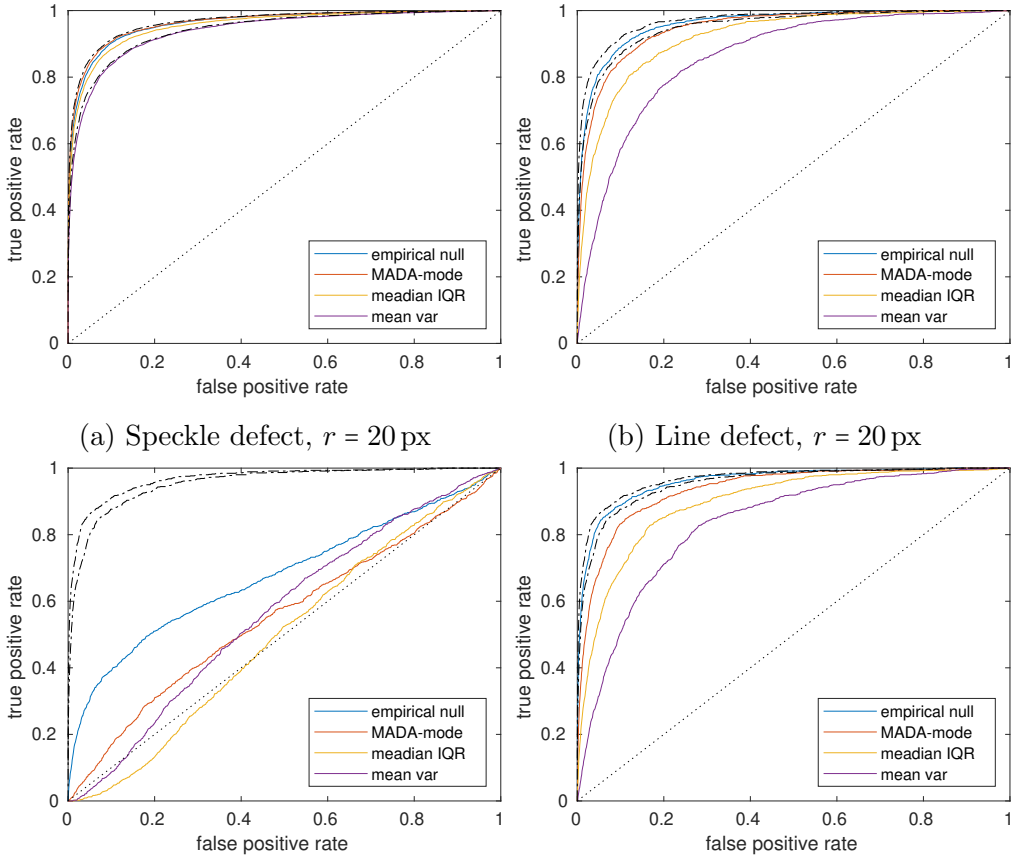
The receiver operating characteristic (ROC) curves (Green and Swets, 1966; Metz, 1978; Hanley and McNeil, 1982; Friedman et al., 2001; Cook, 2007) are shown in Figure 6.35 for the various defects and filters. The ROC curve is a parametric plot, plotting the true positive rate (sensitivity) against the false positive rate (1 – specificity) for varying thresholds. The area under the ROC curve (AUC) is a commonly used statistic to quantify the performance of the test (Friedman et al., 2001). Interpretations of the area do exist (Metz, 1978; Hanley and McNeil, 1982) and discussed thoroughly in Cook (2007).

In terms of the AUC, the empirical null filter improved the performance of hypotheses testing compared with using the unfiltered contaminated image. The performance before contamination cannot be recovered but it was an improvement. For kernel radiuses too small, such as Figure 6.35c, the empirical null filter deteriorated the performance and it would be better off using the contaminated image.

Speckle Defect Experiment

The ROC curves consider all thresholds or specificities used in the hypotheses testing. In the previous experiment, it was found that the variance of the test statistics was different before and after filtering. As a result, for a given threshold, such as the 5% FDR level, the specificity may change ever so slightly after filtering.

An experiment was conducted to investigate how filtering affects hypotheses testing. A $256\,\text{px} \times 256\,\text{px}$ Gaussian image with speckle defect, with density $\pi_1 = 0.1$, and various values of μ_1 were investigated. The AUC, type 1 error, type 2 error and FDR were measured when testing on an image with a defect before and after contamination and then after filtering with the contamination. The AUC was obtained by integrating the ROC curve. A kernel radius of $r = 20\,\text{px}$ was used and was repeated 100 times by simulating another image.



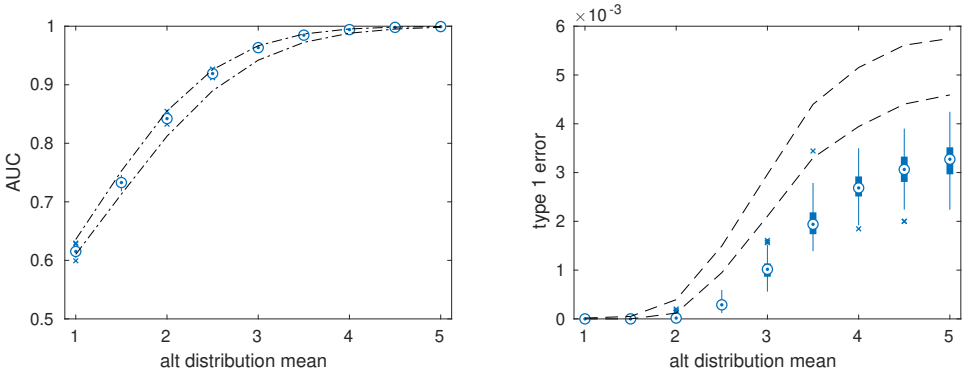
(a) Speckle defect, $r = 20\,\text{px}$

(b) Line defect, $r = 20\,\text{px}$

(c) Square defect, $r = 20\,\text{px}$

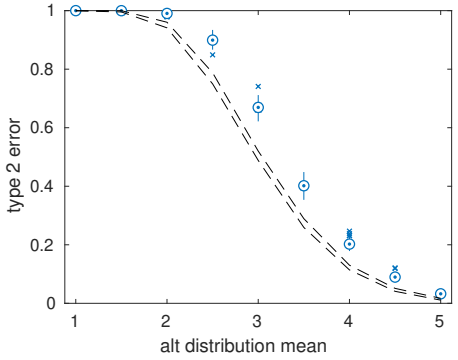
(d) Square defect, $r = 40\,\text{px}$

Figure 6.35: ROC curves for various defected $256\,\text{px} \times 256\,\text{px}$ Gaussian images. The upper/lower dot-dashed lines show the resulting ROC curve when testing on an image without/with contamination respectively. The different curves are the resulting ROC curves after filtering a contaminated image. The speckle defect has density $\pi_1 = 0.1$. The line defect is $5\,\text{px}$ thick. The square defect is $30\,\text{px} \times 30\,\text{px}$ in size. Defected pixels have the distribution $\mathrm{N}(3, 1)$. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$.

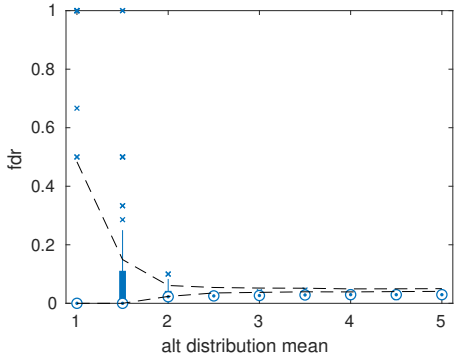


(a) AUC

(b) Type 1 error



(c) Type 2 error



(d) FDR

Figure 6.36: AUC and various errors obtained when conducting hypotheses testing on an empirical null filtered contaminated speckle defected Gaussian image of size 256 px × 256 px. In a), the upper/lower dot-dashed lines shows the resulting mean AUC when testing on a defected image before/after contamination respectively. In b), c) and d), the dashed lines show the 95% empirical confidence interval of the resulting error when testing the defected image before contamination. The filter used a kernel radius of 20 px. Defected pixels have the distribution $\mathrm{N}(\mu_1, 1)$ where μ_1 was varied in this experiment. The speckle defect has density $\pi_1 = 0.1$. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The test was done at the 5% FDR level. The boxplots summarise the 100 repeated simulations of the image.

The results are shown in Figure 6.36. The AUC quantified showed that using the filter improved the performance of the test from the contamination. The type 1 error, or specificity, did decrease after filtering. However, the results illustrated that the FDR is controlled quite well at around 5% after filtering. This meant that in this particular example, FDR control is consistent after filtering. At situations with low detection power, the FDR can fluctuate between 0 and 1 and can take only so many values. For example, if 3 positive pixels were detected, then the FDR can only take values of multiples of 1/3 between and including 0 and 1.

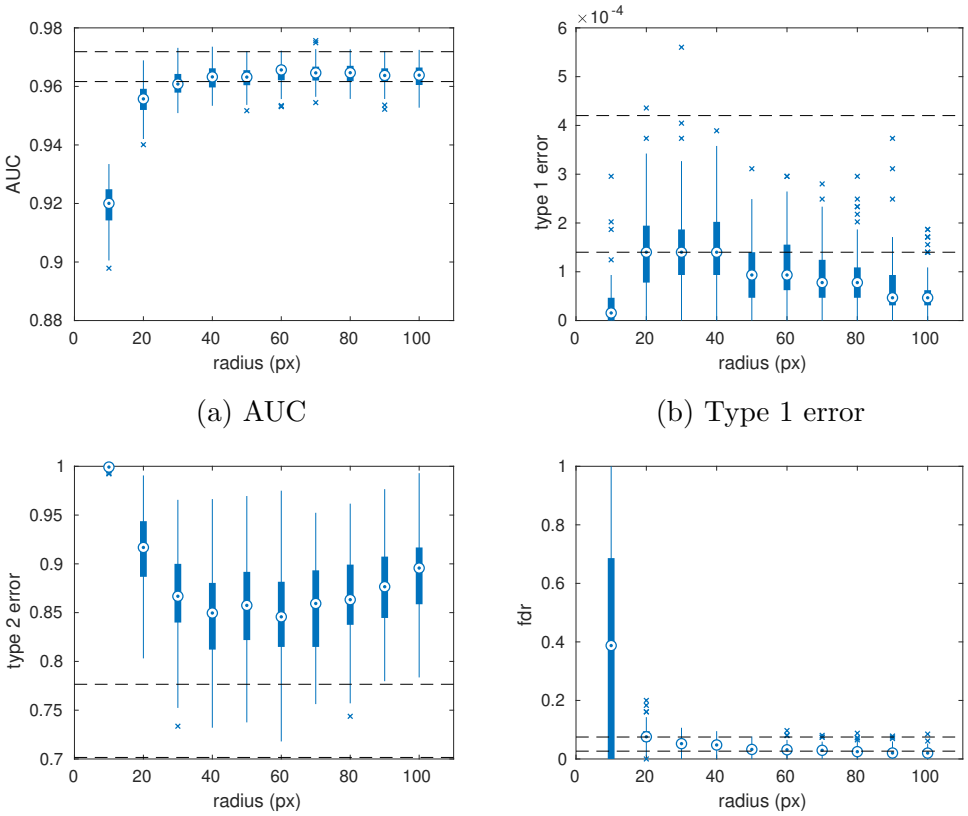
Optimal Kernel Radius Experiment

There is the question of what kernel radius to choose for a given defect. Literature, such as Efron (2004) and Schwartzman et al. (2008b), suggest that, as a rule of thumb, the proportion of non-null statistics should not be larger than $\pi_1 = 0.1$ to satisfy some assumptions made for the empirical null. Given the size of the defect, one could work out the minimum kernel radius by setting a threshold for the maximum proportion of the area of the kernel which contains a defect to 10%. In other words, for a kernel with radius r

$$0.1 > \frac{\text{maximum area of defect captured by the kernel}}{\text{area of kernel}} . \tag{6.46}$$

For a $d \times d$ square defect, this is $r > d\sqrt{10/\pi}$. In the example of $d = 30\,\text{px}$, this is about $r > 54\,\text{px}$. For the line defect with thickness $5\,\text{px}$, the radius is about $r > 32\,\text{px}$. These are rules of thumb because the kernel is not perfectly circular when used on a grid of pixels. It should be pointed out that the implementation can accept non-integer radiuses if desired.

An experiment was conducted where a contaminated defected Gaussian image, of size $256\,\text{px} \times 256\,\text{px}$, was filtered using various kernel radiuses for a fixed alternative distribution $\text{N}(3, 1)$. Hypotheses testing was done on the uncontaminated defected and filtered contaminated defected images. The AUC and various errors were recorded. This was repeated 100 times by simulating the image again. The AUC was obtained by integrating the



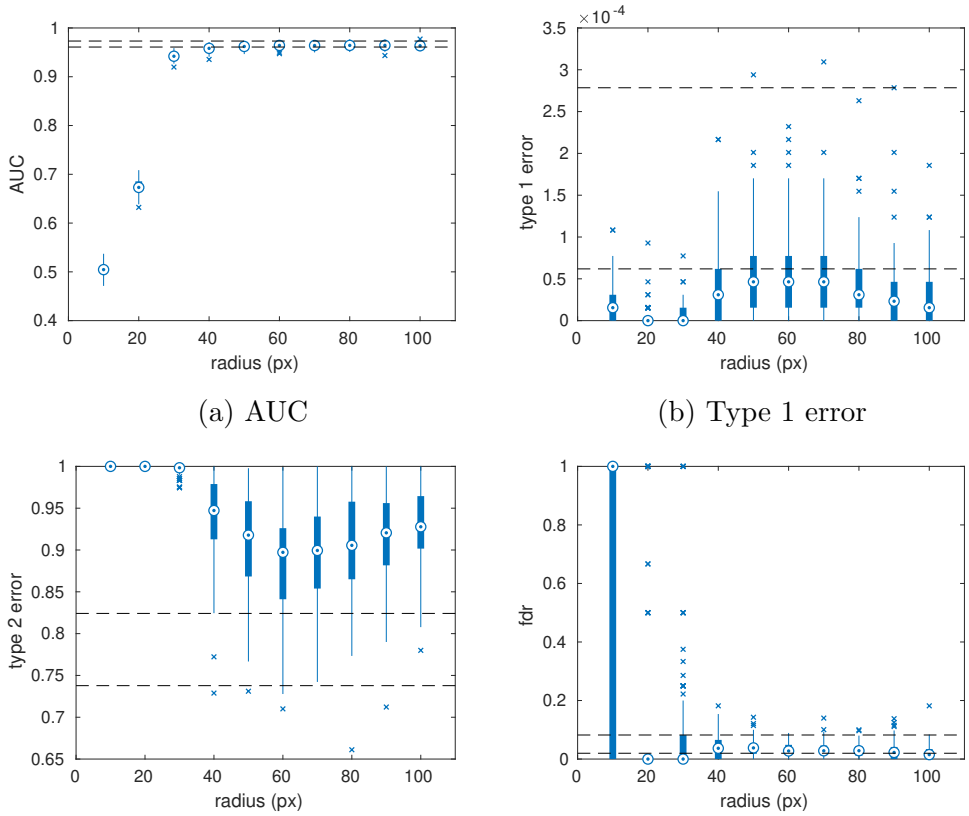
(a) AUC

(b) Type 1 error

(c) Type 2 error

(d) FDR

Figure 6.37: AUC and various errors obtained when conducting hypotheses testing on a filtered contaminated line defected Gaussian image, of size 256 px × 256 px, using various kernel radiuses. The dashed lines show the resulting 95% empirical confidence interval when the hypotheses testing was done on the uncontaminated defected image. Defected pixels have the distribution $\mathrm{N}(3,1)$. The line defect is 5 px thick. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The test was done at the 5% FDR level. The boxplots summarise the 100 repeated simulations of the image.



(a) AUC

(b) Type 1 error

(c) Type 2 error

(d) FDR

Figure 6.38: AUC and various errors obtained when conducting hypotheses testing on a filtered contaminated square defected Gaussian image, of size 256 px × 256 px, using various kernel radiuses. The dashed lines show the resulting 95% empirical confidence interval when the hypotheses testing was done on the uncontaminated defected image. Defected pixels have the distribution $\mathrm{N}(3, 1)$. The square defect is 30 px × 30 px pixels in size. The contamination was such that the null distribution is $Z_{x,y}|H_{0,x,y} \sim \mathrm{N}(\mu_{0,x,y}, \sigma_0^2)$ where $\mu_{0,x,y} = 0.01(x - x_0) + 0.01(y - y_0)$ and $\sigma_0 = 2$. The test was done at the 5% FDR level. The boxplots summarise the 100 repeated simulations of the image.

ROC curve. The results of the uncontaminated defected image were all pooled together to obtain the empirical distribution of the AUC and errors without contamination. Results for the line and square defect are shown in Figures 6.37 and 6.38 respectively.

The results showed that the AUC increased with kernel radius. This highlighted that a kernel with a good radius can perform almost as though there was no contamination. It also appeared that the optimal AUC was achieved when using the kernel radius from the rule of thumb, making these results consistent with the literature.

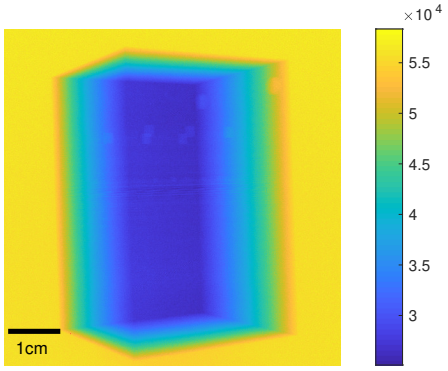
For large enough kernel radius, the FDR was controlled at around 5%. It appeared that a large kernel radius helped preserve FDR control after filtering in these examples. It was observed that there exist a kernel radius which minimised the type 2 error. So a large kernel radius is required for FDR control but too large can lose statistical power. It was noticed that filtering using large kernel radiuses is computationally slower.

Application to Real Projections

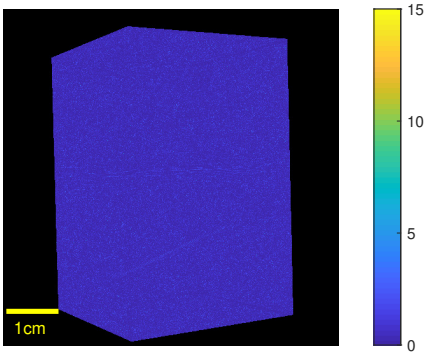
The empirical null filter was applied to the `AbsFilter` projection at 120°. Figures 6.39 and 6.40 shows the resulting inference when using a kernel radius of $r = 10\,\text{px}$ and $r = 130\,\text{px}$ respectively. With a small radius, the empirical null treated the voids as the null which is shown by the empirical null mean capturing the features of the voids. When a large enough radius was used, the larger voids were highlighted by the hypotheses test and the empirical null mean became smoother. This is evidence that this method ironed out false positives observed at the start of the chapter.

False positives were detected on the bottom right of the test sample. When placing the kernel on the bottom right corner, the null distribution is bimodal because each face has a different distribution. Figure 6.41 illustrates this. The geometry of the faces and the kernel used was such that it resulted in one of the faces to be treated as non-null and then tested positive, falsely so.

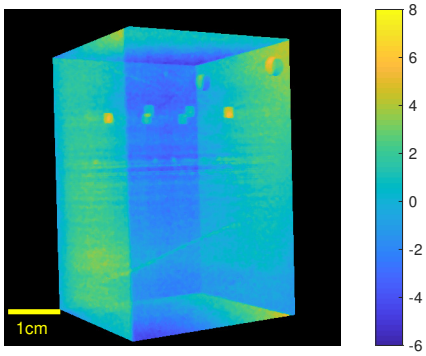
The false positives were tackled by manually segmenting the image fur-



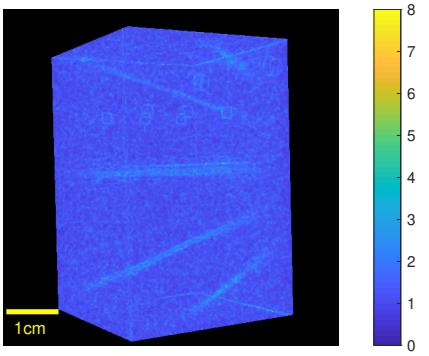
(a) Obtained projection (ADU)



(b) $-\log p$

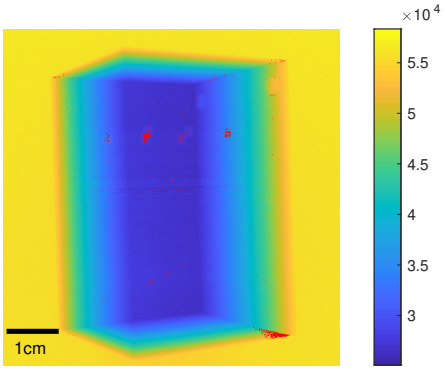


(c) Empirical null mean

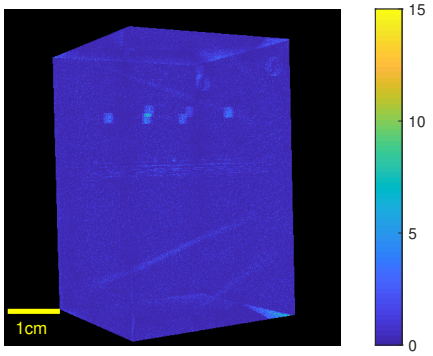


(d) Empirical null std

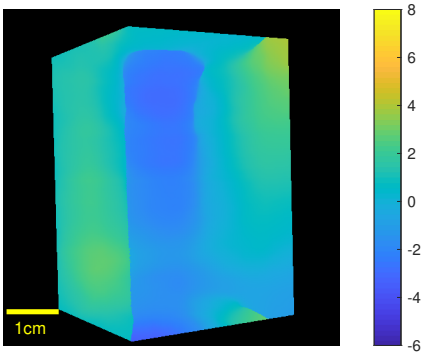
Figure 6.39: Resulting inference when using the empirical null filter on the `AbsFilter` projection at 120°. A kernel radius of $r = 10\,\text{px}$ was used. a) Highlighted in red are positive pixels at the 5% FDR level. b) p-values on the log scale. c) and d) Empirical null mean and empirical null standard deviation.



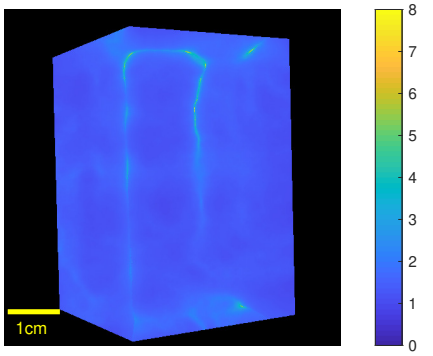
(a) Obtained projection (ADU)



(b) $-\log p$

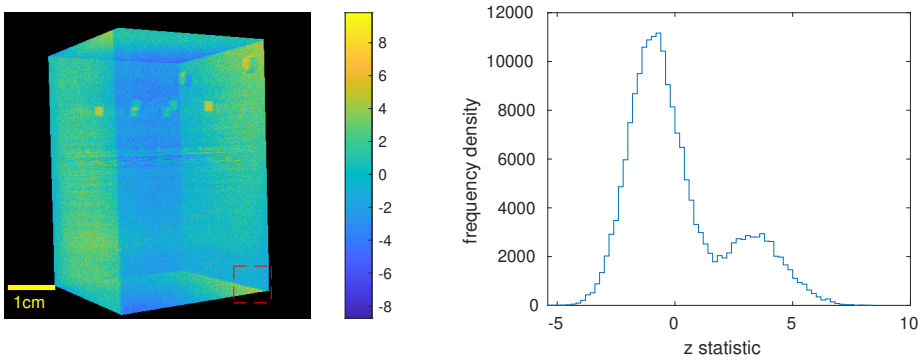


(c) Empirical null mean



(d) Empirical null std

Figure 6.40: Resulting inference when using the empirical null filter on the `AbsFilter` projection at 120°. A kernel radius of $r = 130\,\text{px}$ was used. a) Highlighted in red are positive pixels at the 5% FDR level. b) p-values on the log scale. c) and d) Empirical null mean and empirical null standard deviation.



(a) z image with section highlighted (b) Histogram of z in the kernel

Figure 6.41: A section of the bottom right of an unfiltered z image is shown in a). The z statistics in it are shown as a histogram in b) and indicates a bimodal distribution.

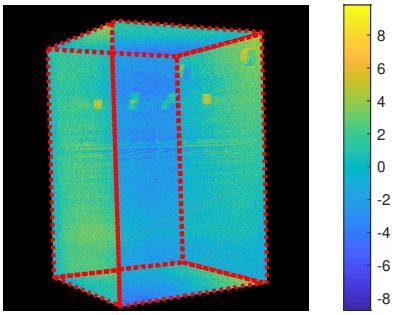


Figure 6.42: The z image was segmented further into 7 segments shown by the dotted red lines.

ther as shown in Figure 6.42. It was segmented using the edges of the z image. The empirical null filter was then used on each segment, or ROI, independently, ignoring any pixels outside the ROI the filter is working on. The resulting filtered segments were stitched together to form the resulting filtered image. By using this method, the resulting inference is shown in Figure 6.43. It can be seen that the empirical null mean did change face to face, resulting in a clear boundary set by the edges. Most of the false positives around the corners and edges have been eliminated.

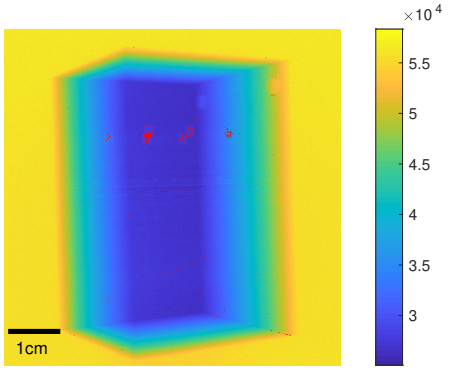
The same procedure was done on the projection at 30°, as shown in Figure 6.44, with similar results.

6.4 Conclusion

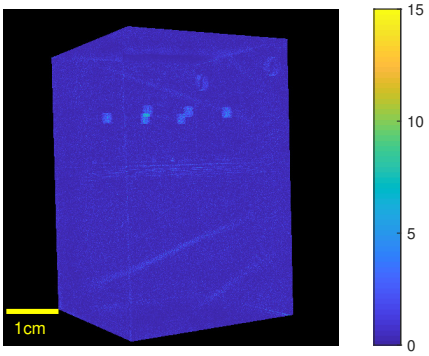
The empirical null filter demonstrated that it can adjust the null hypotheses according to the data to make a sensible inference. In the simulations, it was found that the FDR level is preserved after filtering. Also, the empirical null filter outperformed other methods based on quantiles, as suggested by Efron (2004). A sensible kernel radius is required to avoid treating defects as the null.

In the experiment, the larger voids of diameter 2.4 mm were detected in the test sample. False positives do occur but this is unavoidable because tests were done at the 5% FDR level. Typically in the experiments, false positives were isolated single pixels and probably occurred due to random chance. Clusters of positive pixels should raise suspicion so it would be a good idea to borrow strength from neighbouring pixels. For example, one could create a binary image, assigning a Boolean value whether that pixel was tested positive or not. A binary image filter, such as erode followed by a dilate, can be used to remove isolated positive pixels to emphasise the cluster of positive pixels.

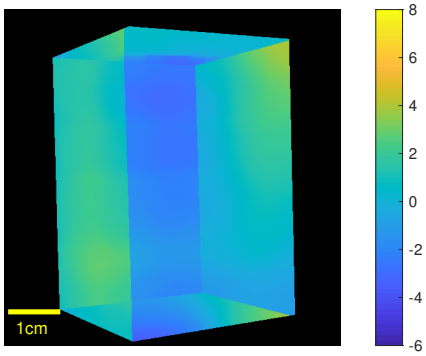
The assessment of the inference could be improved if the location of the defects were known in the projections. This would allow identifying which positives are true and false positives so that the analysis can be quantified using a ROC curve.



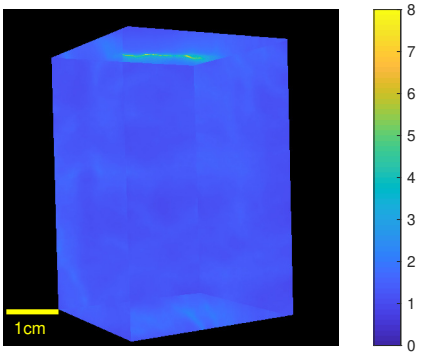
(a) Obtained projection (ADU)



(b) $-\log p$

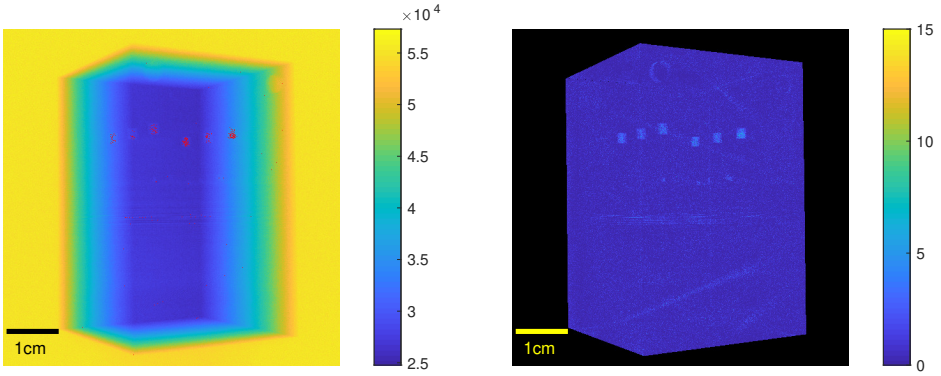


(c) Empirical null mean

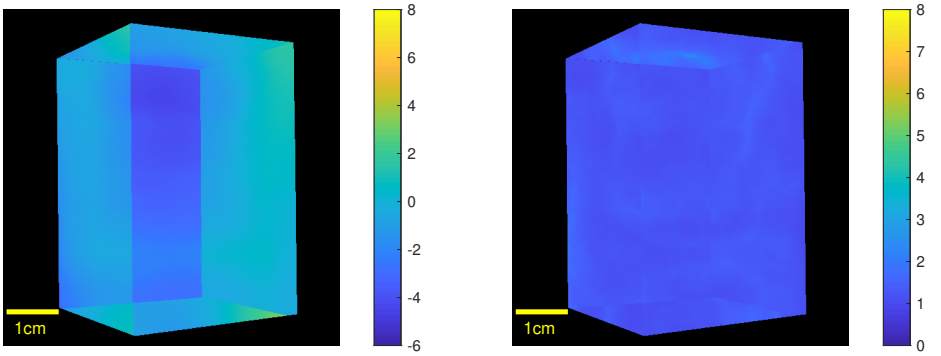


(d) Empirical null std

Figure 6.43: Resulting inference when using the empirical null filter on the `AbsFilter` projection at 120° and on each segment independently. A kernel radius of $r = 130\,\text{px}$ was used. a) Highlighted in red are positive pixels at the 5% FDR level. The test statistics from all segments were combined in the BH procedure. b) p-values on the log scale. c) and d) Empirical null mean and empirical null standard deviation.



(a) Obtained projection (ADU) (b) $-\log p$



(c) Empirical null mean (d) Empirical null std

Figure 6.44: Resulting inference when using the empirical null filter on the `AbsFilter` projection at 30° and on each segment independently. A kernel radius of $r = 130\,\text{px}$ was used. a) Highlighted in red are positive pixels at the 5% FDR level. The test statistics from all segments were combined in the BH procedure. b) p-values on the log scale. c) and d) Empirical null mean and empirical null standard deviation.

For good results, the faces of the test sample were segmented. This was necessary because the empirical null filter assumes that the null parameters varied spatially smoothly and slowly. Manual segmentation was easy to do because a cuboid has 6 faces. However, segmentation of faces cannot be generalised well to AM samples with complicated geometry and curved surfaces. Automatic segmentation may be possible with geometrical information from the CAD model.

The main issue with implementing this method onto the production line was how slow the filter was. The main bottleneck is the evaluation of the density estimate. Each step in the Newton-Raphson required the evaluation of πr^2 data points. As a result, increasing the kernel radius slows down the filter. There may be faster methods for density estimation such as fitting a smoothing spline on the histogram (Efron, 2004), however, that would require tuning the histogram bins as well as the tuning parameters for the spline. On the other hand, the method in Schwartzman et al. (2008b), which uses the histogram count, found that the estimation of the null parameters was insensitive to the histogram binning.

There are a few strategies to accelerate the filter. The empirical null filter may be accelerated using GPUs (Yang et al., 2008; Hwu, 2011; Eklund et al., 2013), however, efforts to implement the filter in *CUDA* and *C++* is fruitless if there may exist a faster method. Instead, accuracy may be sacrificed for speed by only estimating the null parameters for several regularly spaced pixels, the remaining pixels are interpolated. This requires the null parameters to be slowly varying, otherwise, the interpolation may underfit. This may cause problems because of the face to face transition observed in the experiments. It was found that if a bad solution was found for a point, then that solution would spread its bad solution to neighbouring pixels due to the interpolation.

Estimation of the null parameters is essentially robust statistics, estimating the parameters of the null distribution without being affected by non-null statistics. Potential faster methods compared to the empirical null could exist in the literature for robust statistics such as Hampel et al. (1986); Rousseeuw and Leroy (1987); Maronna et al. (2006); Huber and

Ronchetti (2009); Jewson et al. (2018). The use of density estimation for robust estimation is also featured when using loss functions derived from the Hellinger-divergence (Beran, 1977; Jewson et al., 2018) or the beta-divergence (Basu et al., 1998; Jewson et al., 2018) which bears similarities to the empirical null. However, numerical methods are still required to find the mode. Exact estimation of the mode using Bayesian methods is impossible (Heinrich, 2013) which suggest avoiding the use of numerical methods can be difficult.

The EM algorithm (Dempster et al., 1977) could be used to fit a mixture of Gaussians to identify the null distribution and estimate its parameters. However, the power of hypothesis testing and the empirical null comes from the fact that the alternative distribution does not need to be specified.

Chapter 7

Conclusion

To summarise the thesis, Figure 7.1 shows a flowchart of the process of the experiment. An ABS test sample, with purposefully designed voids, was created using a CAD model and manufactured using fused deposition modelling. Replicate x-ray projections were taken of the test sample with voids as well as a simulation of that projection without the voids using *aRTist*. The obtained projections were used to help align the simulated projection. The experiment aimed to develop a statistical method to detect the designed voids.

Shading correction was required to remove spot and panel effects from the obtained projections. This was done by using the greyscale projections at different exposures. By assuming each pixel has a linear response to the power of the x-ray tube, a linear regression was used. This is discussed in Chapter 3.

In Chapter 4, it was attempted to fit a compound Poisson distribution onto the grey values of a projection to quantify the noise. Unfortunately, the model suffered from identifiability issues for high photon counts. However, it was found that the grey value variance has a linear relationship with the grey value. In Chapter 5, various GLM were fitted onto the variance-mean data and it was verified a linear relationship is a good model. The model was used as a tool to predict the variance of a grey value.

In Chapter 6, the replicated projections were split into two. 19 randomly

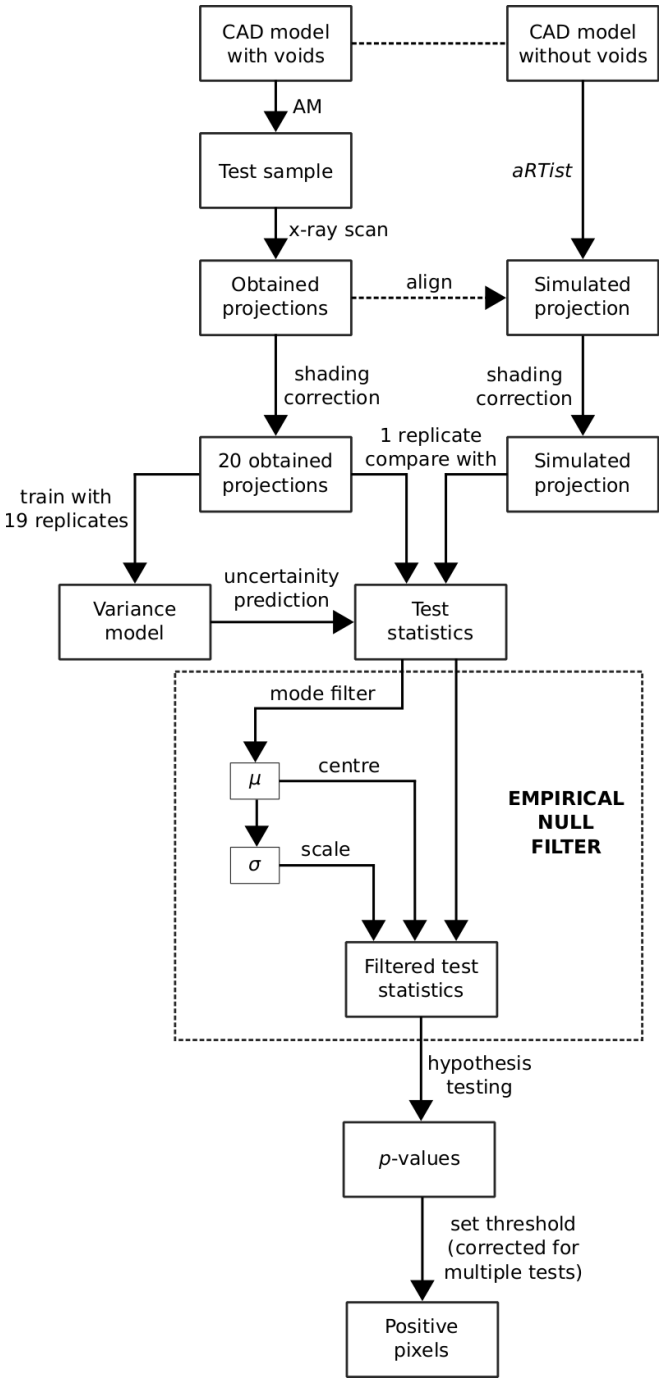


Figure 7.1: Flowchart showing the process of obtaining and comparing a projection of the test sample with the simulated projection. This results in pixels being highlighted as positive for defects.

selected projections were used for the variance-mean model to fit onto. The remaining projection was compared with the simulated projection in the face of uncertainty. The uncertainty was predicted using the variance-mean model. Hypotheses testing was done on the test statistics, one for each pixel in a projection. Unfortunately, the simulation was not perfect which led to model misspecification, thus, false positives. The empirical null filter was developed and used to smooth the test statistics. This allows the hypotheses test to pick out areas which are unusual and unlike its neighbours. The filtered test statistics were converted to p-values so that hypotheses testing can be done to detect the voids.

The main statistical novelty was extending the empirical null (Efron, 2004) to a filter where it can be applied to images. This has been implemented in *ImageJ* (Abràmoff et al., 2004; Schneider et al., 2012; Mateos-Pérez and Pascau, 2013). A by-product is the empirical null mean, effectively the mode filter, which can be used to smooth images.

Two angular projections were looked at in this thesis. The angles were selected by the engineers, in calibration with this experiment, so that all of the voids were visible in a single projection. Defect detection can be tough if a difficult projection angle was used. This experiment could be improved by analysing more angular projections independently to pick up voids which may be hidden at a particular projection angle. Another possible extension is to integrate *aRTist* into the inference, for example, the simulated projection could be realigned to minimise the number of positives. This will help eliminate sources of error due to *aRTist*. A problem with this suggestion is that aligning the simulated projection is a high dimensional problem (Brierley et al., 2018).

Voids in the order of millimetres were detectable using this method. Further work could include test samples made from materials such as titanium alloys (Ti-6Al-4V). This can be challenging because artefacts, such as streaks, can appear when using XCT for metal parts (De Man et al., 1999; Nawaz et al., 2014). Another problem is that because titanium products are created using a power-based method, the powder may become trapped (Brierley et al., 2018) in designed voids. This results in voids appearing with

less contrast in the projections, making it harder to detect. Different types of challenging defects could be investigated, for example, the orientation of cracks can affect the detectability of it (Wooldridge et al., 1997).

This thesis is a contribution to real-time quality control for AM and perhaps inverse problems in imaging (Bertero and Boccacci, 1998) and robust statistics. A powerful multi-core computer could be used to conduct the statistical analysis on a few x-ray projections of AM products on a conveyor belt. This could be competitive with other conveyor belt inspection methods, for example Warnett et al. (2016), by using fewer projections and hopefully less computational time. By using fewer projections, quality control can be faster which advances the development and increases the scope of applications of AM.

Bibliography

3D Systems Inc. (1989). *Sterolithography Interface Specifications.*

3D Systems Inc. (2019a). Our Story. `https://uk.3dsystems.com/our-story`. Accessed: 27/04/2019.

3D Systems Inc. (2019b). What Is An `STL` File? `https://uk.3dsystems.com/quickparts/learning-center/what-is-stl-file`. Accessed: 30/03/2019.

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with *ImageJ*. *Biophotonics International*, 11(7):36–42.

Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, 19(6):716–723.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1):125–127.

Amrhein, S., Rauer, M., and Kaloudis, M. (2014). Characterization of computer tomography scanners using the probability of detection method. *Journal of Nondestructive Evaluation*, 33(4):643–650.

Appleby, J., Mitchell, P. D., Robinson, C., Brough, A., Rutty, G., Harris, R. A., Thompson, D., and Morgan, B. (2014). The scoliosis of Richard III, last plantagenet king of England: Diagnosis and clinical significance. *The Lancet*, 383(9932):1944.

Arcam AB (2019). History. `http://www.arcam.com/company/about-arcam/history/`. Accessed: 27/04/2019.

Atwood, C., Griffith, M., Harwell, L., Schlienger, E., Ensz, M., Smugeresky, J., Romero, T., Greene, D., and Reckaway, D. (1998). Laser engineered net shaping (*LENS*™): A tool for direct fabrication of metal parts. In *Proceedings of the 17th International Congress on Applications of Lasers and Electro-Optics*. Laser Institute of America.

Bartscher, M., Hilpert, U., Goebbels, J., and Weidemann, G. (2007). Enhancement and proof of accuracy of industrial computed tomography (CT) measurements. *CIRP Annals - Manufacturing Technology*, 56(1):495–498.

Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2):147–185.

Bedrick, E. J. and Hill, J. R. (1992). An empirical assessment of saddlepoint approximations for testing a logistic regression parameter. *Biometrics*, 48(2):529–544.

Bell, C. (2014). *Maintaining and Troubleshooting Your 3D Printer*. Apress.

Bellon, C., Deresch, A., Gollwitzer, C., and Jaenisch, G.-R. (2012). Radiographic simulator *aRTist*: Version 2. In *Proceedings of the 18th World Conference on Non-Destructive Testing*. South African Institute for Non-Destructive Testing.

Bellon, C., Gollwitzer, C., Fratzscher, D., Ewert, U., and Jaenisch, G.-R. (2011). Simulation of complex scan paths for 3D reconstruction. In *Proceedings of the International Symposium on Digital Industrial Radiology*

and Computed Tomography. Deutsche Gesellschaft für zerstörungsfreie Prüfung e.V.

Bellon, C. and Jaenisch, G.-R. (2007). *aRTist* - analytical RT inspection simulation tool. In *Proceedings of the International Symposium on Digital Industrial Radiology and Computed Tomography*. Deutsche Gesellschaft für Zerstörungsfreie Prüfung e.V.

Benjamini, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Methodological)*, 72(4):405–416.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25(1):60–83.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, 5(3):445–463.

Bertero, M. and Boccacci, P. (1998). *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing.

Berumen, S., Bechmann, F., Lindner, S., Kruth, J.-P., and Craeghs, T. (2010). Quality control of laser and powder bed-based additive manufacturing (AM) technologies. *Physics Procedia*, 5:617–622.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bland, J. M. and Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal*, 310(6973):170.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Brettschneider, J. A., Thornby, J. A., Nichols, T. E., and Kendall, W. S. (2014). Spatial analysis of dead pixels. *CRiSM Working Paper Series*, 2014(24).

Brierley, N., Bellon, C., and Lazaro Toralles, B. (2018). Optimized multi-shot imaging inspection design. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474.

Brooks, R. A. and Di Chiro, G. (1976). Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Physics in Medicine and Biology*, 21(5):689.

Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Bryan, J. (1990). International status of thermal error research. *CIRP Annals*, 39(2):645 – 656.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge University Press.

Cantatore, A. and Müller, P. (2011). Introduction to computed tomography. Technical report, Technical University of Denmark Mechanical Engineering.

Carlton, H. D., Haboub, A., Gallegos, G. F., Parkinson, D. Y., and MacDowell, A. A. (2016). Damage evolution and failure mechanisms in additively manufactured stainless steel. *Materials Science and Engineering: A*, 651:406–414.

Cerniglia, D., Scafidi, M., Pantano, A., and Rudlin, J. (2015). Inspection of additive-manufactured layered components. *Ultrasonics*, 62:292–298.

Cierniak, R. (2011). *X-ray Computed Tomography in Biomedical Engineering*. Springer Science & Business Media.

Clijsters, S., Craeghs, T., Buls, S., Kempen, K., and Kruth, J.-P. (2014). In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system. *The International Journal of Advanced Manufacturing Technology*, 75(5):1089–1101.

Cole, S. R., Chu, H., and Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: A primer. *American Journal of Epidemiology*, 179(2):252–260.

Compton, A. H. (1923). A quantum theory of the scattering of x-rays by light elements. *Physical Review*, 21(5):483.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115(7):928–935.

Cooper, D., Thornby, J., Blundell, N., Henrys, R., Williams, M. A., and Gibbons, G. J. (2015). Design and manufacture of high performance hollow engine valves by additive layer manufacturing. *Materials and Design*, 69:44–55.

Corcoran, H. C., Brown, S. B., Robson, S., Speller, R. D., and McCarthy, M. (2016). Observations on the performance of x-ray computed tomography for dimensional metrology. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:25–31.

Cormack, A. M. (1973). Reconstruction of densities from their projections, with applications in radiological physics. *Physics In Medicine And Biology*, 18(2):195 – 207.

Cowan, G. (1998). *Statistical Data Analysis*. Oxford University Press.

Craeghs, T., Clijsters, S., Yasa, E., and Kruth, J.-P. (2011). Online quality control of selective laser melting. In *Proceedings of the 22nd Annual International Solid Freeform Fabrication Symposium - An Additive Manufacturing Conference*, pages 212 – 226. Laboratory for Freeform Fabrication and University of Texas.

Crump, S. S. (1991). Fused deposition modeling (FDM): Putting rapid back in prototyping. In *Proceedings of the 2nd International Conference Rapid Prototyping*, pages 358–361.

Crump, S. S. (1992). Apparatus and method for creating three-dimensional objects. US Patent 5121329.

Curran, S. C. (1953). *Luminescence and the Scintillation Counter*. Butterworths Scientific Publications.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25(4):631–650.

Davidson, D., Frjdh, C., O'Shea, V., Nilsson, H.-E., and Rahman, M. (2003). Limitations to flat-field correction methods when using an x-ray spectrum. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 509(1):146 – 150.

De Man, B., Nuyts, J., Dupont, P., Marchal, G., and Suetens, P. (1999). Metal streak artifacts in x-ray computed tomography: a simulation study. *Institute of Electrical and Electronics Engineers Transactions on Nuclear Science*, 46(3):691–696.

Deckard, C. R. (1989). Method and apparatus for producing parts by selective sintering. US Patent 4863538.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–38.

Deresch, A., Jaenisch, G.-R., Bellon, C., and Warrikhoff, A. (2012). Simulating x-ray spectra: From tube parameters to detector output. In *Proceedings of the 18th World Conference on Non-Destructive Testing*. South African Institute for Non-Destructive Testing.

DTM Corp. (1990). The selective laser sintering process, third generation desktop manufacturing. In *Proceedings of the National Conference on Rapid Prototyping*, pages 5–10.

Dunn, P. K. and Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4):267–280.

Dunn, P. K. and Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, 18(1):73–86.

Durnez, J., Moerkerke, B., and Nichols, T. E. (2014). Post-hoc power estimation for topological inference in fMRI. *NeuroImage*, 84:45 – 64.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96.

Efron, B. (2007). Size, power and false discovery rates. *The Annals of Statistics*, 35(4):1351.

Efron, B. (2008). Simultaneous inference: When should hypothesis testing problems be combined? *The Annals of Applied Statistics*, 2(1):197.

Efron, B. et al. (2003). Robbins, empirical bayes and microarrays. *The Annals of Statistics*, 31(2):366–378.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, 23(1):70 – 86.

Efroymson, M. A. (1960). Multiple regression analysis. In Ralston, A. and Wilf, H. S., editors, *Mathematical Methods for Digital Computers*, volume 1, chapter 17, pages 191–203. Wiley.

Eklund, A., Dufort, P., Forsberg, D., and LaConte, S. M. (2013). Medical image processing on the GPU–past, present and future. *Medical Image Analysis*, 17(8):1073–1094.

Elbakri, I. A. (2003). *Statistical Reconstruction Algorithms for Polyenergetic X-ray Computed Tomography*. PhD thesis, The University of Michigan.

Elbakri, I. A. and Fessler, J. A. (2001). Statistical x-ray computed tomography image reconstruction with beam hardening correction. In *Proceedings of SPIE Medical Imaging: Image Processing*, volume 4322.

Elbakri, I. A. and Fessler, J. A. (2002). Statistical image reconstruction for polyenergetic x-ray computed tomography. *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, 21(2):89–99.

Elbakri, I. A. and Fessler, J. A. (2003). Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography. In *Proceedings of SPIE Medical Imaging: Image Processing*, volume 5032.

PerkinElmer Optoelectronics (2006). *XRD 1621 AN/CN Digital X-ray Detector*.

Everton, S. K., Hirsch, M., Stravroulakis, P., Leach, R. K., and Clare, A. T. (2016). Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. *Materials and Design*, 95:431–445.

Feldkamp, L. A., Davis, L. C., and Kress, J. W. (1984). Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619.

Fewell, T. R. and Shuping, R. E. (1977). Photon energy distribution of some typical diagnostic x-ray beams. *Medical Physics*, 4(3):187–197.

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Oliver and Boyd, 14th edition.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1.

Fuchs, T. and Kachelrie, M. (2000). Direct comparison of a xenon and a solid-state CT detector system: Measurements under working conditions. *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, 19(9):941.

Gandossi, L. and Annis, C. (2010). Probability of detection curves: Statistical best-practices. Technical Report 41, European Network for Inspection and Qualification, Task Group on Risk.

Gatto, R. (2010). A saddlepoint approximation to the distribution of inhomogeneous discounted compound Poisson processes. *Methodology and Computing in Applied Probability*, 12(3):533–551.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.

Gibson, I., Rosen, D. W., and Stucker, B. (2010). *Additive Manufacturing Technologies: 3D Printing, Rapid Prototyping, and Direct Digital Manufacturing*. Springer New York.

Godoi, F. C., Prakash, S., and Bhandari, B. R. (2016). 3D printing technologies applied for food design: Status and prospects. *Journal of Food Engineering*, 179:44 – 54.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.

Greskovich, C. and Duclos, S. (1997). Ceramic scintillators. *Annual Review of Materials Science*, 27(1):69.

Griffin, L. D. (2000). Mean, median and mode filtering of images. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 456(2004):2995–3004.

Grossman, B. (2019). Bathsheba Sculpture LLC. `https://bathsheba.com/`. Accessed: 29/03/2019.

Gu, D. and Shen, Y. (2009). Balling phenomena in direct laser sintering of stainless steel powder: Metallurgical mechanisms and control methods. *Materials and Design*, 30(8):2903–2910.

Günaydin, K. and Türkmen, H. S. (2018). Common FDM 3D printing defects. In *Proceedings of the 3rd International Congress on 3D Printing Technologies and Digital Industry*, pages 671–676. 3D Print Turkey.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics*. Wiley.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.

Hanssen, J. (2013). *Fortus 360mc/400mc* accuracy study. Technical report, *Stratasys Ltd.*

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hayden, G. E. and Wrenn, K. W. (2009). Chest radiograph verus computed tomography scan in the evaluation for pneumonia. *The Journal of Emergency Medicine*, 36(3):266 – 270.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., et al. (2001). Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine*, 344(8):539–548.

Heinrich, C. (2013). The mode functional is not elicitable. *Biometrika*, 101(1):245–251.

Herzog, D., Seyda, V., Wycisk, E., and Emmelmann, C. (2016). Additive manufacturing of metals. *Acta Materialia*, 117:371 – 392.

Herzog, F., Bechmann, F., Berumen, S., Kruth, J.-P., and Craeghs, T. (2013). Method for producing a three-dimensional component. US Patent Specification 0168902.

Hoffman, M. D. and Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Hounsfield, G. N. (1973). Computerized transverse axial scanning (tomography): Part 1. Description of system. *The British Journal of Radiology*, 46(552):1016–1022.

Hounsfield, G. N. (1980). Computed medical imaging. *Medical Physics*, 7(4):283–290.

Hsieh, H.-H., Ni, Y.-C., Tseng, S.-P., Tseng, F.-P., and Chen, L.-C. (2015). Compound poisson noise verification for x-ray flat panel imager. In *Proceedings of the IEEE International Conference on Imaging Systems and Techniques*, pages 1–5. Institute of Electrical and Electronics Engineers.

Hsieh, J., Gurmen, O., and King, K. F. (2000). Investigation of a solid-state detector for advanced computed tomography. *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, 19(9):930–940.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. Wiley, 2nd edition.

Hull, C. W. (1986). Apparatus for production of three-dimensional objects by stereolithography. US Patent 4575330.

Hwu, W.-M. W. (2011). *GPU Computing Gems Emerald Edition*. Elsevier.

ImageJ (2018). `ImageJA/RankFilters.java` at `7f965b866c9db364b0b47140caeef4f62d5d8c15` `imagej/ImageJA`. `https://github.com/imagej/ImageJA/blob/7f965b866c9db364b0b47140caeef4f62d5d8c15/src/main/java/ij/plugin/filter/RankFilters.java`. Accessed: 16/10/2018.

Jaenisch, G.-R., Bellon, C., and Ewert, U. (2008). *aRTist*–analytical RT inspection simulation tool for industrial application. In *Proceedings of the 17th World Conference on Non-Destructive Testing*. Chinese Society for Non-Destructive Testing.

Jamieson, R. and Hacker, H. (1995). Direct slicing of CAD models for rapid prototyping. *Rapid Prototyping Journal*, 1(2):4–12.

Jensen, J. L. (1991). Saddlepoint approximations to the distribution of the total claim amount in some recent risk models. *Scandinavian Actuarial Journal*, 1991(2):154–168.

Jewson, J., Smith, J., and Holmes, C. (2018). Principles of Bayesian inference using general divergence criteria. *Entropy*, 20(6):442.

Jin, J. and Cai, T. T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association*, 102(478):495–506.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, pages 127–162.

Jørgensen, B. and Paes De Souza, M. C. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 1994(1):69–93.

Kang, H.-W., Lee, S. J., Ko, I. K., Kengla, C., Yoo, J. J., and Atala, A. (2016). A 3D bioprinting system to produce human-scale tissue constructs with structural integrity. *Nature Biotechnology*, 34(3):312–319.

Kiekens, K., Welkenhuyzen, F., Tan, Y., Bleys, P., Voet, A., Kruth, J.-P., and Dewulf, W. (2011). A test object with parallel grooves for calibration and accuracy assessment of industrial computed tomography (CT) metrology. *Measurement Science and Technology*, 22(11):115502.

Kim, F. H., Villarraga-Gómez, H., and Moylan, S. P. (2016). Inspection of embedded internal features in additively manufactured metal parts using metrological x-ray computed tomography. In *Proceedings of the American Society for Precision Engineering and European Society for Precision Engineering and Nanotechnology Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing*, volume 64.

Kim, G. D. and Oh, Y. T. (2008). A benchmark study on rapid prototyping processes and machines: Quantitative comparisons of mechanical properties, accuracy, roughness, speed, and material cost. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 222(2):201–215.

Kodama, H. (1981). Automatic method for fabricating a three-dimensional plastic model with photo-hardening polymer. *Review of Scientific Instruments*, 52(11):1770–1773.

Kourra, N., Warnett, J. M., Attridge, A., Dibling, G., McLoughlin, J., Muirhead-Allwood, S., King, R., and Williams, M. A. (2018). Computed tomography metrological examination of additive manufactured acetabular hip prosthesis cups. *Additive Manufacturing*, 22:146 – 152.

Kruth, J.-P. (1991). Material incress manufacturing by rapid prototyping techniques. *CIRP Annals*, 40(2):603 – 614.

Kruth, J.-P., Bartscher, M., Carmignato, S., Schmitt, R., Chiffre, L. D., and Weckenmann, A. (2011). Computed tomography for dimensional metrology. *CIRP Annals*, 60(2):821 – 842.

Kruth, J.-P., Leu, M.-C., and Nakagawa, T. (1998). Progress in additive manufacturing and rapid prototyping. *CIRP Annals*, 47(2):525–540.

Kueh, A., Warnett, J. M., Gibbons, G. J., Brettschneider, J. A., Nichols, T. E., Williams, M. A., and Kendall, W. S. (2016). Modelling the penumbra in computed tomography. *Journal of X-ray Science and Technology*, 24(4):583–597.

Langaas, M., Lindqvist, B. H., and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 67(4):555–572.

Larsson, M. and Larsson, M. (2004). Arrangement for the production of a three-dimensional product. Patent Cooperation Treaty WO2004054743A1.

Lasio, G. M., Whiting, B. R., and Williamson, J. F. (2007). Statistical reconstruction for x-ray computed tomography using energy-integrating detectors. *Physics in Medicine and Biology*, 52(8):2247.

Lee, C. and Tarbutton, J. A. (2015). Compliance and control characteristics of an additive manufactured-flexure stage. *Review of Scientific Instruments*, 86(4):045107.

Leuders, S., Vollmer, M., Brenne, F., Tröster, T., and Niendorf, T. (2015). Fatigue strength prediction for titanium alloy TiAl6V4 manufactured by selective laser melting. *Metallurgical and Materials Transactions A*, 46(9):3816–3823.

Li, R., Liu, J., Shi, Y., Wang, L., and Jiang, W. (2012). Balling behavior of stainless steel and nickel powder during selective laser melting process. *The International Journal of Advanced Manufacturing Technology*, 59(9-12):1025–1035. doi:https://doi.org/10.1007/s00170-011-3566-1.

Lifton, J. J., Malcolm, A. A., McBride, J. W., and Cross, K. J. (2013). The application of voxel size correction in x-ray computed tomography for dimensional metrology. In *Proceedings of the Singapore International Non-Destructive Testing Conference and Exhibition*. Non-Destructive Testing Society of Singapore.

Likar, B., Maintz, J. A., Viergever, M. A., Pernus, F., et al. (2000). Retrospective shading correction based on entropy minimization. *Journal of Microscopy*, 197(3):285–295.

Lott, P., Schleifenbaum, H., Meiners, W., Wissenbach, K., Hinke, C., and Bltmann, J. (2011). Design of an optical system for the in situ process monitoring of selective laser melting (SLM). *Physics Procedia*, 12:683 – 690.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

Ma, J., Liang, Z., Fan, Y., Liu, Y., Huang, J., Li, L., Chen, W., and Lu, H. (2012). Variance estimation of x-ray CT sinogram in radon domain. In *Proceedings of SPIE Medical Imaging: Physics of Medical Imaging*, volume 8313.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley.

Mateos-Pérez, J. M. and Pascau, J. (2013). *Image Processing with* ImageJ. Packt Publishing Ltd.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, 16(3):285–292.

Mercelis, P. and Kruth, J.-P. (2006). Residual stresses in selective laser sintering and selective laser melting. *Rapid Prototyping Journal*, 12(5):254–265.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4):283–298.

Michael, G. (2001). X-ray computed tomography. *Physics Education*, 36(6):442. doi:https://doi.org/10.1088%2F0031-9120%2F36%2F6%2F301.

Millikan, R. A. (1916). A direct photoelectric determination of Planck's "*h*". *Physical Review*, 7(3):355.

Moon, T. K. (1996). The expectation-maximization algorithm. *Institute of Electrical and Electronics Engineers Signal Processing Magazine*, 13(6):47–60.

Mumtaz, K. A., Erasenthiran, P., and Hopkinson, N. (2008). High density selective laser melting of *Waspaloy*®. *Journal of Materials Processing Technology*, 195(1-3):77–87.

Münzenmayer, C., Naujokat, F., Mühldorfer, S., and Wittenberg, T. (2003). Enhancing texture analysis by color shading correction. In *Proceedings of the 9th Workshop Farbbildverarbeitung*. Zentrum für Bild-und Signalverarbeitung e.V.

Murray, I., Prescott Adams, R., and MacKay, D. J. (2010). Elliptical slice sampling. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*.

Nawaz, S., Fu, J., and Fan, D. (2014). Metal artifacts reduction in x-ray CT based on segmentation and forward-projection. *Bio-medical Materials and Engineering*, 24(6):3287–3293.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–741.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, chapter 5, pages 113–162. CRC press.

Nelder, J. A. and Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.

Nelder, J. A. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London: Series A*, 231(694-706):289–337.

Ngo, T. D., Kashani, A., Imbalzano, G., Nguyen, K. T., and Hui, D. (2018). Additive manufacturing (3D printing): A review of materials, methods, applications and challenges. *Composites Part B: Engineering*, 143:172 – 196.

Nikl, M. (2006). Scintillation detectors for x-rays. *Measurement Science and Technology*, 17(4):R37–R54.

Nikon Metrology NV (2015a). Inline Quality Control using Computed Tomography. *YouTube*, viewable online `https://youtu.be/b7a71BTvquE`. Accessed: 19/04/2019.

Nikon Metrology NV (2015b). Microfocus Computed Tomography. *YouTube*, viewable online `https://youtu.be/tq9yAKzz9zg`. Accessed: 19/04/2019.

Nikon Metrology NV (2018a). Automated CT - Full Control of Your Production Process. Brochure, retrieved online `https://www.nikonmetrology.com/images/brochures/automated-ct-en.pdf`. Accessed: 19/04/2019.

Nikon Metrology NV (2018b). MCT225 - Metrology CT. Brochure, retrieved online `https://www.nikonmetrology.com/images/brochures/mct225-en.pdf`. Accessed: 19/04/2019.

Niu, H. J. and Chang, I. T. H. (1999). Instability of scan tracks of selective laser sintering of high speed steel powder. *Scripta Materialia*, 41(11):1229–1234.

Ornes, S. (2013). Mathematics in metal. *Proceedings of the National Academy of Sciences*, 110(44):17603–17603.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, 316(7139):1236.

Pham, D. T. and Gault, R. S. (1998). A comparison of rapid prototyping technologies. *International Journal of Machine Tools and Manufacture*, 38(10-11):1257–1287.

Pounds, S. and Cheng, C. (2004). Improving false discovery rate estimation. *Bioinformatics*, 20(11):1737–1745.

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*, 19(10):1236–1242.

Radon, J. (1986). On the determination of functions from their integral values along certain manifolds. *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, 5(4):170–176.

Rale, A. P., Gharpure, D. C., and Ravindran, V. R. (2009). Comparison of different ANN techniques for automatic defect detection in x-ray images. In *International Conference on Emerging Trends in Electronic and Photonic Devices Systems*, pages 193–197. Institute of Electrical and Electronics Engineers.

Reinhart, C. (2008). Industrial computer tomography - a universal inspection tool. In *Proceedings of the 17th World Conference on Non-*

Destructive Testing, volume 1, pages 1561–1571. Chinese Society for Non-Destructive Testing.

Reisinger, S., Kasperl, S., Franz, M., Hiller, J., and Schmid, U. (2011). Simulation-based planning of optimal conditions for industrial computed tomography. In *Proceedings of the International Symposium on Digital Industrial Radiology and Computed Tomography*. Deutsche Gesellschaft für zerstörungsfreie Prüfung e.V.

Reiter, M., Harrer, B., Heinzl, C., Salaberger, D., Gusenbauer, C., Kuhn, C., and Kastner, J. (2011). Simulation aided study for optimising industrial x-ray CT scan parameters for non-destructive testing and materials characterisation. In *Proceedings of the International Symposium on Digital Industrial Radiology and Computed Tomography*. Deutsche Gesellschaft für zerstörungsfreie Prüfung e.V.

Ren, L., Zheng, B., and Liu, H. (2018). Tutorial on x-ray photon counting detector characterization. *Journal of X-ray Science and Technology*, 26(1):1–28.

Revfeim, K. J. A. (1984). An initial model of the relationship between rainfall events and daily rainfalls. *Journal of Hydrology*, 75(1):357 – 364.

Rodriguez, E., Medina, F., Espalin, D., Terrazas, C., Muse, D., Henry, C., Macdonald, E., and Wicker, R. (2012). Integration of a thermal imaging feedback control system in electron beam melting. In *Proceedings of the 23rd Annual International Solid Freeform Fabrication Symposium - An Additive Manufacturing Conference*, pages 945–961. Laboratory for Freeform Fabrication and University of Texas.

Röntgen, W. C. (1896). On a new kind of rays. *Science*, 3(59):227–231.

Rossner, W., Bodinger, H., Leppert, J., and Grabmaier, B. (1993). The conversion of high energy radiation to visible light by luminescent ceramics. *Institute of Electrical and Electronics Engineers Transactions on Nuclear Science*, 40(4):376.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley.

Sachs, E., Cima, M., and Cornie, J. (1990). Three-dimensional printing: Rapid tooling and prototypes directly from a CAD model. *CIRP Annals*, 39(1):201 – 204.

Sage, D. (2018). Local normalization. Biomedical Image Group, `http://bigwww.epfl.ch/sage/soft/localnormalization/`. Accessed: 29/01/2019.

Sage, D. and Unser, M. (2003). Teaching image-processing programming in *Java*. *Institute of Electrical and Electronics Engineers Signal Processing Magazine*, 20(6):43–52.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). *Fiji*: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to *ImageJ*: 25 years of image analysis. *Nature Methods*, 9(7):671.

Schwartzman, A., Dougherty, R. F., Lee, J., Ghahremani, D., and Taylor, J. E. (2009). Empirical null and false discovery rate analysis in neuroimaging. *Neuroimage*, 44(1):71–82.

Schwartzman, A., Dougherty, R. F., Taylor, J. E., et al. (2008a). False discovery rate analysis of brain diffusion direction maps. *The Annals of Applied Statistics*, 2(1):153–175.

Schwartzman, A. et al. (2008b). Empirical null and false discovery rate inference for exponential families. *The Annals of Applied Statistics*, 2(4):1332–1359.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Seibert, J. A., Boone, J. M., and Lindfors, K. K. (1998). Flat-field correction technique for digital detectors. In *Proceedings of SPIE Medical Imagine: Physics of Medical Imaging*, volume 3336.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46(1):561–584.

Sheather, S. J. (2004). Density estimation. *Statistical Science*, 19(4):588.

Shevtsova, I. G. (2014). On the accuracy of the Normal approximation to compound Poisson distributions. *Theory of Probability and its Applications*, 58(1):138–158.

Siddique, S., Imran, M., Rauer, M., Kaloudis, M., Wycisk, E., Emmelmann, C., and Walther, F. (2015). Computed tomography for characterization of fatigue performance of selective laser melted parts. *Materials and Design*, 83:661–669.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, pages 751–754.

Smith, B. D. (1990). Cone-beam tomography: Recent advances and a tutorial review. *Optical Engineering*, 29(5):524 – 534.

Smyth, G. K. and Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *Astin Bulletin*, 32(01):143–157.

Spierings, A. B., Schneider, M., and Eggenberger, R. (2011). Comparison of density measurement techniques for additive manufactured metallic parts. *Rapid Prototyping Journal*, 17(5):380–386.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–147.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):44–47.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Methodological)*, 64(3):479–498.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, pages 2013–2035.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.

Stratasys Ltd. (2019). What is FDM?: Fused Deposition Modeling Technology for 3D Printing. `https://www.stratasys.com/fdm-technology`. Accessed: 27/04/2019.

Stumbo, S., Bottigli, U., Golosio, B., Oliva, P., and Tangaro, S. (2004). Direct analysis of molybdenum target generated x-ray spectra with a portable device. *Medical Physics*, 31(10):2763–2770.

Sun, W., Brown, S. B., Flay, N., McCarthy, M., and McBride, J. (2016a). A reference sample for investigating the stability of the imaging system of x-ray computed tomography. *Measurement Science and Technology*, 27(8).

Sun, W., Brown, S. B., and Leach, R. K. (2012). An overview of industrial x-ray computed tomography. Technical report, National Physical Laboratory.

Sun, W., Brown, S. B., Woolliams, P., McCarthy, M., White, J., and Attallah, M. (2016b). Applications and systematic errors of x-ray computed tomography associated with dimensional metrology. In *55th Annual Conference of the British Institute for Non-Destructive Testing*.

Sun, W., Konstantinidis, A., Billas, I., McCarthy, M., Brown, S. B., and Corcoran, H. (2016c). Characterisation of a scintillation flat panel detector employed by an x-ray computed tomography system: Measurements of MTF, NPS and DQE. In *Proceedings of the European Society for Precision Engineering and Nanotechnology's 16th International Conference and Exhibition*.

Taiwo, O. O., Finegan, D. P., Paz-Garcia, J. M., Eastwood, D. S., Bodey, A. J., Rau, C., Hall, S. A., Brett, D. J. L., Lee, P. D., and Shearing, P. R. (2017). Investigating the evolving microstructure of lithium metal electrodes in 3D using x-ray computed tomography. *Physical Chemistry Chemical Physics*, 19:22111–22120.

Tammas-Williams, S., Zhao, H., Léonard, F., Derguti, F., Todd, I., and Prangnell, P. B. (2015). XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by selective electron beam melting. *Materials Characterization*, 102:47–61.

Taud, H., Martinez-Angeles, R., Parrot, J. F., and Hernandez-Escobedo, L. (2005). Porosity estimation method by x-ray computed tomography. *Journal of Petroleum Science and Engineering*, 47(3-4):209–217.

Thijs, L., Verhaeghe, F., Craeghs, T., Van Humbeeck, J., and Kruth, J.-P. (2010). A study of the microstructural evolution during selective laser melting of Ti–6Al–4V. *Acta Materialia*, 58(9):3303–3312.

Thompson, A., Maskery, I., and Leach, R. K. (2016). X-ray computed tomography for additive manufacturing: A review. *Measurement Science and Technology*, 27(7):1.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Vatani, M., Rahimi, A., Brazandeh, F., and Nezhad, A. S. (2009). An enhanced slicing algorithm using nearest distance analysis for layer man-

ufacturing. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 37, pages 721–726.

Villarraga-Gómez, H., Lee, C., Corbett, T., Tarbutton, J. A., and Smith, S. T. (2015). Assessing additive manufacturing processes with x-ray CT metrology. In *Proceedings of the American Society for Precision Engineering Spring Topical Meeting: Achieving Precision Tolerances in Additive Manufacturing*, volume 60, pages 116–121.

Wang, X., Jiang, M., Zhou, Z., Gou, J., and Hui, D. (2017a). 3D printing of polymer matrix composites: A review and prospective. *Composites Part B: Engineering*, 110:442 – 458.

Wang, Y., Burnett, T. L., Chai, Y., Soutis, C., Hogg, P. J., and Withers, P. J. (2017b). X-ray computed tomography study of kink bands in unidirectional composites. *Composite Structures*, 160:917–924.

Warnett, J. M., Titarenko, V., Kiraci, E., Attridge, A., Lionheart, W. R. B., Withers, P. J., and Williams, M. A. (2016). Towards in-process x-ray CT for dimensional metrology. *Measurement Science and Technology*, 27(3):1–14.

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019). Moving to a world beyond $p < 0.05$. *The American Statistician*, 73(sup1):1–19.

Weeren, R. V., Agarwala, M., Jamalabad, V. R., Bandyopadhyay, A., Vaidyanathan, R., Langrana, N., Safari, A., Whalen, P., Danforth, S. C., and Ballard, C. (1995). Quality of parts processed by fused deposition. In *Proceedings of the 6th Annual International Solid Freeform Fabrication Symposium*, pages 314–321. Laboratory for Freeform Fabrication and University of Texas.

Welkenhuyzen, F., Kiekens, K., Pierlet, M., Dewulf, W., Bleys, P., Kruth, J.-P., and Voet, A. (2009). Industrial computer tomography for dimensional metrology: Overview of influence factors and improvement strategies. In *Proceedings of the 4th International Conference on Optical Mea-*

surement Techniques for Structures and Systems: Optimess2009, pages 401–410.

Whiting, B. R. (2002). Signal statistics in x-ray computed tomography. In *Proceedings of SPIE Medical Imaging: Physics of Medical Imaging*, volume 4682.

Whiting, B. R., Massoumzadeh, P., Earl, O. A., O'Sullivan, J. A., Snyder, D. L., and Williamson, J. F. (2006). Properties of preprocessed sinogram data in x-ray computed tomography. *Medical Physics*, 33(9):3290–3303.

Williams, C. K. I. and Rasmussen, C. E. (1996). Gaussian processes for regression. In *Proceedings of Advances in Neural Information Processing Systems*, pages 514–520.

Withers, C. and Nadarajah, S. (2011). On the compound Poisson-gamma distribution. *Kybernetika*, 47(1):15–37.

Wong, K. V. and Hernandez, A. (2012). A review of additive manufacturing. *International Scholarly Research Network Mechanical Engineering*, 2012.

Wooldridge, A. B., Chapman, R. K., Woodcock, G. S., Munns, I. J., and Georgiou, G. A. (1997). Demonstrating the capability of radiography for detection of large planar defects in thick-section welds. In *Proceedings of the European-American workshop: Determination of Reliability and Validation Methods on Nondestructive Evaluation*. Deutsche Gesellschaft fuer Zerstoerungsfreie Pruefung e.V.

Xie, L. (2008). *X-ray Computed Tomography Image Reconstruction: Energy-integrating Detector and Performance Analysis*. PhD thesis, Washington University, Department of Electrical and Systems Engineering.

Xu, J. and Tsui, B. M. W. (2009). Electronic noise modeling in statistical iterative reconstruction. *Institute of Electrical and Electronics Engineers Transactions on Image Processing*, 18(6):1228–1238.

Yang, E., Ravikumar, P. K., Allen, G. I., and Liu, Z. (2013). On Poisson graphical models. In *Proceedings of Advances in Neural Information Processing Systems*, pages 1718–1726.

Yang, K., Huang, S.-Y., Packard, N. J., and Boone, J. M. (2009). Evaluation of the additive noise of a flat panel detector and its effect on cone-beam CT applications. In *Proceedings of SPIE Medical Imaging: Physics of Medical Imaging*, volume 7258.

Yang, K., Huang, S.-Y., Packard, N. J., and Boone, J. M. (2010). Noise variance analysis using a flat panel x-ray detector: A method for additive noise assessment with application to breast CT applications. *Medical Physics*, 37(7):3527–3537.

Yang, Z., Zhu, Y., and Pu, Y. (2008). Parallel image processing based on *CUDA*. In *Proceedings of the International Conference on Computer Science and Software Engineering*, volume 3, pages 198–201. Institute of Electrical and Electronics Engineers.

Young, I. T. (2000). Shading correction: Compensation for illumination and sensor inhomogeneities. *Current Protocols in Cytometry*, 14(1):2–11.

Zaeh, M. F. and Branner, G. (2010). Investigations on residual stresses and deformations in selective laser melting. *Production Engineering*, 4(1):35–45.

Zeng, G. L. (2010). *Medical Image Reconstruction: A Conceptual Tutorial*. Springer.

Zhang, X., Aliasghari, S., Němcová, A., Burnett, T. L., Kuběna, I., Šmíd, M., Thompson, G. E., Skeldon, P., and Withers, P. J. (2016). X-ray computed tomographic investigation of the porosity and morphology of plasma electrolytic oxidation coatings. *American Chemical Society Applied Materials and Interfaces*, 8(13):8801–8810.

Zhang, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing*, 23(6):743–757.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Methodology)*, 67(2):301–320.

Appendix A

Abbreviations

3D	Three dimensions
ABS	Acrylonitrile butadiene styrene
ADU	Analogue to digital units
AIC	Akaike information criterion
AM	Additive manufacturing
ANOVA	Analysis of variance
Approx.	Approximate
AUC	Area under the receiver operating characteristic curve
BH	Benjamini and Hochberg
BIC	Bayesian information criterion
BW	Black/white
CAD	Computer-aided design
CC BY	Creative Commons Attribution
CPΓ	Compound Poisson-gamma
EM	Expectation maximisation
Expec.	Expected
FDK	Feldkamp-Davis-Kress
FDR	False discovery rate
FWER	Family-wise error rate
Freq.	Frequency
GLM	Generalised linear models

GPU	Graphics processing unit
i.i.d.	Independent and identically distributed
IOP	Institute of Physics
IQR	Interquartile range
MADA-mode	Median absolute deviation around the mode
m.g.f.	Moment generating function
Obs.	Observed
PCER	Per-comparison error rate
p.d.f.	Probability density function
px	Pixels
Q-Q	Quantile-quantile
ROC	Receiver operating characteristic
ROI	Region of interest
Std	Standard deviation
STL	Sterolithography or Standard tessellation language
Var	Variance
XCT	X-ray computed tomography

Appendix B

Compound Poisson

B.1 Moment Generating Function

Let $Y \sim \text{Poisson}(\lambda)$, $U_1, U_2, U_3, \ldots$ be i.i.d. random variables and $X|Y = \sum_{i=1}^{Y} U_i$. Let the m.g.f. of X be $M_X(\theta) = \mathbb{E}\left[e^{\theta X}\right]$. This can be computed using the result for conditional expectations $M_X(\theta) = \mathbb{E}\,\mathbb{E}\left[e^{X\theta}|Y\right]$. This results in

$$\begin{aligned} M_X(\theta) &= \mathbb{E}\,\mathbb{E}\left[\exp\left(\theta U_1 + \theta U_2 + \ldots + \theta U_Y\right)|Y\right] \\ &= \mathbb{E}\,\mathbb{E}\left[e^{\theta U_1} \cdot e^{\theta U_2} \cdot \ldots \cdot e^{\theta U_Y}|Y\right] . \end{aligned} \tag{B.1}$$

Because U_i, for $i = 1, 2, 3, \ldots$, are i.i.d. and each U_i has a m.g.f. $M_U(\theta) = \mathbb{E}\left[e^{U\theta}\right]$, then

$$\begin{aligned} M_X(\theta) &= \mathbb{E}\left(\mathbb{E}\left[e^{\theta U_1}|Y\right] \cdot \mathbb{E}\left[e^{\theta U_2}|Y\right] \cdot \ldots \cdot \mathbb{E}\left[e^{\theta U_Y}|Y\right]\right) \\ &= \mathbb{E}\left[\left(M_U(\theta)\right)^Y\right] \\ &= \mathbb{E}\left[e^{Y \ln(M_U(\theta))}\right] \\ &= M_Y\left(\ln(M_U(\theta))\right) \end{aligned} \tag{B.2}$$

where $M_Y(\theta) = \mathbb{E}\left[e^{Y\theta}\right]$ is the m.g.f. of Y. It can be shown that the m.g.f. of Y is $M_Y(\theta) = \exp\left[\lambda\left(e^{\theta} - 1\right)\right]$, hence

$$M_X(\theta) = \exp\left[\lambda\left(M_U(\theta) - 1\right)\right] . \tag{B.3}$$

Moments of X can be obtained from the m.g.f. by differentiating it and setting θ to zero

$$\begin{aligned} M'_X(\theta) &= \exp\left[\lambda\left(M_U(\theta)-1\right)\right]\cdot\lambda M'_U(\theta) \\ &= M_X(\theta)\lambda M'_U(\theta) \qquad \text{(B.4)} \\ M'_X(0) &= \lambda\,\mathbb{E}\left[U\right] \qquad \text{(B.5)} \end{aligned}$$

which results in

$$\mathbb{E}\left[X\right] = \lambda\,\mathbb{E}\left[U\right] \ . \qquad \text{(B.6)}$$

Conducting the same procedure

$$\begin{aligned} M''_X(\theta) &= M'_X(\theta)\lambda M'_U(\theta) + M_X(\theta)\lambda M''_U(\theta) \\ &= \lambda M_X(\theta)\left[\lambda\left(M'_U(\theta)\right)^2 + M''_U(\theta)\right] \qquad \text{(B.7)} \\ M''_X(0) &= \lambda\left[\lambda\left(\mathbb{E}[U]\right)^2 + \mathbb{E}\left[U^2\right]\right] \qquad \text{(B.8)} \end{aligned}$$

the variance can be obtained

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[X\right] &= M''_X(0) - \left[M'_X(0)\right]^2 \\ &= \lambda\left[\lambda\left(\mathbb{E}[U]\right)^2 + \mathbb{E}\left[U^2\right]\right] - \left[\lambda\,\mathbb{E}\left[U\right]\right]^2 \\ &= \lambda\,\mathbb{E}\left[U^2\right] \ . \qquad \text{(B.9)} \end{aligned}$$

Differentiating the m.g.f. one more time

$$\begin{aligned} M'''_X(\theta) &= \lambda M'_X(\theta)\left[\lambda\left(M'_U(\theta)\right)^2 + M''_U(\theta)\right] \\ &\quad + \lambda M_X(\theta)\left[2\lambda M'_U(\theta)M''_U(\theta) + M'''_U(\theta)\right] \\ &= \lambda^2 M_X(\theta)M'_U(\theta)\left[\lambda\left(M'_U(\theta)\right)^2 + M''_U(\theta)\right] \\ &\quad + \lambda M_X(\theta)\left[2\lambda M'_U(\theta)M''_U(\theta) + M'''_U(\theta)\right] \\ &= \lambda M_X(\theta)\left[\lambda^2\left(M'_U(\theta)\right)^3 + 3\lambda M'_U(\theta)M''_U(\theta) + M'''_U(\theta)\right] \qquad \text{(B.10)} \\ M'''_X(0) &= \lambda\left[\lambda^2\left(\mathbb{E}[U]\right)^3 + 3\lambda\,\mathbb{E}[U]\,\mathbb{E}\left[U^2\right] + \mathbb{E}\left[U^3\right]\right] \qquad \text{(B.11)} \end{aligned}$$

and the third moment about the mean is

$$\begin{aligned} \mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^3\right] &= M'''_X(0) - 3M''_X(0)M'_X(0) + 2\left(M'_X(0)\right)^3 \\ &= \lambda\left[\lambda^2\left(\mathbb{E}[U]\right)^3 + 3\lambda\,\mathbb{E}[U]\,\mathbb{E}\left[U^2\right] + \mathbb{E}\left[U^3\right]\right] \\ &\quad - 3\lambda\left[\lambda\left(\mathbb{E}[U]\right)^2 + \mathbb{E}\left[U^2\right]\right]\lambda\,\mathbb{E}[U] + 2\left(\lambda\,\mathbb{E}[U]\right)^3 \\ &= \lambda\,\mathbb{E}\left[U^3\right] \ . \qquad \text{(B.12)} \end{aligned}$$

B.2 Tweedie Dispersion Exponential Family

Let $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$ and have p.d.f.

$$p_X(x) = \begin{cases} \delta(x)\mathrm{e}^{-\lambda} & \text{for } x = 0 \\ \dfrac{\mathrm{e}^{-x\beta-\lambda}}{x} \displaystyle\sum_{y=1}^{\infty} \frac{\beta^{y\alpha}}{\Gamma(y\alpha)} x^{y\alpha} \frac{\lambda^y}{y!} & \text{for } x > 0 \end{cases} . \tag{B.13}$$

Jørgensen (1987) uses the following parametrisation: $p = \dfrac{2+\alpha}{1+\alpha}$, $\mu = \dfrac{\lambda\alpha}{\beta}$, $\phi = \dfrac{\alpha+1}{\beta^{2-p}(\lambda\alpha)^{p-1}}$. The parameters can be rearranged

$$\lambda = \frac{\mu^{2-p}}{\phi(2-p)} \tag{B.14}$$

$$\alpha = \frac{2-p}{p-1} \tag{B.15}$$

$$\beta = \frac{1}{\phi(p-1)\mu^{p-1}} . \tag{B.16}$$

Substituting these parameters into Equation (B.13), the p.m.f. at zero is

$$\mathbb{P}(X = 0) = \exp\left[-\frac{\mu^{2-p}}{\phi(2-p)}\right] \tag{B.17}$$

and the p.d.f. for $x > 0$ is

$$\begin{aligned} p_X(x) &= \exp\left[\frac{-x}{\phi(p-1)\mu^{p-1}} - \frac{\mu^{2-p}}{\phi(2-p)}\right]\frac{1}{x} \\ &\quad \sum_{y=1}^{\infty}\left[\frac{1}{\phi(p-1)\mu^{p-1}}\right]^{y\alpha} \frac{1}{\Gamma(y\alpha)} x^{y\alpha} \left[\frac{\mu^{2-p}}{\phi(2-p)}\right]^{y} \frac{1}{y!} \\ &= \exp\left[\frac{1}{\phi}\left(x\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right]\frac{1}{x} \\ &\quad \sum_{y=1}^{\infty} \frac{x^{y\alpha}\mu^{y[2-p-\alpha(p-1)]}}{\phi^{y(1+\alpha)}(p-1)^{y\alpha}(2-p)^y y!\Gamma(y\alpha)} . \end{aligned}$$

It should be noted that $2 - p - \alpha(p-1) = 2 - p - \frac{2-p}{p-1}(p-1) = 0$ so that

$$p_X(x) = \exp\left[\frac{1}{\phi}\left(x\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right]\frac{1}{x}\sum_{y=1}^{\infty} W_y(x, p, \phi) \tag{B.18}$$

where

$$W_y(x,p,\phi) = \frac{x^{y\alpha}}{\phi^{y(1+\alpha)}(p-1)^{y\alpha}(2-p)^y y!\Gamma(y\alpha)} \ . \tag{B.19}$$

This is in the form of a distribution in the exponential family

$$p_X(x) = \frac{\exp(x\theta)\, g(x)}{Z(\theta)} \tag{B.20}$$

where $Z(\theta)$ is the partition function and θ is the natural parameter. The partition function has some useful properties. It is a normalisation constant $Z = Z(\theta) = \int_0^\infty \exp(x\theta)g(x)\mathrm{d}x$. The derivative of the log-partition function is

$$\begin{aligned}
\frac{\partial \ln Z}{\partial \theta} &= \frac{1}{Z}\frac{\partial Z}{\partial \theta} \\
&= \frac{1}{Z}\frac{\partial}{\partial \theta}\int_0^\infty \exp(x\theta)g(x)\mathrm{d}x \\
&= \frac{1}{Z}\int_0^\infty x\exp(x\theta)g(x)\mathrm{d}x
\end{aligned}$$

which results in

$$\mathbb{E}[X] = \frac{\partial \ln Z}{\partial \theta} \ . \tag{B.21}$$

The second-order derivative of the log partition function is

$$\begin{aligned}
\frac{\partial^2 \ln Z}{\partial \theta^2} &= \frac{Z\frac{\partial^2 Z}{\partial \theta^2} - \left(\frac{\partial Z}{\partial \theta}\right)^2}{Z^2} \\
&= \frac{1}{Z}\frac{\partial^2 Z}{\partial \theta^2} - \left(\frac{1}{Z}\frac{\partial Z}{\partial \theta}\right)^2 \\
&= \frac{1}{Z}\frac{\partial^2}{\partial \theta^2}\int_0^\infty \exp(x\theta)g(x)\mathrm{d}x - \left(\mathbb{E}[X]\right)^2 \\
&= \frac{1}{Z}\int_0^\infty x^2\exp(x\theta)g(x)\mathrm{d}x - \left(\mathbb{E}[X]\right)^2 \\
&= \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2
\end{aligned}$$

thus

$$\mathbb{V}\mathrm{ar}[X] = \frac{\partial^2 \ln Z}{\partial \theta^2} \ . \tag{B.22}$$

For the compound Poisson-gamma distribution, the natural parameter is

$$\theta = \theta(\mu) = \frac{\mu^{1-p}}{\phi(1-p)} \tag{B.23}$$

and the partition function is

$$Z(\theta) = \exp\left[\frac{\mu^{2-p}}{\phi(2-p)}\right] .$$

Rearranging the natural parameter $\mu = (\theta\phi(1-p))^{1/(1-p)}$, the partition function is

$$Z = \exp\left[\phi^{\frac{1}{1-p}} \cdot \theta^{\frac{2-p}{1-p}} \cdot \frac{(1-p)^{\frac{2-p}{1-p}}}{2-p}\right] . \tag{B.24}$$

The log-partition function is

$$\ln Z = \phi^{\frac{1}{1-p}} \cdot \theta^{\frac{2-p}{1-p}} \cdot \frac{(1-p)^{\frac{2-p}{1-p}}}{2-p} . \tag{B.25}$$

Taking the first-order derivative

$$\begin{aligned}\frac{\partial \ln Z}{\partial \theta} &= \phi^{\frac{1}{1-p}} \cdot \frac{2-p}{1-p} \cdot \theta^{\frac{2-p}{1-p}-1} \cdot \frac{(1-p)^{\frac{2-p}{1-p}}}{2-p} \\ &= \phi^{\frac{1}{1-p}} \cdot \theta^{\frac{1}{1-p}} \cdot (1-p)^{\frac{1}{1-p}} .\end{aligned} \tag{B.26}$$

Taking the second-order derivative

$$\begin{aligned}\frac{\partial^2 \ln Z}{\partial \theta^2} &= \phi^{\frac{1}{1-p}} \cdot \frac{1}{1-p} \cdot \theta^{\frac{1}{1-p}-1} \cdot (1-p)^{\frac{1}{1-p}} \\ &= \phi^{\frac{1}{1-p}} \cdot \theta^{\frac{p}{1-p}} \cdot (1-p)^{\frac{p}{1-p}} .\end{aligned} \tag{B.27}$$

Substitute in Equation (B.23)

$$\begin{aligned}\frac{\partial^2 \ln Z}{\partial \theta^2} &= \phi^{\frac{1}{1-p}} \left[\frac{\mu^{1-p}}{\phi(1-p)}\right]^{\frac{p}{1-p}} (1-p)^{\frac{p}{1-p}} \\ &= \phi\mu^p\end{aligned}$$

and using the result from Equation (B.22)

$$\mathbb{V}\mathrm{ar}[X] = \phi\mu^p \tag{B.28}$$

shows that the compound Poisson-gamma distribution is in the Tweedie dispersion exponential family with $1 < p < 2$.

B.3 Method of Moments

Let $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$. Suppose $\widehat{\mu}_1$ is an estimator of $\mu_1 = \mathbb{E}[X]$ and $\widehat{\mu}_j$ is an estimator of $\mu_j = \mathbb{E}\left[(X - \mathbb{E}[X])^j\right]$ for $j = 2, 3$. It is given (see Section 4.1) that

$$\mu_1 = \frac{\alpha\lambda}{\beta} \tag{B.29}$$

$$\mu_2 = \frac{\alpha(\alpha+1)\lambda}{\beta^2} \tag{B.30}$$

$$\mu_3 = \frac{\alpha(\alpha+1)(\alpha+2)\lambda}{\beta^3} \ . \tag{B.31}$$

The ratios μ_2/μ_1 and μ_3/μ_2 are

$$\frac{\mu_2}{\mu_1} = \frac{\alpha+1}{\beta} \tag{B.32}$$

and

$$\frac{\mu_3}{\mu_2} = \frac{\alpha+2}{\beta} \ . \tag{B.33}$$

Subtracting μ_3/μ_2 from μ_2/μ_1 obtains

$$\frac{\mu_3}{\mu_2} - \frac{\mu_2}{\mu_1} = \frac{1}{\beta}$$

and rearranging

$$\beta = \frac{1}{\frac{\mu_3}{\mu_2} - \frac{\mu_2}{\mu_1}}$$

$$\beta = \frac{\mu_1\mu_2}{\mu_1\mu_3 - \mu_2^2} \tag{B.34}$$

gets a justification for the estimator

$$\widehat{\beta} = \frac{\widehat{\mu}_1\widehat{\mu}_2}{\widehat{\mu}_1\widehat{\mu}_3 - \widehat{\mu}_2^2} \ . \tag{B.35}$$

Rearranging Equation (B.32)

$$\alpha = \frac{\mu_2}{\mu_1}\beta - 1 \tag{B.36}$$

and substituting in Equation (B.34)

$$\begin{aligned}\alpha &= \frac{\mu_2}{\mu_1} \times \frac{\mu_1\mu_2}{\mu_1\mu_3 - \mu_2^2} - 1 \\ &= \frac{\mu_2^2 - \mu_1\mu_3 + \mu_2^2}{\mu_1\mu_3 - \mu_2^2} \\ &= \frac{2\mu_2^2 - \mu_1\mu_3}{\mu_1\mu_3 - \mu_2^2} \end{aligned} \tag{B.37}$$

obtains the estimator

$$\widehat{\alpha} = \frac{2\widehat{\mu}_2^2 - \widehat{\mu}_1\widehat{\mu}_3}{\widehat{\mu}_1\widehat{\mu}_3 - \widehat{\mu}_2^2} \ . \tag{B.38}$$

Finally, rearranging Equation (B.29)

$$\lambda = \frac{\beta\mu_1}{\alpha}$$

and substituting in Equations (B.34) and (B.37)

$$\begin{aligned}\lambda &= \frac{\mu_1\mu_2}{\mu_1\mu_3 - \mu_2^2} \times \frac{\mu_1\mu_3 - \mu_2^2}{2\mu_2^2 - \mu_1\mu_3} \times \mu_1 \\ &= \frac{\mu_1^2\mu_2}{2\mu_2^2 - \mu_1\mu_3} \end{aligned} \tag{B.39}$$

which leads to

$$\widehat{\lambda} = \frac{\widehat{\mu}_1^2\widehat{\mu}_2}{2\widehat{\mu}_2^2 - \widehat{\mu}_1\widehat{\mu}_3} \ . \tag{B.40}$$

B.4 Normal Approximation

Let $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$ with m.g.f. $M_X(\theta) = \exp\left[\lambda\left(\left(\frac{\beta}{\beta-\theta}\right)^{\alpha} - 1\right)\right]$. The m.g.f. in this form is not useful because when considering $\lambda \to \infty$ or $\alpha \to \infty$ then $M_X(\theta) \to \infty$. Also for $\beta \to \infty$, then $M_X(\theta) \to 1$ which is not useful either.

The compound Poisson-gamma random variable X can be standardised to obtain useful limiting results from the m.g.f. Let

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}\text{ar}[X]}} \tag{B.41}$$

then in the form of $Z = bX + a$,

$$b = \frac{\beta}{\sqrt{\alpha(\alpha+1)\lambda}} \tag{B.42}$$

and

$$a = -\sqrt{\frac{\alpha\lambda}{\alpha+1}} \ . \tag{B.43}$$

The m.g.f. of Z is

$$\begin{aligned} M_Z(\theta) &= \mathbb{E}\left[\mathrm{e}^{Z\theta}\right] \\ &= \mathbb{E}\left[\mathrm{e}^{(bX+a)\theta}\right] \\ &= \mathrm{e}^{a\theta} M_X(b\theta) \ . \end{aligned} \tag{B.44}$$

Substituting in a and b

$$\begin{aligned} M_Z(\theta) &= \exp\left(-\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}}\right)\exp\left[\lambda\left(\left(\frac{\beta}{\beta - \frac{\beta\theta}{\sqrt{\alpha(\alpha+1)\lambda}}}\right)^{\alpha} - 1\right)\right] \\ &= \exp\left(-\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}}\right)\exp\left[\lambda\left(\left(\frac{\sqrt{\alpha(\alpha+1)\lambda}}{\sqrt{\alpha(\alpha+1)\lambda} - \theta}\right)^{\alpha} - 1\right)\right] \\ &= \exp\left(-\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}}\right)\exp\left[\lambda\left(\left(1 - \frac{\theta}{\sqrt{\alpha(\alpha+1)\lambda}}\right)^{-\alpha} - 1\right)\right] \ . \end{aligned} \tag{B.45}$$

Using the binomial expansion

$$\left(1 - \frac{\theta}{\sqrt{\alpha(\alpha+1)\lambda}}\right)^{-\alpha} = 1 + \sum_{r=1}^{\infty} \frac{\theta^r \prod_{s=1}^{r}(\alpha+s-1)}{(\alpha(\alpha+1)\lambda)^{r/2} r!} \quad \text{for } \frac{\|\theta\|}{\sqrt{\alpha(\alpha+1)\lambda}} < 1 \ . \tag{B.46}$$

obtains

$$M_Z(\theta) = \exp\left(-\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}}\right)\exp\left[\lambda\sum_{r=1}^{\infty}\frac{\theta^r \prod_{s=1}^{r}(\alpha+s-1)}{(\alpha(\alpha+1)\lambda)^{r/2} r!}\right] \ . \tag{B.47}$$

Writing in full the $r = 1, 2$ terms

$$M_Z(\theta) = \exp\left(-\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}}\right)\exp\left[\theta\sqrt{\frac{\alpha\lambda}{\alpha+1}} + \frac{\theta^2}{2} + \lambda\sum_{r=3}^{\infty}\frac{\theta^r \prod_{s=1}^{r}(\alpha+s-1)}{(\alpha(\alpha+1)\lambda)^{r/2} r!}\right]$$

and a term cancels out

$$M_Z(\theta) = \exp\left[\frac{\theta^2}{2} + \sum_{r=3}^{\infty} \frac{\theta^r \prod_{s=1}^{r}(\alpha+s-1)}{(\alpha(\alpha+1))^{r/2} r!} \lambda^{1-r/2}\right] . \tag{B.48}$$

For large λ

$$\lim_{\lambda\to\infty} M_Z(\theta) = \exp\left[\frac{\theta^2}{2}\right] \tag{B.49}$$

which is the same as the m.g.f. of a standard Normal distribution, therefore

$$\lim_{\lambda\to\infty} Z \sim \mathrm{N}(0,1) . \tag{B.50}$$

This should make sense as for high λ, the Poisson random variable has a high expectation, increasing the number of gamma random variables in a summation. Increasing the number of terms in a summation will trigger the central limit theorem.

For high α

$$\begin{aligned}\lim_{\alpha\to\infty} \frac{\prod_{s=1}^{r}(\alpha+s-1)}{(\alpha(\alpha+1))^{r/2}} &= \lim_{\alpha\to\infty} \frac{\prod_{s=1}^{r}\alpha}{\alpha^r} \\ &= 1 .\end{aligned} \tag{B.51}$$

As a result

$$\lim_{\alpha\to\infty} M_Z(\theta) = \exp\left[\frac{\theta^2}{2} + \sum_{r=3}^{\infty} \frac{\theta^r}{r!} \lambda^{1-r/2}\right] . \tag{B.52}$$

This shows that taking the limit $\alpha \to \infty$ is not enough to get a Normal limiting distribution. The limit must be accompanied with the limit $\lambda \to \infty$ to obtain

$$\lim_{\lambda\to\infty}\lim_{\alpha\to\infty} Z \sim \mathrm{N}(0,1) . \tag{B.53}$$

It should be noted that $M_Z(\theta)$ is independent of β. Thus β does not affect the limiting distribution.

The above results justify the use of the approximation

$$X \sim \mathrm{N}\left(\frac{\lambda\alpha}{\beta}, \frac{\lambda\alpha(\alpha+1)}{\beta^2}\right) \tag{B.54}$$

for large λ. The limiting case where $\lambda \to 0$, $\alpha \to 0$ and $\beta \to 0$ will not be discussed here.

B.5 Saddlepoint Approximation

Let $X \sim \text{CP}\Gamma(\lambda, \alpha, \beta)$ with m.g.f. $M_X(\theta) = \exp\left[\lambda\left(\left(\frac{\beta}{\beta-\theta}\right)^{\alpha} - 1\right)\right]$. Let the cumulant generating function of X be $K_X(\theta) = \ln M_X(\theta)$. The saddlepoint approximation is

$$p_X(x) \approx \left(2\pi K_X''(s)\right)^{-1/2} \exp\left[K_X(s) - sx\right] \tag{B.55}$$

where $s = s(x)$ is the solution to the saddle point equation $K_X'(s) = x$.

The cumulant generating function is

$$K_X(\theta) = \lambda\left[\left(\frac{\beta}{\beta-\theta}\right)^{\alpha} - 1\right] . \tag{B.56}$$

Taking the derivative with respect to θ

$$K_X'(\theta) = \frac{\lambda\alpha\beta^{\alpha}}{(\beta-\theta)^{\alpha+1}} \tag{B.57}$$

and this is known as the saddlepoint equation. The quantity $s = s(x)$ is the solution to the equation $K_X'(s) = x$, that is

$$\frac{\lambda\alpha\beta^{\alpha}}{(\beta-s)^{\alpha+1}} = x$$

with solution

$$s = \beta - \left(\frac{\lambda\alpha\beta^{\alpha}}{x}\right)^{\frac{1}{\alpha+1}} . \tag{B.58}$$

The second-order derivative of the cumulant generating function is

$$K_X''(\theta) = \frac{\lambda\alpha(\alpha+1)\beta^{\alpha}}{(\beta-\theta)^{\alpha+2}} . \tag{B.59}$$

Substituting Equations (B.56) and (B.59) into Equation (B.55)

$$p_X(x) \approx \frac{1}{\sqrt{2\pi}}\left[\frac{(\beta-s)^{\alpha+2}}{\lambda\alpha(\alpha+1)\beta^{\alpha}}\right]^{1/2} \exp\left[\lambda\left(\left(\frac{\beta}{\beta-s}\right)^{\alpha} - 1\right) - sx\right] .$$

Substituting in Equation (B.58)

$$p_X(x) \approx \frac{1}{\sqrt{2\pi}}\left[\frac{\left(\beta - \beta + \left(\frac{\lambda\alpha\beta^{\alpha}}{x}\right)^{\frac{1}{\alpha+1}}\right)^{\alpha+2}}{\lambda\alpha(\alpha+1)\beta^{\alpha}}\right]^{1/2}$$
$$\exp\left[\lambda\left(\left(\frac{\beta}{\beta - \beta + \left(\frac{\lambda\alpha\beta^{\alpha}}{x}\right)^{\frac{1}{\alpha+1}}}\right)^{\alpha} - 1\right) - x\left(\beta - \left(\frac{\lambda\alpha\beta^{\alpha}}{x}\right)^{\frac{1}{\alpha+1}}\right)\right]$$

simplifying further

$$
\begin{aligned}
p_X(x) &\approx \frac{1}{\sqrt{2\pi}}\left[\frac{\left(\frac{\lambda\alpha\beta^\alpha}{x}\right)^{\frac{\alpha+2}{\alpha+1}}}{\lambda\alpha(\alpha+1)\beta^\alpha}\right]^{1/2} \\
&\qquad \exp\left[\lambda\left(\beta^\alpha\left(\frac{x}{\lambda\alpha\beta^\alpha}\right)^{\frac{\alpha}{\alpha+1}}-1\right)-x\beta+\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{\alpha+1}}x^{1-\frac{1}{\alpha+1}}\right] \\
&\approx \frac{1}{\sqrt{2\pi(\alpha+1)}}x^{-\frac{\alpha+2}{2(\alpha+1)}}\left(\lambda\alpha\beta^\alpha\right)^{\left(\frac{\alpha+2}{\alpha+1}-1\right)/2} \\
&\qquad \exp\left[\lambda\left(\beta^\alpha\left(\frac{x}{\lambda\alpha\beta^\alpha}\right)^{\frac{\alpha}{\alpha+1}}-1\right)-x\beta+\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{\alpha+1}}x^{\frac{\alpha}{\alpha+1}}\right] \\
&\approx \frac{\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{2(\alpha+1)}}}{\sqrt{2\pi(\alpha+1)}}x^{-\frac{\alpha+2}{2(\alpha+1)}} \\
&\qquad \exp\left[-\lambda-x\beta+x^{\frac{\alpha}{\alpha+1}}\left(\frac{\lambda\beta^\alpha}{\left(\lambda\alpha\beta^\alpha\right)^{\frac{\alpha}{\alpha+1}}}+\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{\alpha+1}}\right)\right] \\
&\approx \frac{\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{2(\alpha+1)}}\mathrm{e}^{-\lambda}}{\sqrt{2\pi(\alpha+1)}}x^{-\frac{\alpha+2}{2(\alpha+1)}}\mathrm{e}^{-x\beta} \\
&\qquad \exp\left[x^{\frac{\alpha}{\alpha+1}}\left(\left(\lambda\beta^\alpha\right)^{1-\frac{\alpha}{\alpha+1}}\alpha^{-\frac{\alpha}{\alpha+1}}+\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{\alpha+1}}\right)\right] \\
&\approx \frac{\left(\lambda\alpha\beta^\alpha\right)^{\frac{1}{2(\alpha+1)}}\mathrm{e}^{-\lambda}}{\sqrt{2\pi(\alpha+1)}}x^{-\frac{\alpha+2}{2(\alpha+1)}}\mathrm{e}^{-x\beta} \\
&\qquad \exp\left[x^{\frac{\alpha}{\alpha+1}}\left(\lambda\beta^\alpha\right)^{\frac{1}{\alpha+1}}\left(\alpha^{-\frac{\alpha}{\alpha+1}}+\alpha^{\frac{1}{\alpha+1}}\right)\right] .
\end{aligned}
$$

The expression $\alpha^{-\frac{\alpha}{\alpha+1}}+\alpha^{\frac{1}{\alpha+1}}$ can be simplified by putting the two terms over a common denominator

$$
\begin{aligned}
\alpha^{-\frac{\alpha}{\alpha+1}}+\alpha^{\frac{1}{\alpha+1}} &= \alpha^{\frac{1}{\alpha+1}}+\frac{1}{\alpha^{\frac{\alpha}{\alpha+1}}} \\
&= \frac{\alpha^{\frac{1}{\alpha+1}}\alpha^{\frac{\alpha}{\alpha+1}}+1}{\alpha^{\frac{\alpha}{\alpha+1}}} \\
&= \frac{\alpha+1}{\alpha^{\frac{\alpha}{\alpha+1}}}
\end{aligned}
\tag{B.60}
$$

so that the saddlepoint approximation is

$$p_X(x) \approx \frac{(\lambda\alpha\beta^\alpha)^{\frac{1}{2(\alpha+1)}}\,\mathrm{e}^{-\lambda}}{\sqrt{2\pi(\alpha+1)}} x^{-\frac{\alpha+2}{2(\alpha+1)}} \mathrm{e}^{-x\beta} \exp\left[x^{\frac{\alpha}{\alpha+1}} \frac{(\lambda\beta^\alpha)^{\frac{1}{\alpha+1}}(\alpha+1)}{\alpha^{\frac{\alpha}{\alpha+1}}} \right] . \quad \text{(B.61)}$$

B.6 Series Evaluation

Let $X \sim \mathrm{CP}\Gamma(\lambda, \alpha, \beta)$ with p.d.f.

$$p_X(x) = \begin{cases} \delta(x) \exp\left[-\dfrac{\mu^{2-p}}{\phi(2-p)} \right] & \text{for } x = 0 \\ \exp\left[\dfrac{1}{\phi}\left(x\dfrac{\mu^{1-p}}{1-p} - \dfrac{\mu^{2-p}}{2-p} \right) \right] \dfrac{1}{x} \displaystyle\sum_{y=1}^{\infty} W_y(x,p,\phi) & \text{for } x > 0 \end{cases}$$

where $p = \dfrac{2+\alpha}{1+\alpha}$, $\mu = \dfrac{\lambda\alpha}{\beta}$, $\phi = \dfrac{\alpha+1}{\beta^{2-p}(\lambda\alpha)^{p-1}}$ and

$$W_y = W_y(x,p,\phi) = \frac{x^{y\alpha}}{\phi^{y(1+\alpha)}(p-1)^{y\alpha}(2-p)^y y!\Gamma(y\alpha)} . \quad \text{(B.62)}$$

Dunn and Smyth (2005) approximated the sum by truncation

$$\sum_{y=1}^{\infty} W_y \approx \sum_{y=y_{\mathrm{l}}}^{y_{\mathrm{u}}} W_y \quad \text{(B.63)}$$

where $y_{\mathrm{l}} < y_{\max} < y_{\mathrm{u}}$ and $y_{\max}$ is the value of y which maximises W_y. Dunn and Smyth (2005) treated W_y as a continuous and differentiable function of y.

It is easier to differentiate $\ln(W_y)$ where

$$\begin{aligned} \ln(W_y) &= y\alpha\ln(x) - y(1+\alpha)\ln(\phi) - y\alpha\ln(p-1) \\ &\qquad - y\ln(2-p) - \ln(y!) - \ln\Gamma(y\alpha) \\ &= y\ln\left(\frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)} \right) - \ln(y!) - \ln\Gamma(y\alpha) . \end{aligned} \quad \text{(B.64)}$$

Using Stirling's approximation $\ln(n!) \approx \ln\Gamma(n) \approx n\ln(n!) - n$

$$\ln(W_y) \approx y\ln\left(\frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)} \right) - y\ln(y) + y - y\alpha\ln(y\alpha) + y\alpha . \quad \text{(B.65)}$$

Taking the derivative with respect to y

$$\begin{aligned}\frac{\partial \ln(W_y)}{\partial y} &\approx \ln\left(\frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)}\right) - \ln(y) - 1 + 1 \\ &\qquad - \alpha\ln(y\alpha) - \alpha + \alpha \\ &\approx \ln\left(\frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)}\right) - \ln(y) - \alpha\ln(y\alpha)\ . \end{aligned} \tag{B.66}$$

Setting the derivative to zero

$$\begin{aligned} 0 &\approx \ln\left(\frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)y_{\max}^{1+\alpha}\alpha^\alpha}\right) \\ 1 &\approx \frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)y_{\max}^{1+\alpha}\alpha^\alpha} \\ y_{\max}^{1+\alpha} &\approx \frac{x^\alpha}{\phi^{1+\alpha}(p-1)^\alpha(2-p)\alpha^\alpha} \\ y_{\max} &\approx \frac{1}{\phi}\left(\frac{x}{(p-1)\alpha}\right)^{\frac{\alpha}{1+\alpha}}(2-p)^{\frac{-1}{1+\alpha}}\ . \end{aligned}$$

This can be simplified using the fact that $\alpha = \frac{2-p}{p-1}$, $\frac{1}{1+\alpha} = p-1$ and $\frac{\alpha}{1+\alpha} = 2-p$ then

$$y_{\max} \approx \frac{1}{\phi}\left(\frac{x}{2-p}\right)^{2-p}(2-p)^{1-p}$$

and finally

$$y_{\max} \approx \frac{x^{2-p}}{\phi(2-p)}\ . \tag{B.67}$$

To verify that $y_{\max}$ is a maximum, the second-order derivative can be investigated

$$\frac{\partial^2 \ln(W_y)}{\partial y^2} \approx -\frac{1}{y}(\alpha+1) \tag{B.68}$$

to see that

$$\frac{\partial^2 \ln(W_y)}{\partial y^2} < 0 \quad \text{for } y = 1, 2, 3, \ldots \tag{B.69}$$

therefore $y_{\max}$ is a maximum.

Appendix C

Expectation of $\widehat{\sigma}_0^2$

It can be shown that the null variance estimator $\widehat{\sigma}_0^2$ is approximately unbiased when estimating using n samples of $\mathrm{N}(\mu_0, \sigma_0^2)$ random variables. Suppose the random variables are $Z_1, Z_2, \ldots, Z_n$ and are i.i.d. with probability density function $f(z)$. Recall that

$$\widehat{\sigma}_0 = \left[-\frac{\partial^2}{\partial z^2} \ln \widehat{p}(z) \Big|_{z=\widehat{\mu}_0} \right]^{-1/2} \tag{C.1}$$

where

$$\begin{aligned} \frac{\partial^2}{\partial z^2} \ln \widehat{p}_Z(z) = \left[h \sum_{i=1}^{n} \phi\left(\frac{Z_i - z}{h} \right) \right]^{-2} \times \left\{ \left[\sum_{i=1}^{n} \phi\left(\frac{Z_i - z}{h} \right) \right] \right. \\ \times \left[\sum_{i=1}^{n} \phi\left(\frac{Z_i - z}{h} \right) \left(\left(\frac{Z_i - z}{h} \right)^2 - 1 \right) \right] - \left[\sum_{i=1}^{n} \phi\left(\frac{Z_i - z}{h} \right) \right. \\ \left. \left. \left(\frac{Z_i - z}{h} \right) \right]^2 \right\} . \end{aligned} \tag{C.2}$$

Apply an approximation such that

$$\begin{aligned} \mathbb{E}\left[-\widehat{\sigma}_0^{-2} \right] \approx \left[h \sum_{i=1}^{n} \mathbb{E}\left[\phi\left(\frac{Z_i - z}{h} \right) \right] \right]^{-2} \times \left\{ \left[\sum_{i=1}^{n} \mathbb{E}\left[\phi\left(\frac{Z_i - z}{h} \right) \right] \right] \right. \\ \times \left[\sum_{i=1}^{n} \mathbb{E}\left[\phi\left(\frac{Z_i - z}{h} \right) \left(\left(\frac{Z_i - z}{h} \right)^2 - 1 \right) \right] \right] - \left[\sum_{i=1}^{n} \mathbb{E}\left[\phi\left(\frac{Z_i - z}{h} \right) \right. \right. \\ \left. \left. \left. \left(\frac{Z_i - z}{h} \right) \right] \right]^2 \right\} . \end{aligned} \tag{C.3}$$

The following function of Z was studied

$$g_t(Z) = \phi\left(\frac{Z - z_0}{h}\right)\left(\frac{Z - z_0}{h}\right)^t \ . \tag{C.4}$$

The expectation is

$$\mathbb{E}\left[g_t(Z)\right] = \int_{-\infty}^{\infty} \phi\left(\frac{z - z_0}{h}\right)\left(\frac{z - z_0}{h}\right)^t f(z)\mathrm{d}z \ . \tag{C.5}$$

By substituting $u = (z - z_0)/h$ then

$$\mathbb{E}\left[g_t(Z)\right] = \int_{-\infty}^{\infty} h\phi(u)u^t f(z_0 + uh)\mathrm{d}u \tag{C.6}$$

which a Taylor series can be used to expand $f(z_0 + uh)$

$$\mathbb{E}\left[g_t(Z)\right] = \int_{-\infty}^{\infty} h\phi(u)u^t \sum_{r=0}^{\infty} \frac{f^{(r)}(z_0)}{r!}(uh)^r \mathrm{d}u \tag{C.7}$$

to get

$$\mathbb{E}\left[g_t(Z)\right] = \sum_{r=0}^{\infty} \frac{f^{(r)}(z_0)h^{r+1}}{r!} M^{(t+r)}(0) \tag{C.8}$$

where $M(\theta) = \int_{-\infty}^{\infty} \mathrm{e}^{\theta u}\phi(u)\mathrm{d}u$ is the moment generating function of the standard Normal distribution. It is left as an exercise to show that

$$M(0) = 1 \tag{C.9}$$
$$M^{(1)}(0) = 0 \tag{C.10}$$
$$M^{(2)}(0) = 1 \tag{C.11}$$
$$M^{(3)}(0) = 0 \tag{C.12}$$
$$M^{(4)}(0) = 3 \tag{C.13}$$
$$M^{(5)}(0) = 0 \tag{C.14}$$
$$M^{(6)}(0) = 15 \ . \tag{C.15}$$

Then for $t = 0, 1, 2$

$$\mathbb{E}\left[g_0(Z)\right] = f(z_0)h + \frac{f^{(2)}(z_0)h^3}{2} + \frac{3f^{(4)}(z_0)h^5}{4!} + O(h^7) \tag{C.16}$$
$$\mathbb{E}\left[g_1(Z)\right] = f^{(1)}(z_0)h^2 + \frac{3f^{(3)}(z_0)h^4}{3!} + O(h^6) \tag{C.17}$$
$$\mathbb{E}\left[g_2(Z)\right] = f(z_0)h + \frac{3f^{(2)}(z_0)h^3}{2} + \frac{15f^{(4)}(z_0)h^5}{4!} + O(h^7) \ . \tag{C.18}$$

The derivatives of $f(z)$ are

$$f^{(1)}(z) = \frac{f(z)}{\sigma_0^2}\left[-(z-\mu_0)\right] \tag{C.19}$$

$$f^{(2)}(z) = \frac{f(z)}{\sigma_0^2}\left[\frac{(z-\mu_0)^2}{\sigma_0^2} - 1\right] \tag{C.20}$$

$$f^{(3)}(z) = \frac{f(z)}{\sigma_0^4}\left[3(z-\mu_0) - \frac{(z-\mu_0)^3}{\sigma_0^2}\right] \tag{C.21}$$

$$f^{(4)}(z) = \frac{f(z)}{\sigma_0^4}\left[3 - \frac{6(z-\mu_0)^2}{\sigma_0^2} + \frac{(z-\mu_0)^4}{\sigma_0^4}\right] . \tag{C.22}$$

Then substituting these into Equations (C.16), (C.17) and (C.18) obtains

$$\begin{aligned}\mathbb{E}\left[g_0(Z)\right] = f(z_0)h + \frac{h^3}{2}\frac{f(z_0)}{\sigma_0^2}\left[\frac{(z_0-\mu_0)^2}{\sigma_0^2} - 1\right] \\ + \frac{3h^5}{4!}\frac{f(z_0)}{\sigma_0^4}\left[3 - \frac{6(z_0-\mu_0)^2}{\sigma_0^2} + \frac{(z_0-\mu_0)^4}{\sigma_0^4}\right] + O(h^7)\end{aligned} \tag{C.23}$$

$$\begin{aligned}\mathbb{E}\left[g_1(Z)\right] = -h^2\frac{f(z_0)}{\sigma_0^2}(z_0-\mu_0) + \frac{h^4}{2}\frac{f(z_0)}{\sigma_0^4}\left[3(z_0-\mu_0) - \frac{(z_0-\mu_0)^3}{\sigma_0^2}\right] \\ + O(h^6)\end{aligned} \tag{C.24}$$

$$\begin{aligned}\mathbb{E}\left[g_2(Z)\right] = f(z_0)h + \frac{3h^3}{2}\frac{f(z_0)}{\sigma_0^2}\left[\frac{(z_0-\mu_0)^2}{\sigma_0^2} - 1\right] \\ + \frac{15h^5}{4!}\frac{f(z_0)}{\sigma_0^4}\left[3 - \frac{6(z_0-\mu_0)^2}{\sigma_0^2} + \frac{(z_0-\mu_0)^4}{\sigma_0^4}\right] + O(h^7) .\end{aligned} \tag{C.25}$$

The definition for $\mathbb{E}\left[g_t(Z)\right]$ can be used to simplify Equation (C.3) to

$$\mathbb{E}\left[\widehat{\sigma}_0^2\right] = \frac{-\left(h\,\mathbb{E}\left[g_0(Z)\right]\right)^2}{\mathbb{E}\left[g_0(Z)\right]\left[\mathbb{E}\left[g_2(Z)\right] - \mathbb{E}\left[g_0(Z)\right]\right] - \left(\mathbb{E}\left[g_1(Z)\right]\right)^2} . \tag{C.26}$$

Substituting in Equations (C.23), (C.24) and (C.25) and setting $z_0 = \widehat{\mu}_0$ obtains

$$\mathbb{E}\left[\widehat{\sigma}_0^2\right] = \frac{\sigma_0^2 + h^2\left[\frac{(\widehat{\mu}_0-\mu_0)^2}{\sigma_0^2} - 1\right] + O(h^4)}{1 + h^2\left[\frac{(\widehat{\mu}_0-\mu_0)^2}{\sigma_0^4} - \frac{2}{\sigma_0^2}\right] + O(h^4)} \tag{C.27}$$

where the following may be useful

$$\mathbb{E}\left[g_0(Z)\right]\left[\mathbb{E}\left[g_2(Z)\right]-\mathbb{E}\left[g_0(Z)\right]\right]=h^4\frac{f^2(z_0)}{\sigma_0^2}\left[\frac{(z_0-\mu_0)^2}{\sigma_0^2}-1\right] + h^6\frac{f^2(z_0)}{\sigma_0^4}\left[2-\frac{4(z_0-\mu_0)^2}{\sigma_0^2}+\frac{(z_0-\mu_0)^4}{\sigma_0^4}\right]+O(h^8)\ . \quad \text{(C.28)}$$

The result implies that $\widehat{\sigma}_0^2$ is an unbiased estimator of σ_0^2 up to the first-order approximation. Higher-order terms consist of at least even-order polynomials of $(\widehat{\mu}_0-\mu_0)$ and h. As a result, any bias in $\widehat{\mu}_0$ would contribute to the bias of $\widehat{\sigma}_0^2$. It is not clear if $\widehat{\sigma}_0^2$ is a consistent estimator because h depends on n and the result for $\mathbb{V}\mathrm{ar}\left[\widehat{\sigma}_0^2\right]$ is difficult to obtain.