BAYESIAN INVERSION TECHNIQUES FOR SYNTHETIC APERTURE RADAR IMAGING

A thesis submitted to the University of Manchester
for the degree of Master of Philosophy
in the Faculty of Science and Engineering

2022

Ella M. Cooper
Department of Mathematics
School of Natural Sciences

Contents

Word count 18688

List of Figures

The University of Manchester

Ella M. Cooper
Master of Philosophy
Bayesian Inversion Techniques for Synthetic Aperture Radar Imaging
September 26, 2022

The presence of congestion in the electromagnetic (EM) spectrum can have a detrimental effect on the usability of conventionally formed synthetic aperture radar (SAR) images. This is particularly an issue in scenarios involving interference; or where it is desirable to penetrate walls or foliage by transmitting in the low frequency range (which is a notoriously active part of the EM spectrum). To alleviate the effects of excess noise appearing in the recorded data, Bayesian inference techniques can be used during the image formation process, which incorporate prior knowledge about the scene. The techniques aim to obtain an image which is consistent with both prior knowledge and the recorded data, in addition to retaining information on the certainty of the solution.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. This thesis is subject to Crown Copyright © Dstl 2022 (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and Dstl has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=2442), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/about/regulations) and in The University's Policy on Presentation of Theses.

Acknowledgements

I would like to express my gratitude to my supervisors, Prof Kody Law and Prof William Lionheart, for allowing me the opportunity complete this project and providing guidance to me along the way. I would also like to thank my colleagues, Josh and Declan, who's support has been invaluable.

Chapter 1

Introduction

Radar imaging is a crucial sensing modality in many applications, such as topography, oceanography, geology, forestry, remote sensing, infrastructure monitoring, and environment monitoring. However, it is typically not achievable to yield quality radar images in scenarios involving:

- ★ foliage penetration,
- ★ through wall imaging,
- ★ urban environments,
- ★ congested electromagnetic environments.

A problem common across these scenarios is the presence of excess clutter in the spectrum; arising from other active radars, TV signals, or where the signal has bounced multiple times before returning for example. A comparison of SAR (synthetic aperture radar) images with and without a high level of interference is presented in Figure 1.1. From these images, it can be seen how an image produced in the presence of a congested electromagnetic spectrum conveys significantly less information about a scene.

As part of this thesis, methods have been developed which apply Bayesian techniques at various stages of the image formation process. These techniques have shown promise at reducing the detrimental effects that the presence of congestion in the EM (electromagnetic) spectrum induces in radar imaging. The techniques presented fall into two main strands:

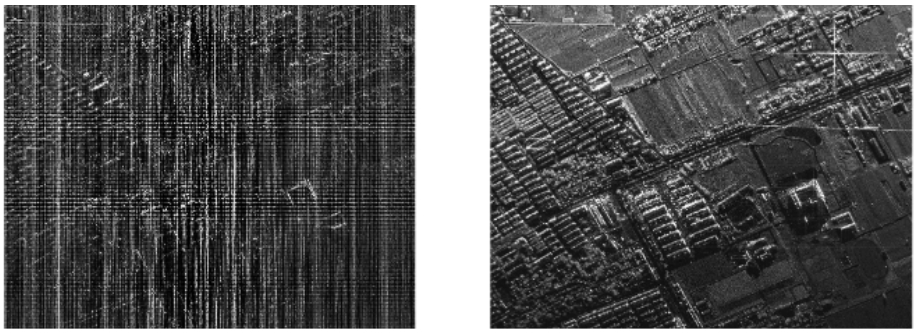

Figure 1.1: An example of images with and without interference. These two images show the same scene, one showing the detrimental impact a congested electromagnetic environment can have on an image. (Image from [1], Copyright © Society for Industrial and Applied Mathematics, reprinted with permission, all rights reserved.)

- ★ development of forming SAR images directly from measured phase history data (i.e. raw data),
- ★ development of improving the usability of poor quality or degraded SAR images, formed via existing techniques, particularly making use of multi-aspect collections.

Contrary to conventional imaging techniques, [8], which are usually back-projection based, the methodology developed here pose the problems within SAR imaging as an inverse problem. An inverse problem refers to the process of estimating data which is not directly measurable itself, but is related in some way to data which is directly measurable. Within radar imaging this translates to wanting information about a scene based on some radar measurements that have been directly measured.

The two strands mentioned have both been explored using Bayesian solving approaches, beginning with using simulated data, and progressing to test against real data sets.

This thesis is comprised of an explanation of the inverse problem relevant to this thesis, then a literature review detailing what Bayesian approaches have been used in the radar imaging (and other imaging) domains, followed by a detailed description of the specific problems that were considered, the methodology which was used is then presented, and finally results using both simulated and real data have been shown.

Chapter 2

Bayesian Inference and Inverse Problems

In this thesis, the problems will all be posed as inverse problems, such as, $\mathbf{y} = A\mathbf{x} + \epsilon$, where the aim is to find $\mathbf{x}$; for this problem to be well posed it should satisfy, as in [57],

- ★ a solution exists,
- ★ the solution is unique,
- ★ a systems behavior changes continuously with the initial conditions.

When any one of these requirements fail, the problem can be considered ill-posed. In the problems arising in this thesis, many are ill-posed.

In the case that the problem is ill-posed, it is not possible to simply use A^{-1} to solve the problem due to the noise term. A solution to this would be to use the least squares solution:

$$\mathbf{x}_{\text{least squares}} = \underset{\mathbf{x}}{\operatorname{argmin}} ||\mathbf{y} - A\mathbf{x}||_2^2 \tag{2.1}$$

However, this does not always produce a unique solution, as stated in theorem 1, [58].

Theorem 1. The least squares problem always has a solution. The solution is unique if and only if the null space of A is trivial.

It is possible to further minimize over the multiple solutions in order to obtain a unique solution; however this will only address existence and uniqueness requirements, not stability.

It is for this reason that it is desirable to regularise the solution, which is done by incorporating some prior knowledge about the solution into the solving. This is done by adding a regularising term onto Equation 2.4 which aims to both stabilize the solution and input relevant knowledge that is known about the solution already, ahead of the solving. A popular regularisation is Tikhonov regularisation, [59]. This has a solution of

$$\mathbf{x}_{\text{Tikhonov}} = \underset{\mathbf{x}}{\operatorname{argmin}} \, ||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda ||H\mathbf{x}||_2^2, \tag{2.2}$$

where λ can be considered a scalar tuning parameter and H represents a choice of behavior which is penalized in the solutions, often $H = I_n$.

This is also known as l_2 regularisation. The first term in 2.2 can be thought of as a data fitting term, minimizing the data fit error. The second term in 2.2 can be thought of as the solution constrainer, minimising the behaviour dictated by H (simply minmising the solution when this is the identity). The tuning parameter λ can be scaled to obtain an optimal balance between these two terms.

It is also important to consider whether the problem is well-conditioned or ill-conditioned, [60]. The condition number of a function (or in the case of this thesis a matrix) indicates how much change in the output can be expected from a change in the input. It is a measure of how sensitive the function is to errors.

Consider the SAR observation matrices from equation 4.11, there are many parameters that determine the radar system as a whole, and these parameters in turn determine how well-conditioned the SAR observation matrix is. For example, take the resolution of the image that is trying to be made, if an inappropriately high resolution is trying to be reconstructed (one which is more than can be physically expected from the data from the radar) then this will make the problem more ill conditioned. Figure 2.1 shows how the condition number varies for different resolutions. It can be seen that as the pixel resolution decreases to roughly equal to the wavelength of the radar; the condition number increases rapidly. However for resolutions multiple times the wavelength, the condition number is significantly lower. All other parameters of the radar were kept constant for the purpose of this comparison.

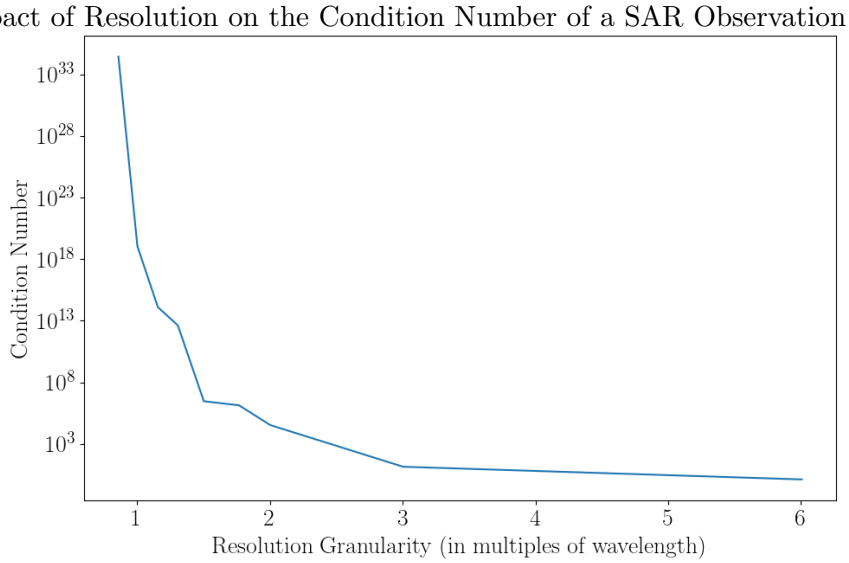

Figure 2.1: Plot displaying the impact desired resolution has on the condition number of a SAR observation matrix. The x axis denotes the size of the pixels, written as a multiple of the wavelength. SAR observation matrices are described more in Section 4.

A similar behavior can be seen when looking at the range the radar system is at from the scene. Figure 2.2 shows the relationship of the distance between the radar and the scene, and the condition number of the SAR observation matrix. Here it can be seen that, as the range is increased the condition number increases. It is well known that the power of a radar wave attenuates with an inverse fourth-route relationship to range [61], so with the addition of the lower SNR achieved at greater distances, the problem would be even harder than this plot displays.

This behavior is what would be expected from the SAR observation matrix, essentially as parameters are selected which result in a harder problem, the more ill-conditioned the problem becomes, (becoming more sensitive to changes or errors in the input). This demonstrates the mathematical difficulty of the types of problem being considered.

In this section, initial solving approaches are presented and interpreted in a Bayesian sense, and explored using simulated data.

2.1 Analytic Solving Approaches

In section 2, two solving approaches to the inverse probelm,

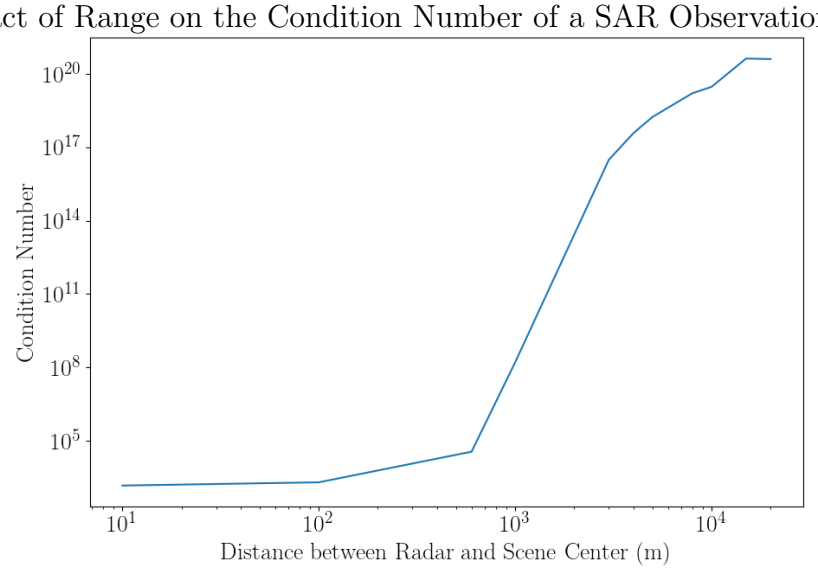

Figure 2.2: Plot displaying the impact range between the scene center and the radar has on the condition number of a SAR observation matrix. SAR observation matrices are described more in Section 4.

$$\mathbf{y} = A\mathbf{x} + \epsilon, \tag{2.3}$$

were mentioned (here again, $\mathbf{y}$ is observed data, A is a forward model, $\mathbf{x}$ is the unknown scene, and ϵ is the noise or error). The two solving approaches being least squares, Equation 2.4 and Tikhonov, Equation 2.2. Here these two solving approaches are explored in a Bayesian setting.

2.1.1 Least Squares

Take the least squares solution,

$$\mathbf{x}_{\text{least squares}} = \underset{\mathbf{x}}{\operatorname{argmin}} \, ||\mathbf{y} - A\mathbf{x}||_2^2 \tag{2.4}$$

From this equation, it is possible find an analytical solution to this minmisation problem. The task is to find the $\mathbf{x}$ which minimises $||\mathbf{y} - A\mathbf{x}||_2^2$. To find this solution, a temporary value W is defined as argument to be minimised (for ease of reading), which by definition of the Euclidean norm is equal to

$$W = ||\mathbf{y} - A\mathbf{x}||_2^2 = (\mathbf{y} - A\mathbf{x})^T(\mathbf{y} - A\mathbf{x}) \tag{2.5}$$

where the T indicates the adjoint, if all values are real then this would be the matrix transpose [1]. Expanding out this expression gives

$$W = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T A\mathbf{x} - \mathbf{x}^T A^T\mathbf{y} + \mathbf{x}^T A^T A\mathbf{x} \tag{2.6}$$

$$= \mathbf{y}^T\mathbf{y} - 2\mathbf{x}^T A^T\mathbf{y} + \mathbf{x}^T A^T A\mathbf{x}. \tag{2.7}$$

And now to find a minimum point of W, it is necessary to first differentiate with respect to $\mathbf{x}$, in order to find a point where the gradient is equal to zero. Hence,

$$\frac{\partial W}{\partial \mathbf{x}} = -2A^T\mathbf{y} + 2A^T A\mathbf{x} \tag{2.8}$$

This expression equals zero (and so is a minimum or maximum point) only when $\mathbf{x}$ satisfies

$$\mathbf{x} = (A^T A)^{-1}A^T\mathbf{y}. \tag{2.9}$$

To determine if this is a minimum or maximum point, the second derivative of W needs to be calculated,

$$\frac{\partial^2 W}{\partial \mathbf{x}^2} = 2A^T A = 2||A||_2^2 > \mathbf{0}. \tag{2.10}$$

If A is not the zero matrix, then this inequality will always hold. $\mathbf{x} = (A^T A)^{-1}A^T\mathbf{y}$ minimises W, and is the least squares solution.

The least squares solution to an inverse problem also has a Bayesian interpretation. To derive this result, first Bayes' Theorem is presented.

Theorem 2. (Bayes' Theorem) Given two events C and B (where $p(B) \neq 0$) Bayes' Theorem states that

$$P(C|B) = \frac{P(B|C)P(C)}{P(B)}.$$

Proof. The probability of two events C and B happening is equal to

$$P(C \cap B) = P(C)P(B|C) \tag{2.11}$$

and the probability of the two events C and B happening is also equal to

$$P(C \cap B) = P(B)P(C|B) \tag{2.12}$$

hence the equating the two leads to Bayes Theorem,

$$P(C|B) = \frac{P(B|C)P(C)}{P(B)} \tag{2.13}$$

□

[1] Though the problems in this thesis do occur in the complex domain, the problem can be posed in the real domain by splitting the problem and stacking it.

It is possible to frame Bayes' Theorem in terms of the inverse problem to solve, where $\mathbf{x}$ and $\mathbf{y}$ are the two events. Hence,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{2.14}$$

in order to obtain $\mathbf{x}$, the distribution on the LHS of Equation (2.14) needs to be obtained; called the **posterior distribution**. Following this, the $\mathbf{x}$ which maximises this distribution (which can be thought of the $\mathbf{x}$ which has the highest probability of producing the data $\mathbf{y}$) is then a solution of the problem of interest.

For the least squares estimation, $p(\mathbf{x})$, called the **prior distribution** is assumed to be constant (uniform distribution over the whole solution domain), and $p(\mathbf{y})$ is not dependent on $\mathbf{x}$; because of this and the goal being to find the solution which maximizes the distribution, they are both omitted for now as they will not impact the least squares solution.

The remaining distribution is $p(\mathbf{y}|\mathbf{x})$, which is equivalent to $\ell(\mathbf{x};\mathbf{y})$, where ℓ is the likelihood function. The likelihood function is often viewed as a function of the parameters ($\mathbf{x}$) given a sample of data ($\mathbf{y}$). This likelihood function is now what needs maximising. Because $\mathbf{y}$ represents a vector of values, the likelihood function can be written as

$$\ell(\mathbf{x};\mathbf{y}) = \ell(\mathbf{x}; y_1, y_2, ..., y_n), \tag{2.15}$$

where $y_1, y_2, ...y_n$ are the n elements of the vector $\mathbf{y}$. Due to the equivalence between the likelihood and the probability density function, the following holds, assuming the pixel values are independent of eachother,

$$\ell(\mathbf{x};\mathbf{y}) = p(y_1, y_2, ..., y_n|\mathbf{x}) \tag{2.16}$$

$$= p(y_1|\mathbf{x})...p(y_n|\mathbf{x}) \tag{2.17}$$

$$= \prod_{i=1}^{n} p(y_i|\mathbf{x}). \tag{2.18}$$

$$= p(\mathbf{y}|\mathbf{x}) \tag{2.19}$$

From the equation $\mathbf{y} = A\mathbf{x}+\epsilon$, if it is assumed that the noise, ϵ is additive Gaussian white noise, then the observed data $\mathbf{y}$ has a Gaussian distribution

$$\mathbf{y} \sim \mathcal{N}(A\mathbf{x}, I\sigma^2), \tag{2.20}$$

where I is the identity matrix and σ^2 is the variance of the noise. In practice this assumption will be incorrect, however it appears sufficient in [7] in clean data. The thesis aims to evaluate suitability of using this assumption with degraded/interference riddled data.

If the assumption that ϵ is additive Gaussian noise has been made, then the probability distribution for each y_i will be a Gaussian distribution with mean $A_i\mathbf{x}$, (this indicates that given an $\mathbf{x}$, the expected value of y_i to be close the value of $A_i\mathbf{x}$)

$$p(y_i|\mathbf{x}) = c\exp\left(-\frac{1}{2}(y_i - A_i\mathbf{x})^2\right), \tag{2.21}$$

where c is a constant and A_i represents the i^{th} row of A. Hence now it is the case that,

$$\ell(\mathbf{x};\mathbf{y}) = \prod_{i=1}^{n} c\exp\left(-\frac{1}{2}(y_i - A_i\mathbf{x})^2\right) \tag{2.22}$$

$$= e^{(\sum_{i=1}^{n} -\frac{1}{2}(y_i - A_i\mathbf{x})^2)} = \exp\left(-\frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2\right). \tag{2.23}$$

The $\mathbf{x}$ which maximises the likelihood function is a solution to the inverse problem and can be known as the maximum likelihood (ML) estimate. To now find the $\mathbf{x}$ which maximises this function (and hence will be a solution to the problem), the natural logarithm, and matrix algebra are used to arrive at a solution. The maximum likelihood estimator is

$$\mathbf{x}_{\text{Maximum Likelihood}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmax}} \exp\left(-\frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2\right). \tag{2.24}$$

As the natural logarithm is a strictly increasing function, the exponential from this expression is able to be removed; to arrive at,

$$\mathbf{x}_{\text{Maximim Likelihood}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmax}}(-\frac{1}{2})||\mathbf{y} - A\mathbf{x}||_2^2. \tag{2.25}$$

Dividing through by $-\frac{1}{2}$ (note the negative and the switch from maximum to minimum here) gives

$$\mathbf{x}_{\text{Maximum Likelihood}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmin}} ||\mathbf{y} - A\mathbf{x}||_2^2. \tag{2.26}$$

The $\mathbf{x}$ which maximises equation 2.4 is equivalent to the $\mathbf{x}$ which minimises $||y - Ax||_2^2$, hence

$$\mathbf{x}_{\text{Least Squares}} = \mathbf{x}_{\text{Maximum Liklihood}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmin}} ||\mathbf{y} - A\mathbf{x}||_2^2. \tag{2.27}$$

This solution is derived only from the the likelihood distribution, no prior knowledge about the solution is incorporated into the solution. Due to this, and the inverse problem being ill-posed, this solution can be unstable; which means that small changes in the input can cause large changes in the solution, as explored in [62]. The equivalence between the maximum likelihood formulation and least squares problem in the classical setting, is explored in [63].

2.1.2 Tikhonov

As mentioned in section 2, it is possible to regularise the least squares solution in order to penalize a particular behavior of the solution, making it more stable simultaneously. From section 2, it can be expressed as

$$\mathbf{x}_{\text{Tikhonov}} = \underset{\mathbf{x}}{\operatorname{argmin}} ||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||H\mathbf{x}||_2^2. \tag{2.28}$$

Here, λ is a tuning parameter, and H is a matrix which represents what choice of behavior is going to be penalized in the solutions. Like with the least squares solution, it is possible to find an analytical solution to the minimisation problem

$$\mathbf{x}_{\text{Tikhonov}} = (A^TA + \lambda H^TH)^{-1}A^T\mathbf{y}. \tag{2.29}$$

The derivation of this follows almost identically the formulation of the least squares solution.

The Tikhonov solution also has a Bayesian interpretation. As the maximum likelihood estimate can become unstable, particularly when there is a high level of noise, or if the problem is highly ill-posed, prior knowledge which is known about the solution can be input into the formulation. In a maximum a posteriori (MAP) estimate, each parameter of $\mathbf{x}$ from Equation (2.3) (in our cases each parameter will be a scattering

coefficient/final pixel value) has distribution indicating what is assumed about their value before any data is observed, known as a prior distribution.

The posterior distribution for an inverse problem is often displayed as proportionality of a likelihood distribution and a prior distribution,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \ell(\mathbf{x};\mathbf{y})p(\mathbf{x}). \tag{2.30}$$

Here, $p(\mathbf{x})$ is the prior distribution. The choice of the prior distribution determines certain factors about the $\mathbf{x}$ which maximise the posterior distribution. Recall that the prior distribution was emitted from least squares / maximum likelihood formulation, as a uniform prior had been assumed over the whole solution domain. However, it is not often the case that every possible solution is equally likely. Because of this, it is possible to make the prior distribution increase the probability on solutions which are considered more likely.

$$\mathbf{x}_{\text{Maximum a Posteriori}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmax}}\, \ell(\mathbf{y};\mathbf{x})p(\mathbf{x}). \tag{2.31}$$

If the prior distribution is a Gaussian distribution, then the solution $\mathbf{x}_{\text{Maximum a Posteriori}}$ which satisfies Equation (2.31 is known as the Maximum A Posteriori (MAP) estimate).

When the prior distribution is Gaussian, this can be written as

$$\mathbf{x} \sim \mathcal{N}(\mu, (\lambda H^T H)^{-1}), \tag{2.32}$$

where H is known as the regularization operator. Deriving the maximising $\mathbf{x}$, the process is much the same as for the ML estimate; however from Equation (2.14), the distribution $p(\mathbf{x})$ is not ignored. Instead it is used to input prior knowledge that may be known about the solution. The computation of $\ell(\mathbf{x};\mathbf{y})$ remains the same, however instead of maximising just $\ell(\mathbf{x};\mathbf{y})$, it is now $\ell(\mathbf{x};\mathbf{y})p(\mathbf{x})$, that is desirable to maximise. Putting these distributions into Equation 2.31 gives,

$$\mathbf{x}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}}{\operatorname{argmax}} \exp\left(-\frac{1}{2}(c_1||\mathbf{y} - A\mathbf{x}||(_2^2 + c_2||H(\mathbf{x}-\mu)||_2^2)\right). \tag{2.33}$$

The c_1 and c_2 stem from the variance of each of the likelihood and prior distribution. In the maximum likelihood derivation this was ommited as a constant as there was only

one variance it did not impact on the solution, however for this solution as there are two Gaussian distributions it is necessary to keep the ratio of the two in the solution. From here, much like the maximum likelihood estimate, this maximisation problem can be turned into a minimisation problem.

$$\mathbf{x}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}}{\text{argmax}} \exp\left(-||\mathbf{y} - A\mathbf{x}||_2^2 - \lambda||H\mathbf{x}||_2^2\right) \tag{2.34}$$

$$= \underset{\mathbf{x}\in\mathbb{R}}{\text{argmin}} ||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||H\mathbf{x}||_2^2. \tag{2.35}$$

The parameter λ, the tuning parameter, comes from c_1 and c_2, ($\lambda = \frac{c_1}{c_2}$), and is put in front of the regularisation term to align with convention. Equation (2.35) has solution

$$\mathbf{x}_{\text{MAP}} = \mathbf{x}_{\text{Tikhonov}} = (A^T A + \lambda H^T H)^{-1} A^T \mathbf{y} \tag{2.36}$$

Figure 2.3 shows normal distributions with zero mean and varying standard deviations, when using this as a sparsity promoting prior for a Bayesian inference it is easy to see that a prior with a smaller standard deviation results in a much higher peak at zero, and therefore lower tails. This means that the smaller the standard deviation of the (normally distributed) prior, the more sparsity promoting it is.

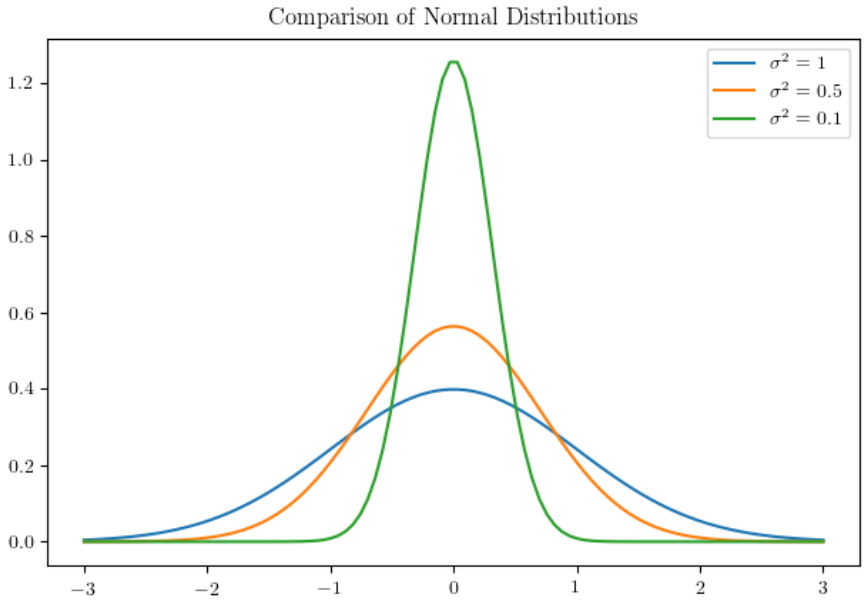

Figure 2.3: Comparison of normal distributions with variance σ^2

A taller peak at zero in a prior assigns a higher probability that this will be the recovered value. For this type of prior, the variance is inversely proportional to the

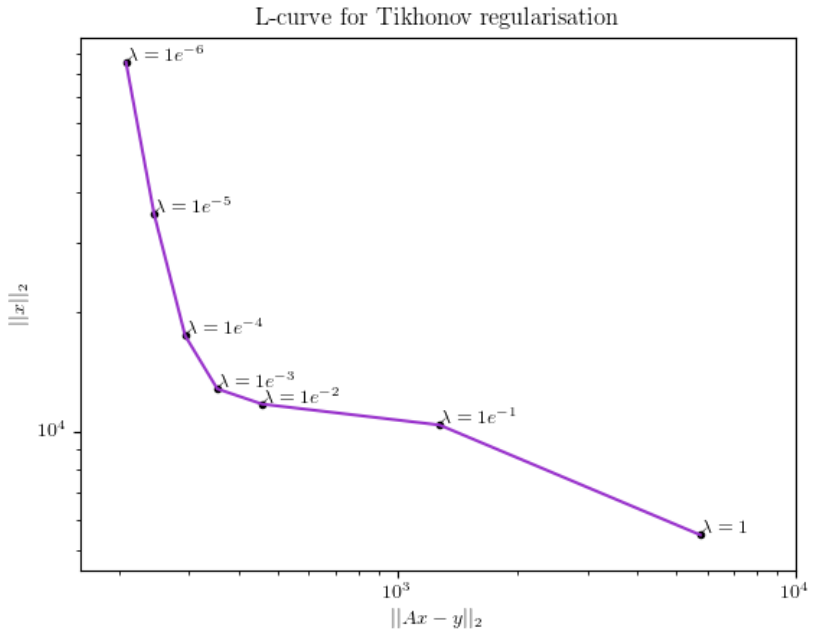

Figure 2.4: L-curve plot for Tikhonov regularisation.

"regularisation parameter". Hence, a smaller variance parameter results in a higher penalty term λ.

The performance of an inversion scheme can be degraded if too high of a regularisation parameter is used since you will have a solution which fits the prior but not the data that has been observed. Figure 2.4 displays an L-curve plot for Tikhonov regularisation. This compares the magnitude of $||A\mathbf{x} - \mathbf{y}||_2$ (the data fitting term) and $||x||_2$ (the regularising term). The optimal λ is often the λ most in the corner of the curve; $\lambda = 1e^{-3}$ in this example. The MAP estimate is the same as the Tikhonov solution. The equivalence of these solutions is explored more in [64], and an example of success that has been achieved in image formation in radar imaging using these techniques is [65].

Chapter 3

Literature Review

3.1 Imaging outside of Radar

This section explores the current state-of-the-art concepts and methods using Bayesian inference that has been adopted outside of radar imaging. In particular, this will look at the forward modelling used, the selection choice for distributions, and method of solving; both image formation and image improvement techniques are considered.

3.1.1 Neurology Imaging

In neurology, inverse problems arise when producing an image of a brain, the most common problems being mapping brain activity and obtaining the physical properties of the brain. Bayesian methods have been proposed for solving both.

In [9] the neural dynamics are studied. Typically, a system of differential equations is used to model the dynamics of brain activity, which map neural dynamics θ to measured data $\mathbf{y}$ via

$$\mathbf{y} = f(\theta) + \epsilon, \tag{3.1}$$

where f here is a forward observation model. The approach in [9], computes θ, the neural dynamics, from $\mathbf{y}$, the measured data, and ϵ being additive Gaussian noise. A Gaussian likelihood function has been used, which is centered around the forward observation model which has been proposed for brain activity dynamics. They have also opted to use Gaussian priors over the neuronal connecting and haemodynamic parameters (the parameters represented by θ) associated with the inverse problem.

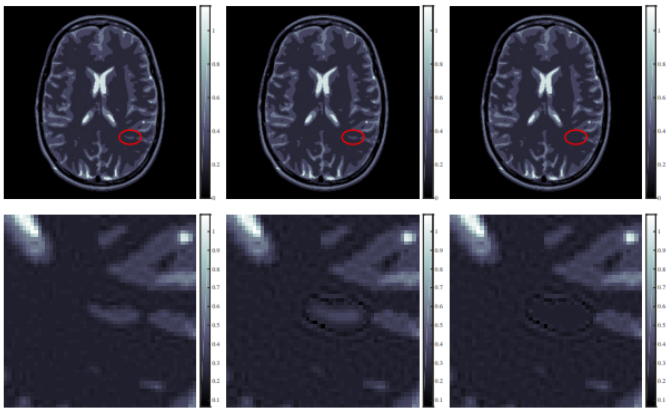

Figure 3.1: Structure of interest in uncertainty quantification. Image via Pereyra et al [2].

The noise has been modelled as additive Gaussian noise. Variational Bayes (VB) has then been used to do the computation and form the image, however they noted that VB, whilst being computationally preferable over MCMC methods is not always robust to local maxima.

In [10] a different brain imaging technique is presented, for magnetic resonance imaging (MRI), to recover unknown information θ from noisy measurements y through the relation in equation 3.1 ($\mathbf{y} = f(\theta) + \epsilon$).

To do this, both the prior and likelihood distributions have been expressed in exponential form, which makes finding the MAP (maximum a posteriori) estimate a simpler and less computationally intense minimisation problem. In medical imaging, it is likely to be more desirable to produce an image quicker (say with a MAP estimate) rather than do a more computationally heavy procedure due to the nature of the imaging needs.

This idea has been built upon in [2] where an uncertainty quantification (UQ) method has been proposed, shown in figure 3.1. This uses the MAP estimate to produce an initial image, and structures of interest (such as tumours in this case) are isolated and the uncertainty in these structures is quantified. Along with the MAP estimate, the scheme produces a set of likely images taken from what is known as a posterior credible region [11], to compare the particular structure in each image in the set. This approach is allows the uncertainty to be quantified without performing a computationally heavy procedure to obtain the full posterior distribution. The computational advantage increases as the dimensions of the problem at hand increases-

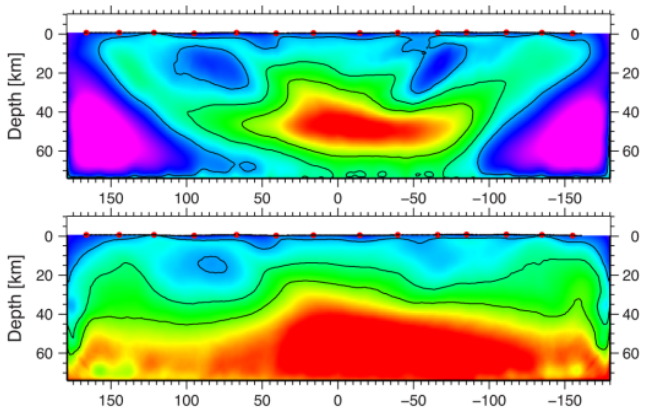

Figure 3.2: The result of an inverse problem in seismology which is a representation of the make up of the earth's crust, from [3]

with the example in [2] noting that MCMC methods would have taken 10^6 iterations whereas the UQ proposed took 10^3 iterations (with iterations for each method having similar computational cost). This approach is not restricted to neurology imaging, and has also been applied in astronomy in [2].

Compressed sensing has been considered in the context of MRI in [12]. For more details on compressed sensing see section 3.2.1. The motivation for this stems from the desire for a quicker acquisition time and an improved reconstruction quality. The approach used in [12] was able to obtain reconstructions with smaller errors when compared to other widely used methods.

3.1.2 Seismology

In seismology, inverse problems arise when wanting to find out information about the earth's seismic structure, by sending and receiving a seismic wave through the seismic structure in order to produce an image of the earth's crust and mantle. An example of the output of this is shown in figure 3.2.

In 2012 a general framework for Bayesian inversion for this problem was suggested in [4]. Since this, it has been well adopted within Bayesian inversion in seismology research; with researchers making different adaptations to it in [3] and [13], amongst others.

The approach suggested in [4] formulates the inverse problem as equation 3.1 ($\mathbf{y} = f(\theta) + \epsilon$), where θ is the set of parameters which they are interested in recovering,

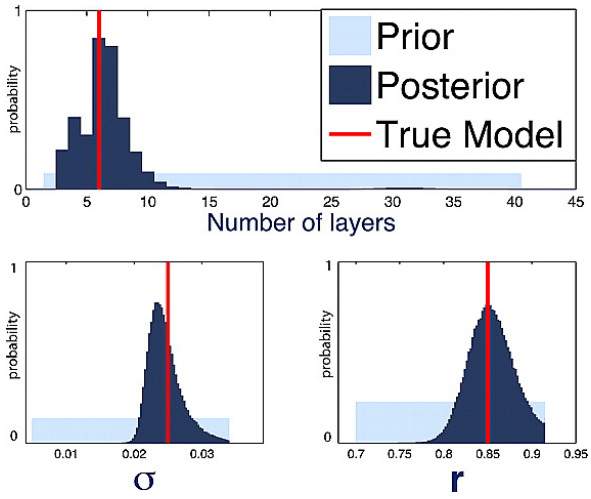

Figure 3.3: A graphical representation of MCMC sampling, from [4]. Here σ and r are parameters from the unknown variable.

y is the set of observed data (the received seismic waves), and f is a forward model relating the parameters to the received RF wave.

In the Bayesian framework this has been modelled using a Gaussian likelihood centred around the forward model. They have also employed uniform, hierarchical priors on the set of parameters. MCMC was used to explore the posterior, shown graphically in figure 3.3, has been used in this case to compute the solution, where the posterior mean can be taken as the intensity value for each pixel, to produce an image. The nature of seismology imaging allows for a longer acquisition time, as in many cases the need is not as urgent as in other applications. This makes the drawbacks of MCMC less important to this application. The advantage of using MCMC was noted, in that multiple likely final images were produced which each fit the data well. From this a reference image and its standard deviation was able to be obtained. The performance was said to be similar to conventional tomographic methods. The advantage of having a reference image along with a standard deviation would be desirable for applications within radar, as this could enable a quantification of the outcome's confidence.

3.1.3 Dental Imaging

Inverse problems occur in dental X-ray imaging, as in [14], where projection images of the teeth are taken from a few sparsely distributed 2D projection images and a 3D x-ray is reconstructed. The reconstruction of the 3D structure based on a few projection images becomes ill-posed, hence a Bayesian framework is being considered for the

reconstruction. The model used here is as in equation 3.1, $\mathbf{y} = f(\theta) + \epsilon$, where y are projection measurements, θ is a vector of attenuation values, A matrix approximates the projection measurement. Here θ is the set of parameters to be recovered.

The model observation noise is assumed to be Gaussian noise and hence the likelihood distribution is Gaussian. The prior distribution for the attenuation parameters has been modelled using both the total variation prior [15], which enforces sparsity in the derivative of the solution, in addition to the Laplace prior. These favour piecewise regular solutions with small variation in the attenuation parameter within each region, and also promote sparsity within the reconstruction. This choice of likelihood distribution and prior produces a conjugate posterior distribution, which in turn corresponds to a closed form expression of the posterior distribution.

This approach resulted in improved image quality over currently-used methods, and it was noted that MCMC methods were not likely to be feasible for the computation of this in the near future due to their high computational cost. It appears that image quality is determined purely visually, and hence it may be somewhat subjective.

3.1.4 Summary

To summarise, a variety of similar Bayesian inference methods have been suggested for use in areas other than radar imaging. The main bulk of the literature in this area represents work purely in a research stage. The main differences between approaches in these areas comes in the computation or sampling of the posterior distribution, with the different methods being more or less suited to particular applications, mainly due to timing demands.

3.2 Radar Imaging: Current State

This section reviews the Bayesian frameworks which have been adopted in current research within the radar imaging area.

3.2.1 Compressive Sensing

There is a substantial amount of research into Bayesian approaches for solving inverse problems within compressive sensing, [16, 17, 18, 19]. The general framework of each

of the approaches is constant across current research. The method used in [19] is outlined.

In [19] a Bayesian framework for compressive sensing is defined. The basic concept behind compressive sensing is to recover a signal which has been under-sampled. This is generally done via sparse reconstruction. This can be expressed as an inverse problem

$$\mathbf{y} = A\theta + \epsilon \tag{3.2}$$

where $\mathbf{y}$ is the under-sampled measured signal, θ refers to some weights (signal basis coefficients which the original non compressed signal can be easily produced from) and A is a forward model (in the case of compressive sensing this is generally computed via a SAR observation matrix and an undersampling matrix). Hence the inverse problem is to recover θ which is unknown (θ could be the full signal, or the scene being imaged depending on what is being recovered, in the case that the full signal is being recovered A will only represent the under-sampling and not the SAR itself), and $\mathbf{y}$ which is known. The motivation for compressive sensing is to have a shorter acquisition time, at the cost of a longer recovery time.

The framework which they have adopted to solve this consists of a **hierarchical Gamma-Gaussian likelihood** (a Gaussian likelihood which is dependent on hyperparameters, which are subsequently gamma distributed) and a sparsity promoting prior. For information about the gamma distribution see [20]. The likelihood distribution for $\mathbf{y}$ is

$$p(\mathbf{y} \mid \theta, \beta) = (2\pi\beta^{-1})^{-\frac{n}{2}} \exp\left(-\frac{\beta}{2}||\mathbf{y} - A\theta||_2^2\right) \tag{3.3}$$

where β is a precision parameter (for a Gaussian maximum-entropy probability model for the components in the error ϵ), and has a gamma distribution

$$p(\beta \mid a_0, b_0) = \Gamma(\beta, a_0, b_0), \tag{3.4}$$

where $n \in \mathbb{N}$ is the dimension and hyperparameters a_0 and b_0 are set with a non-informative distribution, like uniform distribution for example,

$$p(\alpha_0) = \begin{cases} \frac{1}{b-a}, & \text{if } a > \alpha < b \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

A non-informative distribution ensures that the parameters a_0, b_0, β are not restrictive to the model. The full likelihood distribution can be expressed as

$$p(\mathbf{y} \mid \theta) = p(\mathbf{y} \mid \theta, \beta)p(\beta \mid a_0, b_0)p(a_0)p(b_0). \tag{3.6}$$

In this case, it was decided to promote sparseness using an automatic relevance determination (ARD) prior [21]. This prior promotes sparseness on the signal basis coefficients θ, by favouring that they are equal to zero. The ARD prior can be defined by

$$p(\theta \mid \alpha, \beta) = \prod_{n=1}^{N} (2\pi)^{-\frac{1}{2}} (\beta\alpha_n)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\beta\alpha_n\theta_n^2\right), \tag{3.7}$$

where β is as above, and α is Gamma distributed;

$$p(\alpha \mid c_0, d_0) = \prod_{n=1}^{N} \Gamma(\alpha_n \mid c_0, d_0), \tag{3.8}$$

with c_0 and d_0 chosen to have non-informative priors. Note that this is not the only formulation of a sparsity promoting prior that is used in current research.

A MAP estimate was then used to reconstruct the signal. Many compressive sensing applications are of high dimension, hence MCMC will often be impracticable and so the MAP estimate for this problem is preferable in many cases.

In [22], a Bayesian compressive sensing framework specifically designed for SAR is proposed. Non-Bayesian compressive sensing techniques for SAR have been studied, amongst others, in [23, 24]. In [22], spotlight mode SAR was considered, and this was represented in a model as in equation 3.2 ($\mathbf{y} = A\theta + \epsilon$) where y is the under sampled signal, θ is the unknown to recover and A is a known sensing matrix (composed of the product of an observation matrix and an under-sampling matrix), much like in section 3.2.1.

In order for them to promote sparsity in the unknown parameter, the Laplace density function was considered, however a zero mean Gaussian prior was chosen instead for conjugacy with the Gaussian likelihood. Their method was compared to both the original image and to a non-Bayesian compressive sensing method from experimental results. They showed that the Bayesian Compressive Sensing method outperforms the others on the metrics which were chosen for comparison.

In [5], a passive multi-static SAR imaging method has been suggested, based on Bayesian Compressive Sensing. The passive system has made use of broadcasting and

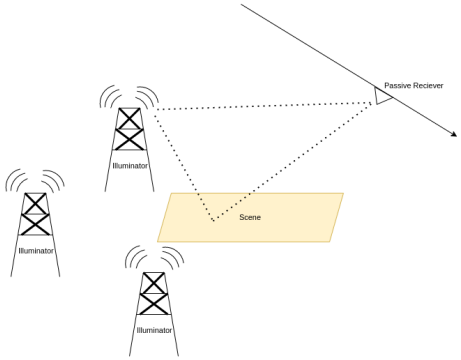

Figure 3.4: A passive multi-static SAR setup, reproduced from [5].

navigation signals, as shown in figure 3.4.

They have derived a forward model for multi-static passive SAR, and have then converted the sparse target imaging into a group sparse signal reconstruction problem, which they have tackled similarly to that described above. Their approach produced higher resolution images in comparison to other state-of-the-art techniques produced.

A Bayesian approach has been suggested for speckle reduction in [25]. SAR images are often affected by multiplicative speckle, due to the nature of the scattering phenomenon. This can be expressed as in equation 3.2, ($\mathbf{y} = A\theta + \epsilon$), where y is the observed (and already formed) SAR image, A is the multiplicative noise, and θ is the SAR image de-noised. Here the additive noise has been ignored as it is a less significant term. The main contribution of the paper is that is it is deduced that a SAR image can be accurately modelled by an alpha-stable distribution [25]. A MAP estimate is used to compute the solution here.

3.2.2 Inverse Synthetic Aperture Radar Imaging

Bayesian methods have been proposed for uses in ISAR (Inverse Synthetic Aperture Radar) imaging, exploiting the sparse nature that ISAR images generally adopt, in order for sparse signal recovery algorithms to be used. An implementation of sparse signal recovery for ISAR imaging has been presented in [26]. As ISAR images are generally sparse, and only contain small regions of targets, sparse signal recovery algorithms are applicable. A Bayesian framework was proposed using a logarithmic Laplacian sparsity promoting prior, defined in [26]. The inverse problem can be expressed as equation 3.2, $\mathbf{y} = A\theta + \epsilon$, where $\mathbf{y}$ and θ are vector forms of the range

profiles and the ISAR image respectively, and A is a matrix representing an inverse fast Fourier transform multiplied with phase error.

In the Bayesian framework the additive noise ϵ is assumed to be complex Gaussian distributed with zero mean. Hence the likelihood function is normally distributed centred around $A\theta$, with variance represented by a covariance matrix. As the ISAR image represented by θ is assumed to be sparse, a sparsity promoting prior is placed on θ.

In this approach the MAP estimate has been used to reconstruct the image. This Bayesian MAP ISAR imaging approach has yielded improvement over other methods in low SNR settings, in the testing carried out in [26] on both simulated and measured data. The proposed next step for this approach is to consider targets with complex motion.

In [27] a different ISAR imaging algorithm was proposed for imaging targets with complex motion under low SNR. The method is composed of the following steps:

1. A de-noising method improves the SNR of the range profiles, which accumulates the aligned range profiles non-coherently to obtain a window for noise suppression.

2. An optimal coherent processing interval (CPI) selection algorithm finds an interval where the Doppler shift is relatively stationary.

3. A sparse aperture imaging method within the Bayesian framework is then used to reconstruct the well-focussed ISAR image, as the CPI often contains not enough pulses to produce well focused ISAR images.

After the CPI selection algorithm has been applied, the problem can be expressed as equation 3.2, $\mathbf{y} = A\theta + \epsilon$, where y is a complex vector representing the range profiles from the CPI algorithm, A is the partial Fourier matrix, and θ represents the ISAR image to be found.

To model this, a complex gamma-Gaussian hierarchical distribution was chosen for the likelihood function. This enables the hyperparameters describing the noise to best fit the data observed. The prior chosen to put on θ in this case was a Laplacian scale mixture (LSM) [28] prior in order to promote sparsity in the ISAR image.

In this approach, variational Bayes was chosen over MCMC due to its efficiency of implementation. This method has shown improvements over other Bayesian frameworks for ISAR imaging, but was not explicitly compared to an approach which did not use a Bayesian method to solve the inverse problem.

3.2.3 Radar Applications

This section details potential benefits to application areas within radar which are of particular interest to the radar community.

Through wall radar imaging

Bayesian compressive sensing has been suggested in [29, 30, 31], for through wall radar imaging (TWRI). The scene which is behind a wall is assumed to be sparse, making compressive sensing applicable. Bayesian compressive sensing has been used in TWRI to both directly reconstruct an imaged scene using reconstruction algorithms, and to restore a full data set in order for conventional image formation methods to be used.

In [18], an image formation method is proposed for multi-polarization through-the-wall radar imaging. In the proposed method a Bayesian framework is adopted, among other techniques:

1. The radar signals are recovered using Bayesian compressed sensing in the wavelet domain.

2. A subspace projection method mitigates the front wall reflections.

3. Principle component analysis is used to remove noise.

4. A linear model relates the compressed wavelet coefficients to the image, where Bayesian compressed sensing is employed again to form the images.

This method has been seen to form clearer images, and reduced background clutter when compared to a delay-and-sum beam forming method and a non-Bayesian compressed sensing method.

Bayesian methods for this have also seen improvements in target location, and clutter suppression when used on simulated data and real TWRI data [31]. It is

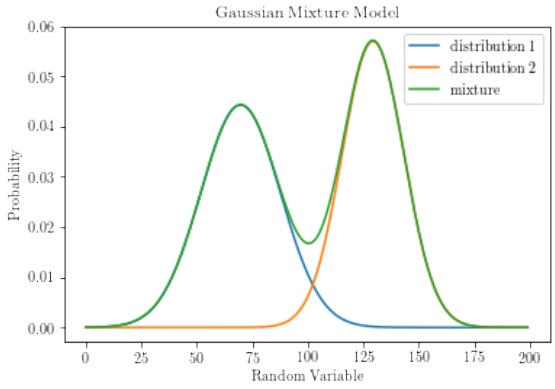

Figure 3.5: Graphical example of a Gaussian Mixture Model, reproduced from [6]

suggested that these methods will reduce data acquisition time, and improve TWRI target detection and classification[30].

Enhanced Sensor Protection

Different approaches within radar have been suggested to counteract jamming. In [32], any jamming present in the signals has been modelled within the noise.

The key difference in this approach is that a Gaussian Mixture Model (see figure 3.5) has been used to describe the unknown noise. By using a sufficient number of Gaussian distributions, a much wider variety of noise can be modelled than by using a single Gaussian distribution. The use of this approach transforms jamming suppression into a problem of parameter estimation. The proposed method has been shown to provide an effective way to perform high-resolution radar imaging with barrage jamming when tested experimentally. It is proposed in [32] that this approach could be extended to non steadily-moving targets.

In [33], another method has been identified, which states it looks promising for interference and jamming suppression. An RCI (Radar Coincidence Imaging) method has been proposed applying a Bayesian technique to recover approximation coefficients relating to the image. In basic simulation, this method has shown advantages (such as high resolution and smaller error in interference-like scenarios) over other techniques but is noted to suffer from a high computational requirement.

High Clutter

In [34] a space-time adaptive processing (STAP) [35] algorithm utilising Bayesian inference has been suggested, with the intention of detecting slow moving targets with strong clutter. Conventional STAP algorithms have shown effectiveness in this area [36]. The new approach being proposed uses sparse recovery to obtain information about the adjacent cells whilst current cells are being tested. The inaccuracy of this information causes the performance of the STAP to degrade, hence why this area of the algorithm was targeted for improvement.

In this scenario, the inverse problem can be cast in the form of ref equation, where θ is a vector indicating the information about the cells that is of interest, A is a forward model relating the vector of interest θ to the measured snapshot $\mathbf{y}$. With the goal being to reconstruct θ, a general sparse reconstruction formalism has been followed, with zero-mean Gaussian priors and a Gaussian likelihood centred on the forward model A. This approach was tested against other STAP algorithms, and was shown to produce comparable results.

The approach outlined in section 3.2.3, proposed in [32] has also been suggested to have advantages in high clutter environments. This is due to the nature of the noise model that they have adopted.

Operating radar within high clutter environments is of particular interest due to its applications in foliage penetration and urban environments.

Moving Targets

In [37] a SAR algorithm for multiple ground moving target imaging has been proposed. Time frequency representation is a method for moving target estimation, of which Lv's distribution (LVD) [38] has achieved good performance in dealing with multiple moving targets. However LVD has some practicality issues: the limited accuracy of the Doppler chirp rate can cause the image to become unfocused.

In [37] a SBL framework is employed to obtain a focused image. The inverse problem here can be expressed in the form of ref equation, with θ representing the final SAR moving image, y the observed signal, and A composed of the product of a phase variation matrix and a matrix determined by the Doppler chirp rate. Using a Gaussian likelihood, a hierarchical Laplace prior and VB to compute the posterior distribution,

[37] have realised potential improvements by the use of this method. However, a high signal-to-clutter-noise ratio (SCNR) is needed for an enhanced moving target signature to be obtained. This has been compared to other LVD methods via the correlation between the target image truth and the processed image, with different SCNR inputs.

Difficult Targets

Within radar, difficult targets can refer to targets which are shaped in a particular way to reduce their radar cross section (often referred to as low observable platforms) and thus reducing the radar response that they produce. Because of this, approaches taken to image targets with a low signal to noise ratio (SNR) are of interest. Methods which are robust to low SNR are of interest to the defense community due to the desire to image low observable platforms. This has been approached using methods, like compressive sensing, seen previously in this thesis. In [39], a Bayesian compressive sensing framework has been developed to produce a well-focused SAR image under low SNR conditions. This approach utilities a sparsity promoting prior distribution which is capable of suppressing the noise. A similar approach has been proposed in [40].

In [41] an algorithm with a high denoising capability has also been proposed, using Bayesian MAP estimation, for producing an ISAR image. It has been noted that the proposed method has a high robustness, and an ability to produce a well focused image under low SNR.

3.2.4 SAR Forward Model

Here, the particular types of forward modeling used in synthetic aperture radar imaging are looked at. When considering the SAR image formation as an inverse problem, it is necessary to describe relationship between the unknown data and the known data by means of a forward model.

These forward models can be split into two main categories: linear and non-linear.

Within the non-linear category is full wave inversion, being the most accurate modeling available. This involves solving Maxwell's equations, as is done in [42, 43, 44, 45]. Generally this is done for small scene sizes, as computationally it is restrictive.

In practice, current state-of-the-art imaging algorithms tend to be back-projection based, which assumes a linear relationship among other assumptions. A delayed signal

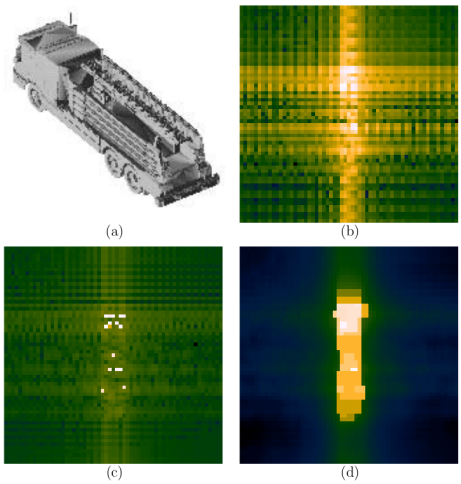

Figure 3.6: An example of some images obtained in [7], © 2001 IEEE, using linear forward modeling.

forward model, which shares the same assumptions as back-projection is used in [7, 46, 47, 48]. By using a linear model such as this; some assumptions which have been made include:

- ★ stationary phase centre assumption- assumes that the radar itself has not moved in the time that the transmitted pulse takes to travel to and from the scene.
- ★ single scattering assumption- assumes that the transmitted wave bounces off one target in the scene and returns straight back to the receiver immediately (no multi bounce).
- ★ plane wave assumption- assumes that at the time the radar wave reaches the scene, it is a plane wave, rather than a curve centred around the transmitter.

These simplifying assumptions have been shown to be wrong, [49]; however, it is often the case that the effects of making these assumptions are considered negligible. These assumptions are generally used in practice. An example of results obtained by using this sort of forward modeling is displayed in Figure 3.6.

A 2D Fourier transform can also be used as a forward model for SAR image formation, as is the case in [50, 51, 52, 53] as it can approximate the relationship between the unknown data and the known data, however this restricts the geometry of the radar collections somewhat harshly, and generally produces a worse image than a conventional technique such as back-projection.

3.3 Summary

This review has outlined the approaches taken by research into medical, seismology and radar imaging; with the aim of reviewing methods which could provide benefits in radar imaging. There has been a sizeable amount of research into Bayesian techniques applicable to radar imaging, including approaches which could yield benefits in the particular areas of interest. Techniques showing benefits in through wall radar, clutter and interference suppression, and techniques robust to low SNR have been reviewed.

An advantage to using Bayesian methods for inverse problems is that the posterior distribution not only produces a solution, but also information which allows quantification of the uncertainty of the solution. The ability to do this is of particular interest in the defence community. With that said, not a lot of the research into Bayesian methods for radar imaging specifically utilises the other information provided by the posterior distribution. Approaches in medical imaging and seismology have touched on this more, with an uncertainty quantification method for particular structures being suggested in [10] and a set of likely images being produced in [4].

The main drawback to using Bayesian methods is the associated computational expense, especially if MCMC is used. This drawback is particularly detrimental for inverse problems within imaging due to the dimensionality of the problems. Less computationally demanding variants of MCMC is an ongoing subject of current research.

Chapter 4

Synthetic Aperture Radar

There are two main applications considered in this thesis, as mentioned,

- ★ development of forming SAR images directly from measured phase history data (i.e. raw data),
- ★ development of improving the usability of poor quality or degraded SAR images, particularly making use of multi-aspect collections.

This section describes how these have two problems have been modeled, as well as posing them as inverse problems.

4.1 Image formation

In SAR imaging the aim is to produce a human-interpretable image from data acquired from a radar. Radar systems can detect how far away an object is by transmitting an electromagnetic wave and measuring what time has elapsed before the wave has returned (after reflecting off of said object). In producing an image using this notion, the resolution is restricted by some physical aspects of the radars transmitter and receiver.

Consider a system as shown in figure 4.1. The maximum resolution in the azimuth direction is dependent on the beam-width (which is in turn dependent on aperture size). The maximum resolution in the range direction (ignoring the movement of the radar from figure 4.1 initially) is dependent on the pulse length. See [49] for more details on this.

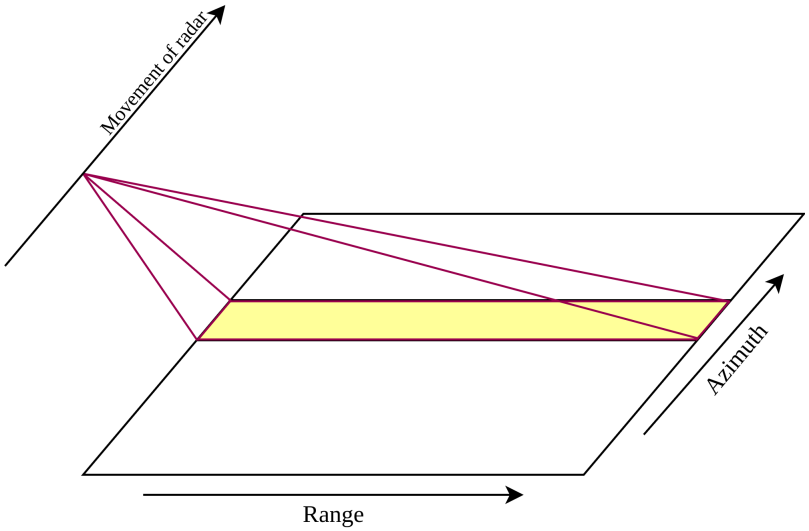

Figure 4.1: Diagram displaying the set up of a SAR system.

To overcome the restriction in the range direction, techniques such as pulse compression is used. This usually means emitting a long pulse (needing a low amount of power), and later compressing it to a shorter pulse in order to get the resolution benefits of a short pulse length, [54]. To overcome the restriction in the azimuth direction, it is possible to synthetically make the aperture longer by making use of the motion of a moving platform, and doing multiple transmits and receives with the radar, hence the term Synthetic Aperture Radar.

By employing these sorts of techniques to allow for a higher resolution image, the relationship between a human readable image of the scene and the data acquired from the radar becomes more complicated.

Consider a scene to image, which has a reflectivity function $f : D \to \mathcal{C}$, where $D \subset \mathcal{R}^2$. This is assuming the scene is 2-dimensional, however this methodology does allow to extend to three dimensions.

A pulse transmitted by the radar is called a wave, and a linear frequency modulated wave (LFM) is defined as

$$s(t) = \begin{cases} e^{j(\omega_0 t + \alpha t^2)}, & \text{for } |t| \leq \frac{T}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where ω_0 is the carrier frequency and 2α is the chirp rate, t is the time, and T is the pulse length.

If this waveform is transmitted in the direction of the scene, the expected return from this scene at time t and pulse n would be

$$q(x_0, y_0, t, n) = f(x_0, y_0)s(t - \tau_0), \tag{4.2}$$

where τ_0 is the time delay of the signal, related to the distance the particular scatterer $(x_0, y_0) \in D$ is from the radar; $\tau_0 = 2r/c$ where r is the distance between the radar and the scatterer and c is the speed of light. This assumes a monostatic radar, but this could trivially be altered to $\tau_0 = (r_{\text{receiver}} + r_{\text{transmitter}})/c$.

Equation 4.2 is the return expected from one scatterer in the scene, the sum of all reflectivities of scatterers which are at a distance of $R + u_0$ (where R is the distance from the radar to the center of the scene) to the radar (along the line $u = u_0$) can be expressed as

$$p(n, u) = \int\int_D f(x, y)\delta(u - x\cos(\theta) - y\sin(\theta))dxdy. \tag{4.3}$$

Here, θ is the angle between the x axis and the u axis when the radar is at position n, in figure 4.2. The selection of scatterers at a distance of $R + u_0$ to the radar would actually lie along an arced line; however an assumption has been made that the arc is negligible due to the distance between the radar and the scene, (for a closer spaced geometry this assumption may fail and cause issues).

Hence the expected return in signal from the scatterers positioned along $u = u_0$ would be

$$q(t, n, u_0) = p(n, u_0)s(t - \frac{2(R + u_0)}{c}) \tag{4.4}$$

and then the response received over the whole range of u would then be

$$q(t, n) = \int_D p(n, u)s(t - \frac{2(R + u)}{c})\mathrm{d}u \tag{4.5}$$

Subbing Equation 4.1 back into this,

$$q(t, n) = \int_D p(n, u)\exp j(\omega_0(t - \frac{2(R + u)}{c} + \alpha(t - \frac{2(R + u)}{c})^2))\mathrm{d}u. \tag{4.6}$$

Demodulating this response and multiplying through by the reference chirp leads to (after omitting terms associated with a higher order, due to the born approximation [55]),

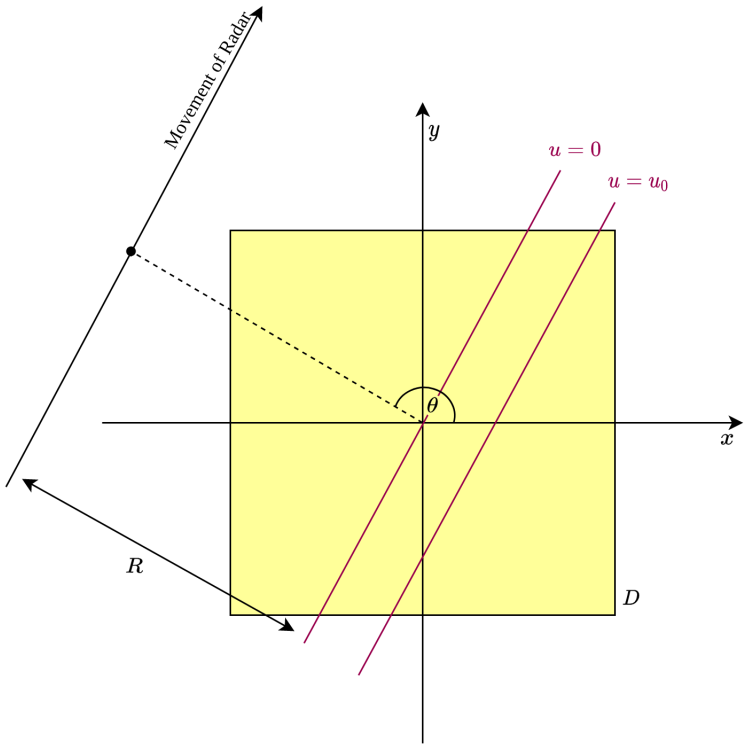

Figure 4.2: Diagram of geometry of imaging scene.

$\exp\left(-j(\omega_0(t-\tau_0)+\alpha(t-\tau)^2)\right)$ yields

$$q(t,n) = \int_D p(n,u)\exp\left(-j\frac{2u}{c}(\omega_0+2\alpha(t-\tau_0))\right)\mathrm{d}u. \tag{4.7}$$

Substituting Equation 4.3 into this arrives at an expression for the received signal as a function of the scene $f(x,y)$,

$$q(t,n) = \int\int\int_D f(x,y)\delta(u-x\cos(\theta)-y\sin(\theta))\exp\left(-j\frac{2u}{c}(\omega_0+2\alpha(t-\tau_0))\right)\mathrm{d}x\mathrm{d}y\mathrm{d}u, \tag{4.8}$$

$$= \int\int_D f(x,y)\exp\left(-j\frac{2u}{c}(\omega_0+2\alpha(t-\tau_0))(x\cos(\theta)-y\sin(\theta))\right)\mathrm{d}x\mathrm{d}y. \tag{4.9}$$

This describes the relationship between the scene reflectivity and the demodulated signal received by the radar, both described here as continuous functions. In reality the signal received by the radar can only be sampled at a certain rate, often referred to as fast times; and any screen displaying an interpretation of the scene will inevitably only have so many pixels. There is also a limit on the resolution of a scene you can reasonably expect obtain from the radar data to consider.

Because of this, it is necessary to discretize Equation 4.9, which can be expressed as

$$\mathbf{y} = A\mathbf{x}, \tag{4.10}$$

Here $\mathbf{y}$ is a vector of the received signal from the radar, containing the value at each sample, stacked by each looking angle (each n from the previous equations). $\mathbf{x}$ is a vector representing the scattering coefficients of the discretized scene, again stacked by ether row or column. A is a matrix representing the relationship in Equation 4.9. An example of how the matrix A could be formed is described in Appendix A.

The starting data, $\mathbf{y}$, is often referred to as phase history data, this is simply raw data obtained from the radar. It contains multiple discretely sampled pulses which have traveled from the radar, to the scene, and back to the radar again.

It is now important to consider the data which is acquired from the radar is unlikely to be completely clean and free of noise, because of this, a noise term is added to Equation 4.10,

$$\mathbf{y} = A\mathbf{x} + \epsilon, \tag{4.11}$$

where the noise term ϵ refers to system noise of the radar and electromagnetic noise in the spectrum. This describes the approach taken to modeling SAR image formation.

4.2 Image Improvement

Improving SAR images which were formed via conventional techniques is the other application explored in this thesis. Consider the case where radar a image already exists, though it contains unwanted features from contributions such as:

- ★ interference,
- ★ unmodelled EM interactions,
- ★ a system's hardware induced noise,
- ★ reconstruction artifacts.

These can cause particularly detrimental effects to the usability of the resultant image. However, said images do contain information about the true scene, and it is possible to extract that information from the poor quality image.

This task can be written as an inverse problem, as in equation 4.11, ($\mathbf{y} = A\mathbf{x} + \epsilon$) where this time $\mathbf{y}$ refers to the existing poor quality radar image, $\mathbf{x}$ the true scene, ϵ noise, and A refers to a relationship model between the poor quality radar image and the true scene.

The problem which is considered in this thesis makes use of an image degradation matrix to model the relationship between the true scene, and the flawed radar image of the scene, which already exist.

An image kernel is a matrix which can be used to apply effects to an image. For example given a 3 by 3 matrix (a common size of image kernel) and a grey-scale image, each pixel value in the image and it's surrounding 8 pixels are multiplied element wise with the matrix. These 9 values are then summed together, which results in the new pixel value. This method described is commonly called convolving an image. The different kernels produce particular effects in the resultant images, for example Gaussian blurring or edge detection, [56].

This approach could be used to obtain an improved image of the scene; however, in the situation where multiple of these radar images exist showing the same scene, they are related to the true scene that is there by some relationship model which is an extension of equation 4.11. Again, this relationship can be exploited to try and obtain a superior image of the scene, which suppresses the effect of the unwanted features.

This problem can be expressed as an inverse problem, such as

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_q \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_q \end{pmatrix} \mathbf{x} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_q \end{pmatrix} \tag{4.12}$$

where each $\mathbf{y}_i$ is a vector representing the existing SAR images which contain the detrimental features, each A_i is a degradation forward model, each ϵ_i is a noise vector, and $\mathbf{x}$ is a vector representing the true scene. This relationship is presented graphically in 4.3. Each A_i could also align the images, however this is not the case in the examples

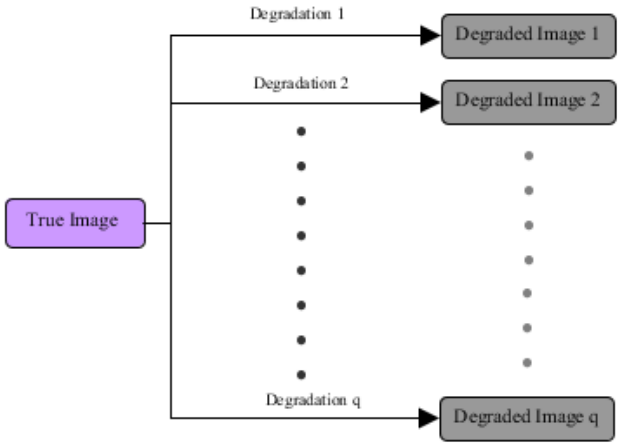

Figure 4.3: Visual representation of forward model from equation 4.12

in this thesis.

Chapter 5

Regularising for Radar Imaging Problems

This chapter implements the solving approaches defined in Chapter 2, particularly the analytic solutions for the ML (least squares) estimate and the MAP (Tikhonov) estimate (using a Gaussian sparsity prior). For the MAP estimate, the parameter for λ was chosen to be that sits in the most left hand corner of Figure 2.4. A higher choice of parameter will favour an even sparser solution, though this may lead to the solution not fiting the measured data particularly well. A lower choice of this parameter will lead to a less sparse solution which alligns to the measured data particularly well, but may be considerably more noisy. The point in Figure 2.4 which is most in the corner of the 'L-curve' is considered an optimal compromise between the two.

5.1 The Impact of Regularisation

The performance of the maximum liklihood/least squares method and the MAP/Tikhonov solution is explored using simulated data and models. Figure 5.1 displays the absolute mean squared error of solutions obtained using these methods to solve inverse problem of the form $\mathbf{y} = A\mathbf{x} + \epsilon$, in the presence of varying amounts of noise. The figure shows that as the noise level increases, the error of the MAP estimate increases at a slower rate than the error of the ML estimate. As this is done purely using simulated data, it is what is known as an inverse crime, [66]; however this allows for explicit computation of errors.

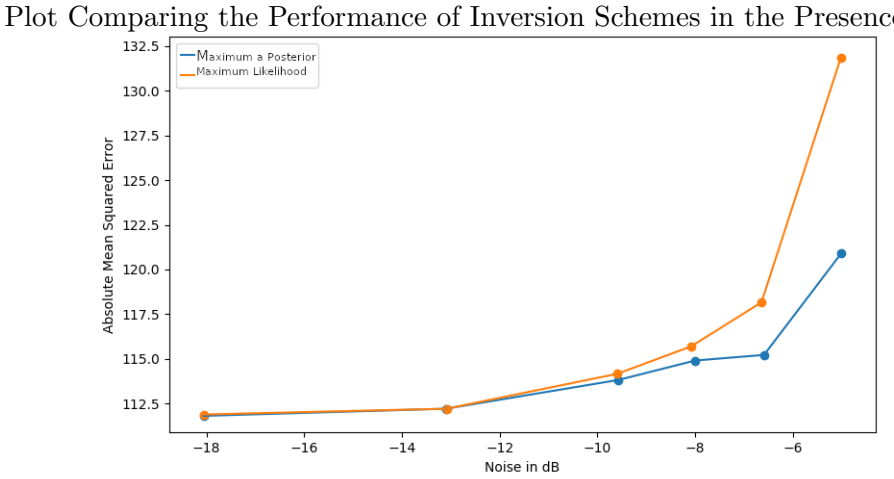

Figure 5.1: Plot comparing the error in Maximum Likelihood (orange) and Maximum a posteriori (blue) reconstructions.

This figure shows the purpose of regularising. To display the effect of regularising on the two applications the thesis is targeting; forming SAR images and improving the usability of poor quality SAR images, simulated comparisons have been carried out.

5.1.1 SAR Image Formation

To test the least squares and Tikhonov method at solving a SAR image formation problem, some simulated phase history data has been produced. These two methods are also compared with a basic back-projection method [8], a current standard choice for forming SAR images. With simulated data it is easily possible to calculate the errors, so an objective statistical comparison was also carried out. This will be less possible when using experimental real data as the ground truth will not be known as specifically.

For this, mean-squared-error was chosen as the metric to do the comparisons. This was due to it's simplicity in both implementation and interpretation. Due to the nature of the simulated experiments that were tested with this, the metric is also not exposed to the risk that the solution may be correct but offset. It is however, not the only choice for a comparison metric, for example a structural similarity index would be a good choice also.

The simulated phase history data is produced on a smaller scale, for ease of computing. The scene which is simulated is a simple scene of a single scattering pixel, on a 15x15 grid. Additive Gaussian noise has been added to the phase history data.

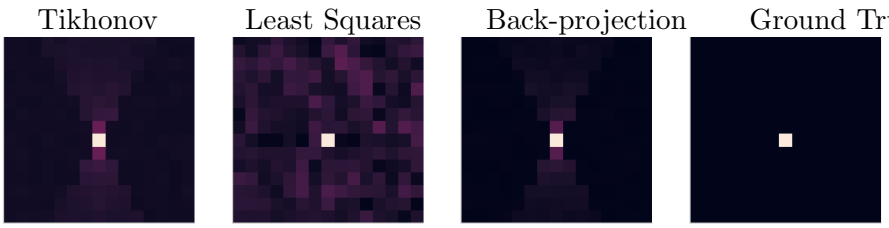

Figure 5.2: Pictures showing the reconstruction of simulated scenes using back-projection, least squares and Tikhonov regularisation. Note that the maximum likelihood estimate is equivalent to the least squares solution and the MAP estimate is equivalent to the Tikhonov solution. These images were produced using the algebraic solutions from chapter 2.

Figure 5.2 shows the reconstructions of this scene using least squares, Tikhonov and back-projection, and the original ground truth scene is also displayed. Here some differences in the reconstructions are apparent; the least squares construction has a higher background noise level, where the Tikhonov and back-projected images have less noise but do appear to have reconstruction artifacts stemming from the scatterer in the scene.

To explore these difference further, the error has been calculated for multiple image constructions for each method. Figure 5.3 shows the spread of the Absolute-Mean-Squared-Error (AMSE) for multiple reconstructions for the ML, MAP and back-projection method. In this figure, it appears that the ML estimate has a higher error compared to the MAP and back-projection methods, and that the MAP errors are marginally lower with a smaller range.

The difference in errors for the different methods can be explored objectively using frequentist statistical tests. Hypothesis testing is a technique that can be used to pick out of two mutually exclusive hypotheses which is more likely, specifically these are:

- ★ Null hypothesis, commonly referred to as H_0
- ★ Alternate hypothesis, commonly referred to as H_1

If there is sufficient evidence, the null hypothesis can be rejected. Whether there is sufficient evidence is decided upon by a test statistic, which is usually a single value which indicates behaviors which would differentiate between the null and alternate hypothesis. The test statistic has a sampling distribution (which is shown as an F-distribution in figure 5.4, but could be others such as a normal or t-distribution). The area under the curve, labeled the rejection region, is also called the p-value [67].

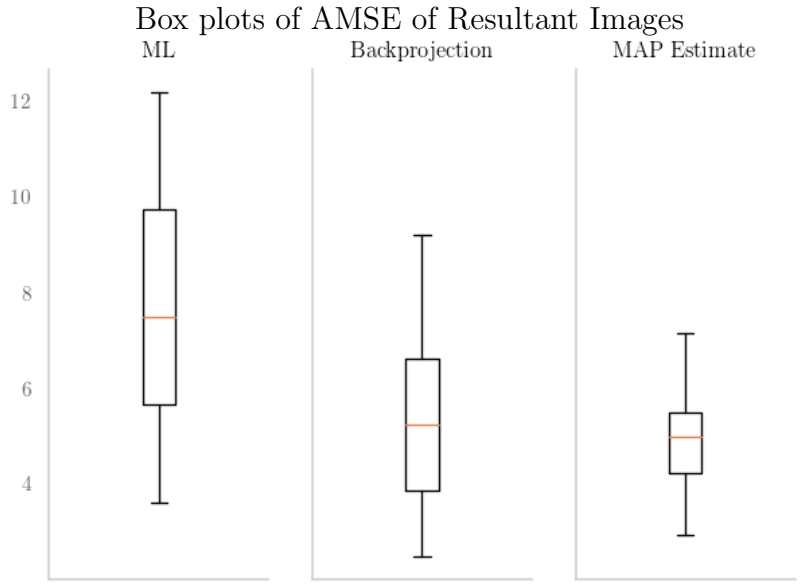

Figure 5.3: Shows boxplots of the AMSE of the formed SAR images for each method.

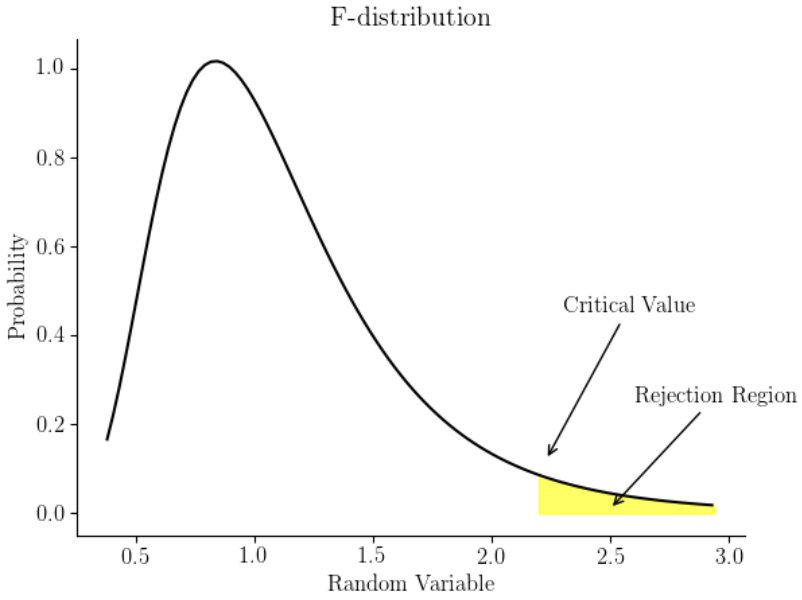

Figure 5.4: F-distribution with critical value shown and rejection region shaded. This is shown as a one-tailed test, i.e. the rejection region is only on one side of the distribution.

A test statistic compares the data to data that would be expected if the null hypothesis were true. A p-value is the probability of obtaining the data by chance, if the null hypothesis were true. The p-value goes hand in hand with a significance level which is the threshold at which the null hypothesis is rejected or accepted. This is decided on before the hypotheses test, ie. if a significance level of 0.01 is used then the null hypothesis would be rejected if the p-value obtained is less than 0.01.

An ANOVA (ANalysis Of VAriance) test is a statistical test which tests for difference among means of populations, with hypotheses of:

- ★ H_0: The means of all the population groups are equal.
- ★ H_1: The means of all the population groups are not equal.

The ANOVA test will be used to determine if the differences in the AMSE groups shown in figure 5.3 are statistically significant. Before doing this, the assumptions that an ANOVA test makes are checked, to ensure validity of the testing. The assumptions being:

- ★ samples are independent,
- ★ each sample is from a normally distributed population
- ★ the population standard deviations of the groups are all equal, (homoscedasticity).

First, an initial hypothesis test is conducted for each population, to test for normality. A Shapiro-Wilk test [68], with significance level of 0.05 is used for this. (This tests the null hypothesis that a sample set comes from a normally distributed population). For the three sets tested here, the obtained p-values are:

Method	P-value (to 3sf)
Maximum Likelihood	0.966
Maximum a Posteriori	0.275
Backprojection	0.980

As these are all well above the threshold, it is reasonable to assume that the sets are normally distributed. A similar test was also done to test that the variances are not too dissimilar, as figure 5.3 suggest that it could be the case that the variances

are too different. However, a Levene test concluded that this is not the case. Hence all the assumptions that are assumed when performing an ANOVA test are satisfied.

To perform the ANOVA test if the difference between the three sets of MSE values is statistically significant, an analysis of variance (ANOVA) was conducted. This tests the null hypothesis that the means of each of the sets are equal. Specifically here, a repeated measures ANOVA was used. For this, 20 scenes were simulated by putting 3 point scatterers at uniformly random coordinates within the scene, and Gaussian noise was added also, with a mean which was also drawn from a uniform distribution between two parameters. Each of these scenes were then used to simulate some phase history data, and the three methods (back-projection, maximum a posterior, and maximum likelihood) were used to obtain the original scene. As there were 20 samples for each group, this resulted in a statistical power of 89%. A repeated measures ANOVA was necessary as the same starting data was being used for each each method. The obtained p-value for this repeated measures ANOVA was 0.000744.

As this is smaller than 0.05, the null hypothesis is rejected. This demonstrates that the differences between the three means is statistically significant, it does not specify between which the difference occur. A Bonferroni post-hoc test was used to determine this, with significance level 0.05. The null hypothesis for this test states that the means of all the sets are equal. The test compares all the combinations of the means of each set, in order to determine in which combinations have differences. The results from this test are:

Methods Compared	P-value Adjusted (to 1sf)	Reject
Back-projection and Maximum a Posterior	0.9	False
Back-projection and Maximum Likelihood	0.000	True
Maximum a Posterior and Maximum Likelihood	0.001	True

Hence the two statistically significant differences here occur within the backprojection - ML combination and the ML-MAP combination, as indicated by the ‘p-adj’ column of p-values. This proves that there is a statistically significant improvement gained by regularising the solution, when comparing a MAP estimate to the ML estimate, tested to a significance level of 0.05.

This displays that in the presence of noise, regularising the solution (or using a prior distribution) is important in the image formation problem. It is noted however

Figure 5.5: Figures displaying simulated backprojected images of the same scene imaged using differing flight paths.

that these comparisons have been done reconstructing SAR images data using the same modeling assumptions which created the phase history data. This also indicates that the Gaussian MAP approach is equalling the performance of back projection on this test.

5.1.2 SAR Image Improvement

A similar comparison has been done using a synthetic implementaion of the image improvement problem. Synthetic images have been produced of scenes containing a few single scatterers (phase history data has been simulated and then backprojection has been used to form the starting images). For each scene, 3 scatterers were uniformly randomly placed in the scene, and the angle of the straight flight path was again drawn from a uniform distribution. Gaussian noise was also added to the images before the back-projection, with a mean which was again drawn from a uniform distribution. The set up of this simulation involves different linear flight paths flying alongside the scene, in order to obtain scenes representing a multi-look/ multi-aspect dataset. A level of noise was also added to the phase history data. Figure 5.5 displays 4 of these images, the images have been rotated and aligned to match up with eachother, the artifacts surrounding each of the scatterers is different in each image.

From this set of images, creating a master image including all the information in these is explored, by posing the problem as in equation 4.12. Both a least squares and a Tikhonov solvers are used to solve this inverse problem. Figure 5.6 displays two reconstructions of the scene, and the ground truth scene which was used to simulate the data. This shows that the Tikhonov/MAP estimate produces a scene which has less noise around the scatterers.

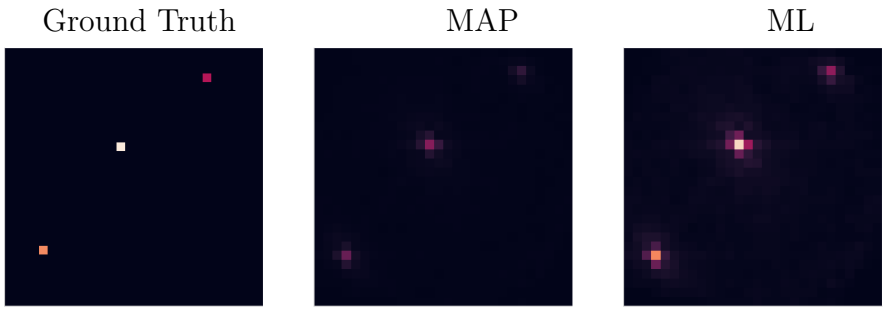

Figure 5.6: Figures showing the ground truth scene, along with MAP and ML reconstructions of the scene. Methodology follows the algebraic solutions from chapter 2

Due to the simulated nature of this, it is again possible to easily compare the error of these approaches as a ground truth is available, these images are compared to simply summing all the input scenes also. The errors of each of the master images which were created are displayed as boxplots in Figure 5.7.

Here it can be seen that the error in the MAP estimate is considerably lower that the error of the sum of the original set of images, and the ML images. This is likely due to the regularization in the presence of the noise which was added due to the phase history data; and demonstrates the importance of regularising. This difference is also tested for significance, following a similar format to above.

As a Sharipo-Wilk test (with significance level of 0.05) confirms that it is reasonable to confirm that the three populations are considered reasonable;

Method	P-value (to 3sf)
Maximum Likelihood	0.563
Maximum a Posteriori	0.494
Control	0.565

and a Levene test confirmed homoscedasticity, hence it is appropriate to perform an ANOVA to test if the difference between these three groups is statistically significant. Doing so and testing the null hypothesis, that the means of each population are equal, at a significance level of 0.05, yielded 0.00354 (to 3sf).

Hence the null hypothesis can be rejected, demonstrating that the differences between the means of the three groups is statistically significant. Again a Bonferonni post-hoc test was used to determine where the statistically significant differences lie, with 20 samples in each group.

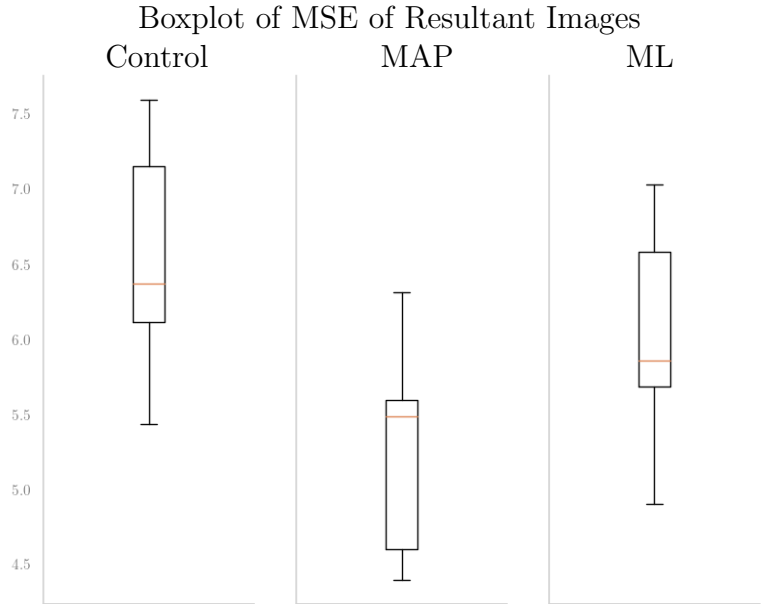

Figure 5.7: Boxplots showing the MSE of the obtained images using the MAP and ML approaches. The control plot represents the average MSE of the sum of the original set of images.

Methods Compared	P-value Adjusted (to 1sf)	Reject
Control and Maximum a Posterior	0.000	True
Control and Maximum Likelihood	0.3	False
Maximum a Posterior and Maximum Likelihood	0.04	True

It can be seen that there is a statistically significant difference between the PSNR values for the control group and the MAP group, and for the MAP group against the ML group. This indicates that the regularising / prior distribution choice has a significant effect on the amount of error in the obtained images.

This section demonstrates (using Tikhonov regularisation) for examples, the impact that regularising solutions has on the two main problems that are tackled in this thesis.

5.2 Image Improvement Forward Modelling

This section looks at some results obtained using real datasets, to display the performance of the image improvement techniques in a more realistic scenario, and to compare the performance of the forward modeling choices. In particular, images with interference (or other artifacts) are used to test against.

The Image Improvement method is applied to a real dataset, (Gotcha Volumetric Dataset [70]), with interference added synthetically. By adding interference onto clean

trials data, this is a compromise between the limited interference in purely real datasets with no ground truth, and purely simulating the data. The resultant images display interference which replicates that seen on real datasets, and this method gives an interference free reference image for objective testing of image quality improvement. There is no multi-aspect to the data used here, it is consistent with a multi-look scenario.

Like the previous experiments, the multi-image problem being considered is of the form

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_q \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_q \end{pmatrix} \mathbf{x} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_q \end{pmatrix} \tag{5.1}$$

with the goal to recover the unknown variable $\mathbf{x}$. The MAP estimate only is considered for this comparison. The identity matrix is used to model A, note that without the inclusion of a prior, which promotes lower values in $\mathbf{x}$, this would simply be the average of each of the $\mathbf{y}_i$'s. Also considered is a convolution forward operator; both of these approaches are compared in 5.2.3. The convolution operator used here is based on a Gaussian kernel; as this forward map applied to an image will have a blurring effect, the inverse of this will have a sharpening effect.

5.2.1 Synthetic Interference

Synthetic interference is added to real radar data to provide a ground truth image for objective image quality improvement metrics. Existing ground-truth free metrics do not provide robust indication of interference reduction in radar imaging. To obtain a realistic interference pattern, noise was added to the phase history signals themselves, before the image is formed via backprojection. Examples of this are displayed in Figure 5.9, where the interference pattern follows lines in the azimuth direction in the middle image which is consistent with the streaked effect seen in Figure 1.1. The right hand image in 5.9 shows an unstructured noisy image, also consistent with artifacts seen in radar imagery.

The middle image from Figure 5.8 was produced by adding noise with a variance of 95dB to the received signal from a single pulse, as displayed in Figure 5.8. The change

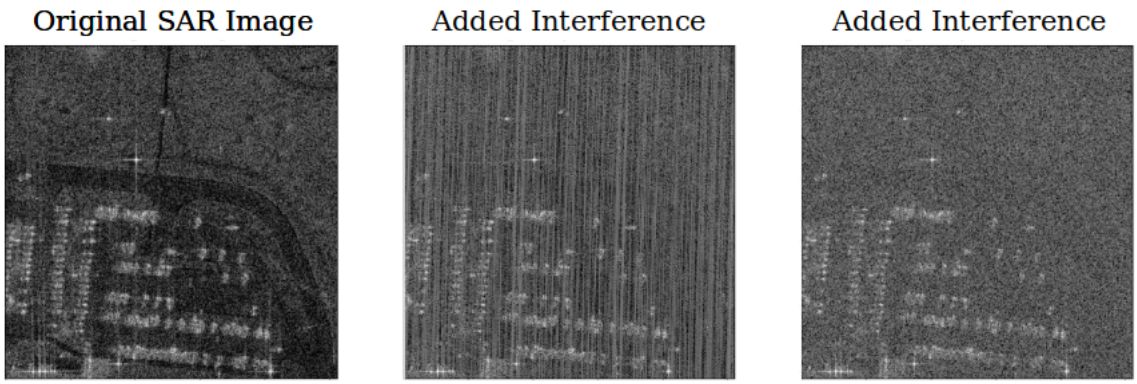

Figure 5.8: Example of synthetically-added interference pattern. The first image shows a scene with a car park surrounded by various roads (darker areas). The middle image shows the same scene but with synthetically-added interference displaying as vertical lines over the whole image. The right image shows the same scene again but with a different interference synthetically-added. In the latter two images the car park area of the scene is still identifiable, but nothing else.

in intensities on the y-axis demonstrates the level of noise added. The right hand image from Figure 5.8 was produced in a similar manner, but noise with a variance of 67dB was added to all of the collection vectors of the phase history data. The level of noise was chosen to obscure the SAR images to a similar level to what was seen in Figure 1.1, the noise levels used in the remainder of this section are similar to these numbers.

Qualitatively, the resultant images appear close to the real degraded SAR images presented earlier. Hence both techniques have been used to test the image quality improvement methods. Figure 5.9 displays a single pulse in each of these cases, both before and after filtering, which is a common step in backprojection. The ramp filtering used here is chosen due to the inclusion in the RITSAR package, [71]. Though it is not the only option for filtering, matched filtering is another common filtering technique used, [49]. Filtering in the presence of interference is not something which has been explored in this piece of work.

5.2.2 Image Results

In this section interfered images are presented with recovered images for visual review, the recovered images are again produced using the algebraic solutions from chapter 2. The results are compared more rigorously in section 5.2.3. The results presented here

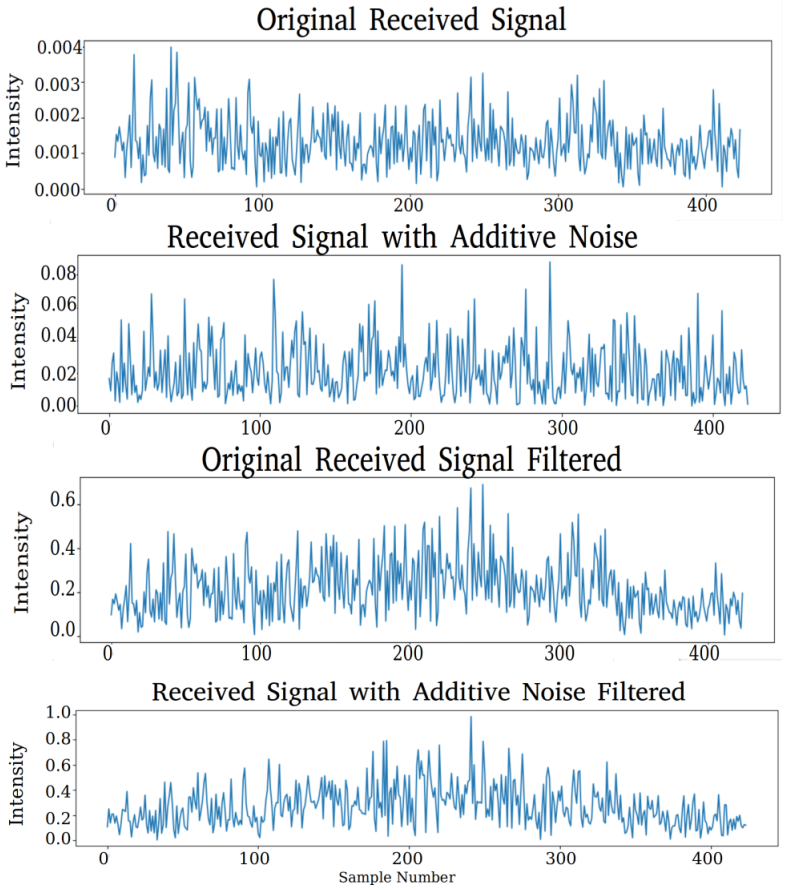

Figure 5.9: Comparison of the absolute value of phase history data for a single pulse, with and without synthetically-added interference, (filtered using a ramp filter). The intensity of the signal increases with the noise as the floor of the noise level is higher than that of the signal.

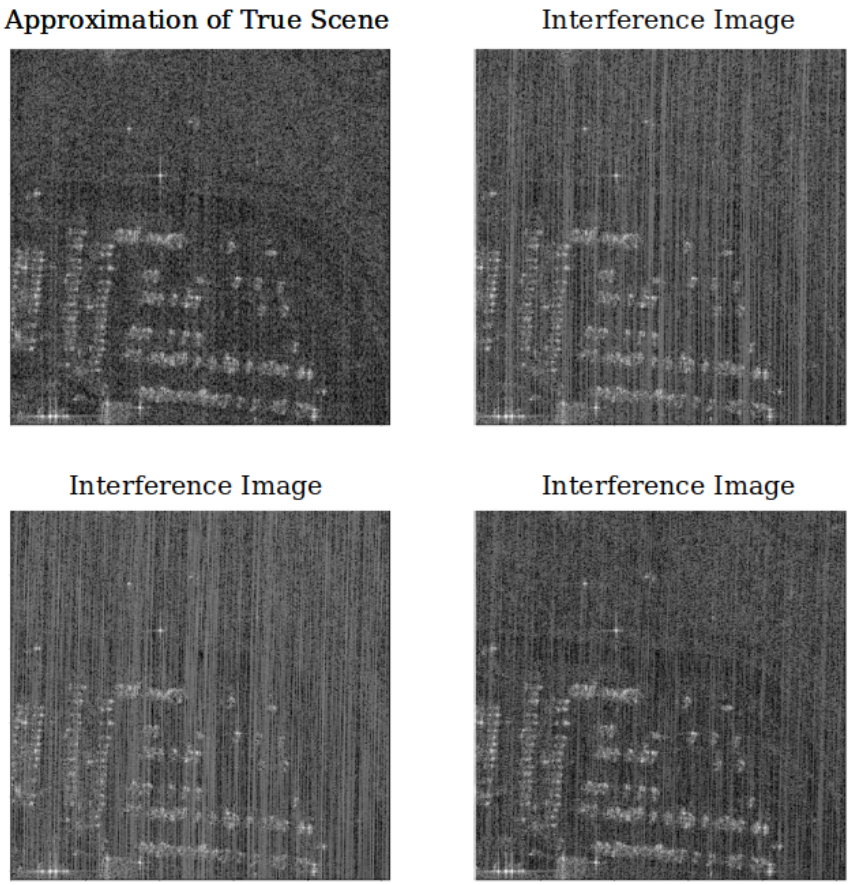

Figure 5.10: Comparison of 3 interfered images with the inferred bayesian approximation of the true scene (the original image is shown in Figure 5.8). The forward modelling operator used here is the identity operator. The images show that the recovered image has less dominant interference. The level of noise added to each image varied, hence the bottom right interfered image displays a more obscured scene.

use three interfered images, aiming to produce one master improved image from the three.

Figure 5.10 presents 3 images produced using interfered datasets, and the inferred Bayesian approximation of the true scene, (the identity operator was used as the forward map in this case). The vertical streaking is muted and more of the scene is identifiable in the recovered image, (e.g. the road). The average interference added into these images had a variance level of $95dB$ added to a single pulse for each image, the same base phase history data was used, but different pulses had the noise added to them, section 5.2.1).

Similarly, Figure 5.11 shows a recovered scene which has been obtained using the convolution operator. This also shows the vertical streaking more muted.

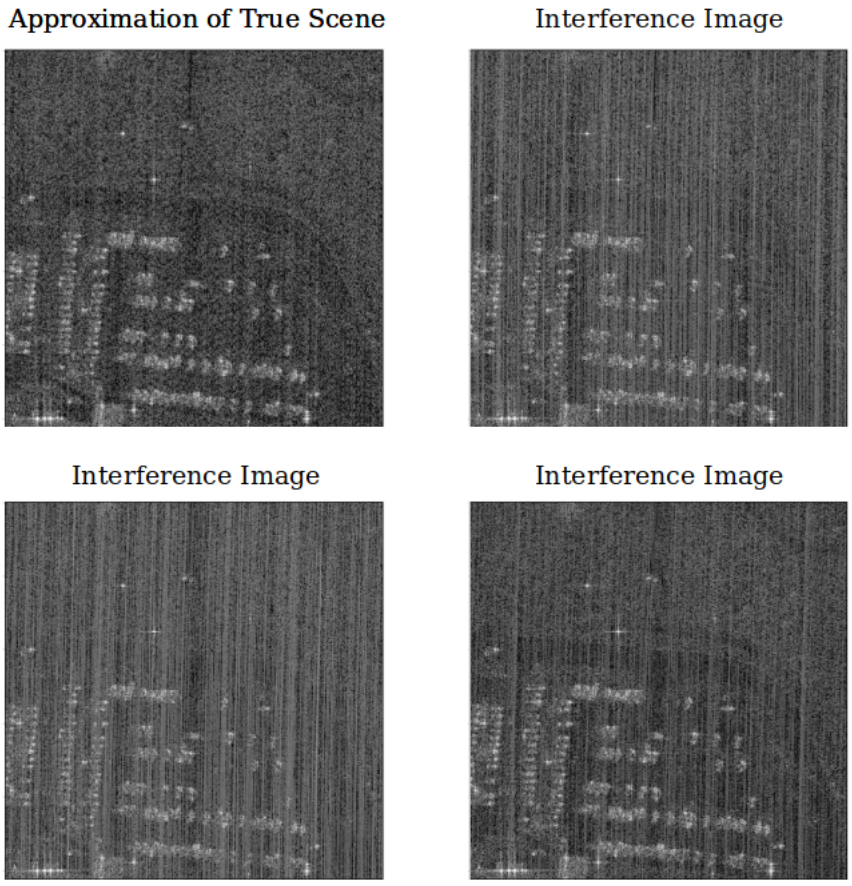

Figure 5.11: Comparison of 3 interfered images and the inferred Bayesian approximation of the true scene (the original image is shown in Figure 5.8). The forward modelling operator used is a convolution forward operator.

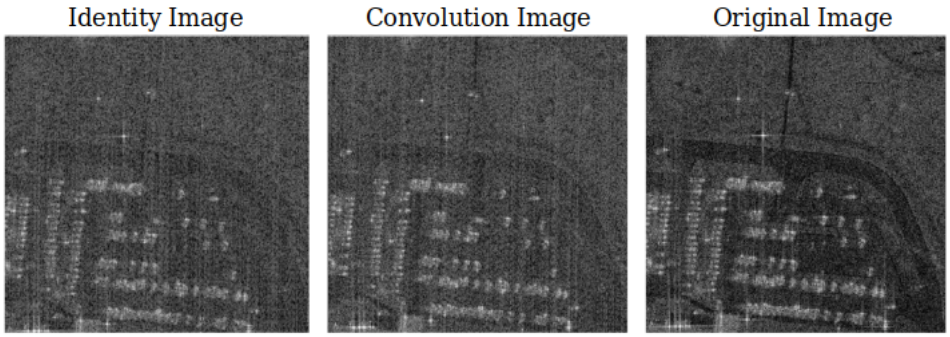

Figure 5.12: Comparison of 2 recovered images, using the different forward operators with the original SAR image.

The recovered images obtained from both the identity and convolution forward modelling approaches are compared to each other and the original SAR image in Figure 5.12. The 2 recovered images look similar, however some areas show strong differences (in particular the road coming from the top of the image). The following section makes use of an objective comparison metric to compare these further.

5.2.3 Quantitative Results

In this section, the difference between the original ground truth images and the interfered images is compared to the difference between the ground truth image and the recovered image (peak signal to noise ratio (PSNR) is used for this). PSNR is defined as

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{5.2}$$

where MSE is the mean squared error, and MAX_I is the maximum possible pixel value of the images. To compare the SAR images, each image was represented by a power image using the same scale (meaning each image had the same maximum and minimum pixel value). These images were compared to the noise-free ground truth SAR image, when also represented as a power image. Each noisy SAR image was appointed a PSNR value which represented the level of interference in that image. This metric was used to measure improvement in the SAR images (a higher PSNR means less noise in the image).

A comparison between both approaches in presented in Figure 5.13. It shows that the recovered images show improvement over the interfered images at each of the noise levels. It shows a small difference between the two methods in this PSNR plot. It

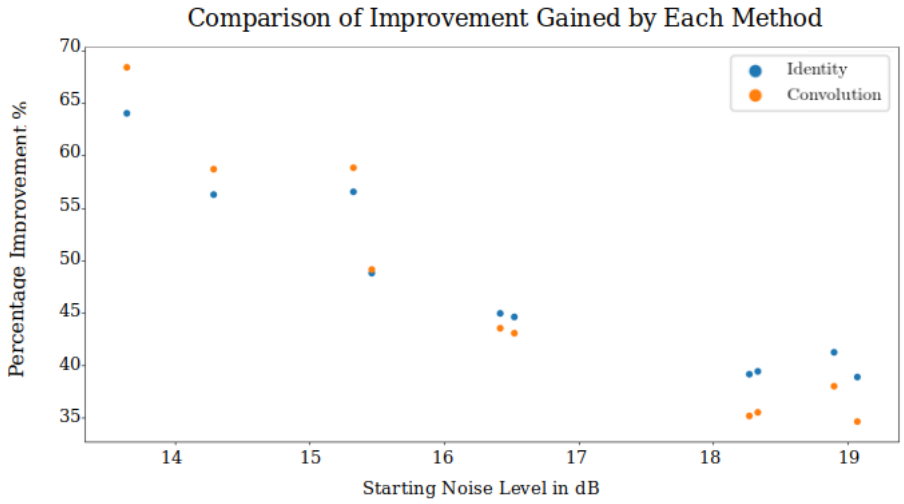

Figure 5.13: Comparing the improvement of the recovered images for forward model being the identity operator and the convolution operator. This plot suggests that the different methods may have a different impact depending on the level of noise, however specific tests would need to be conducted to confirm or reject this.

is worth noting that the convolution operator isn't necessarily expected to obtain an image which is identical to the original image, the nature of using the convolution operator and inverting means the inferred image should be sharper.

5.2.4 Statistical Comparison

To compare these two approaches further, hypothesis testing has been conducted. The null hypotheses considered for this are:

- ★ Identity operator approach has no impact on the PSNR
- ★ Convolution operator approach has no impact on the PSNR
- ★ Identity operator and convolution operator have the same impact on PSNR.

The PSNR value of the MAP estimate images from each approach is compared with the average PSNR of the interfered initial images (which is the control) for a set of 20 images. In total there were 3 sets of 20 PSNR values (for the identity operator approach, convolution operator approach and the control set). These were made by comparing 60 interfered SAR images with a noise-free ground truth image to obtain a single number representing the interference level of each image.

It was necessary to determine which distribution each set of PSNR values follow. Figure 5.14 displays a box plot of the three PSNR sets, and Figure 5.15 displays a QQ (quantile-quantile) plot of the PSNR sets. From these figures it is plausible that the sets all follow a normal distribution, the QQ plots do diverge from the normal line, however it is not significant enough cause issues.

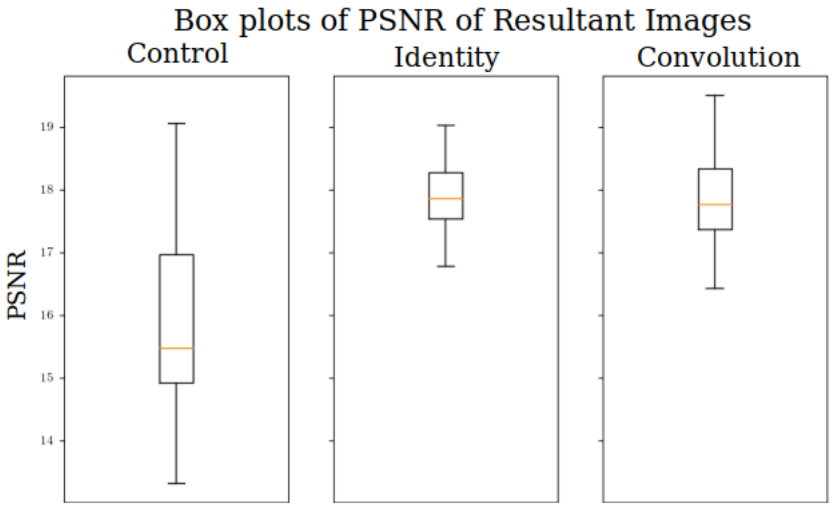

Figure 5.14: Box plot for PSNR values for control set, convolution operator approach and identity operator approach.

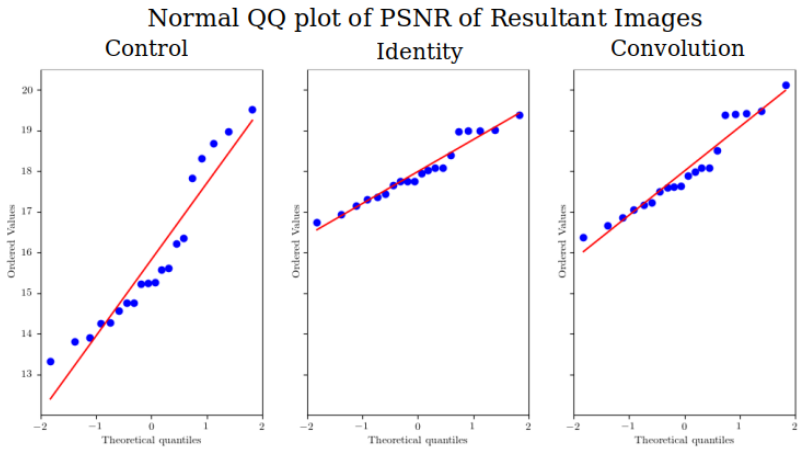

Figure 5.15: QQ plot for PSNR values for control set, convolution operator approach and identity operator approach.

To confirm the normality a Shapiro-Wilk test [68], with significance level of 0.05 is used. This tests the null hypothesis that a sample set comes from a normally distributed population. For the three sets tested here the obtained p-values were:

As all three of these values are higher than the chosen significance level, there is

Method	P-value (to 3sf)
Control	0.227
Identity	0.275
Convolution	0.228

no evidence that these sets did not come from normally distributed populations and so the null hypotheses are not rejected.

It is now the assumption that the sets of PSNR values are normally distributed. To test if there is a difference between the three sets, an analysis of variance (ANOVA) was conducted, which tests the null hypothesis that the means of each of the sets are equal. Specifically, a repeated measures ANOVA [72] is employed here (again with a significance level of 0.05), due to the same starting images being tested with each method. This results in a power level of 98%.

The p-value obtained from the repeated measure ANOVA was zero at 5 sig. figs, is lower than the significance level so the null hypothesis that the means of each set are equal can be rejected. It is now necessary to determine between which of the sets do the differences occur, a Bonferroni post-hoc test was used, with significance level 0.05. The null hypothesis for this test states that the means of all the sets are equal. The test compares all the combinations of the means of each set, in order to determine in which combinations have differences. The results from this test are:

Methods Compared	P-value Adjusted (to 1sf)	Reject
Control and Convolution	0.001	True
Control and Identity	0.001	True
Convolution and Identity	0.9	False

The p-values are lower than the significance level for two of the three possible combination, these represent the sets which contain the differences. The null hypothesis can be rejected, that the identity operator has no impact on the PSNR, and the convolution operator has no impact on the PSNR, with a significance level of 0.05. The hypothesis that the identity and convolution approaches have the same impact on the PSNR cannot be rejected. It is concluded that both the identity and convolution approaches presented provide statistically significant improvements in the images (to a significance level of 0.05), and as shown in section 5.2.2 it is concluded that the improvement is also practically significant. This demonstrates that both approaches can be considered going forward.

Chapter 6

More Bayesian Solving Approaches

Section 5 has detailed some solving approaches for inverse problems and explored their performance on simulated problems. The methodology presented so far has allowed for testing and demonstrating on some small simulated problems. However, the analytic solution to the inverse problems have not allowed for any uncertainty quantification. When obtaining a MAP estimate, some methodologies (that are described in later sections) allow the variances as well as the means of the multivariate Gaussian distribution to be retained. With this information, it is possible to quantify the uncertainty of the solution in some sense. Figure 6.1 displays the posterior distribution of a single pixel value in a reconstructed scene. This shows the spread of likely values. A posterior with a higher variance would be a lot shorter and wider.

Retaining information on the certainty of the solutions is important in the applications being considered in this thesis. The use cases of radar imaging generally involve decision makers using the information displayed by the images to help inform decisions. This end use case should be kept in mind when developing imaging techniques. For this reason having information regarding the certainty of the images produced is important.

The schemes presented so far have been using only Gaussian priors. Whilst these priors have proved to be successful in section 5, it is desirable to be able to use a wider variety of priors. For example, a Laplacian prior is thought to have stronger sparsity inducing properties than a sparse Gaussian prior. Both a Laplacian distribution and a Gaussian distribution (with zero means) are displayed in Figure 6.2. This shows that the Laplacian distribution has a taller peak at zero, but also higher tails.

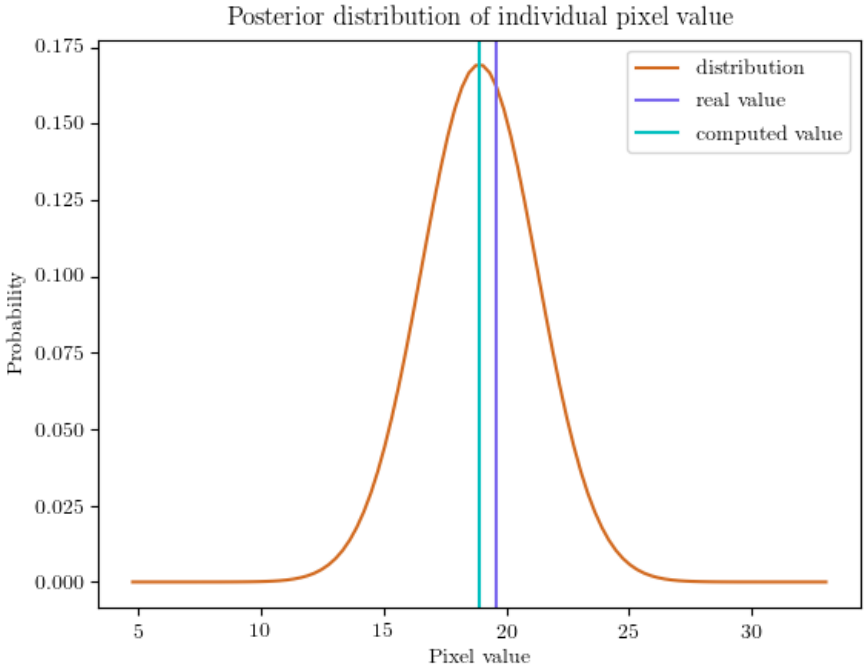

Figure 6.1: Posterior distribution of a pixel value, showing the real value and the computed value (this is the absolute value of the scene which has been transformed to be within a range of $0 - 255$.)

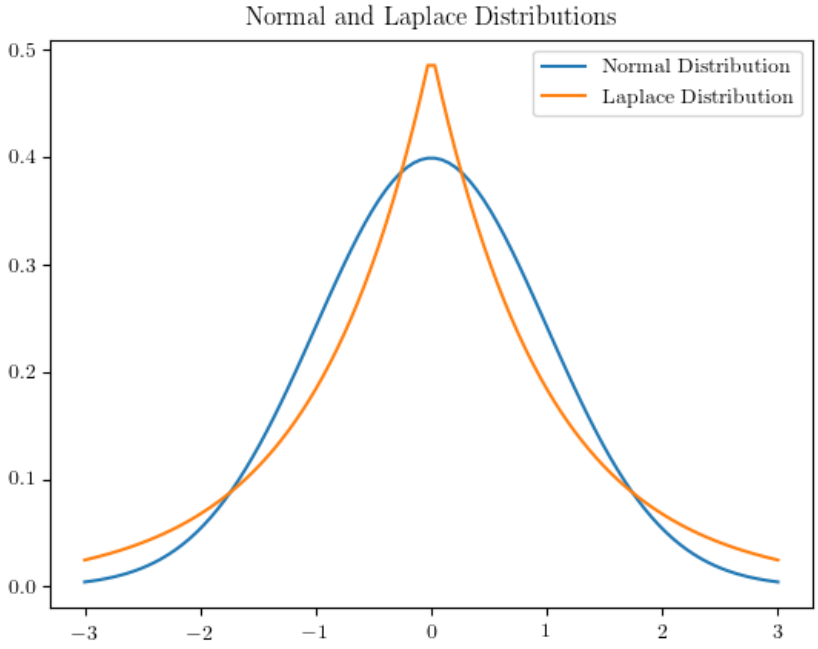

Figure 6.2: Laplacian distribution with mean zero, and a Gaussian distribution with mean zero.

Additionally it is sometimes the case where the dimensionality of a problem becomes restrictive, and it is not possible to compute the solutions (or the forward model matrices) defined in section 5 using reasonable computing power. The sizes of the forward models for image formation are largely determined by the specs of the radar, (i.e. the sampling frequency and synthetic aperture size) and this can easily be a size which is too restrictive.

It is for these reasons, that the methodology considered going forward considers the following aspects a priority:

- ★ the ability to use a variety of priors as well as Gaussian
- ★ the ability to retain the information regarding the certainty of the solution
- ★ the ability to sequentially update the solution for very large scale problems (i.e. splitting the problem into smaller problems).

6.1 Kalman

First explored is a Kalman based approach, which is an iterative solution.

In an iterative/ online approach, the solution estimate is updated sequentially, and the forward modelling matrix, and observed data are only dealt with one row at a time (or a block of multiple rows which is smaller than the total problem). Which can be written as

$$\mathbf{y}_k = A_k\mathbf{x} + \epsilon_k \tag{6.1}$$

From [73], a recursive estimator for this system can be written in the form of

$$\mathbf{x}_k = \mathbf{x}_{k-1} + K_k(\mathbf{y}_k - A_k\mathbf{x}_{k-1}), \tag{6.2}$$

where K_k is a gain matrix, which provides how much to change the current estimate for x_{k-1}. The optimal gain matrix is derived below. Given that $\epsilon \sim \mathcal{N}(0, R)$, it holds that $\mathbb{E}(\epsilon_k) = 0$.

The estimation error at a given estimate $\mathbf{x}_k$ is

$$r_k = \mathbf{x} - \mathbf{x}_k, \tag{6.3}$$

Subbing in the Equation 6.2 for x_k and rearranging then gives

$$r_k = (I - K_k A_k) r_{k-1} - K_k \epsilon_k. \tag{6.4}$$

Hence if $\mathbb{E}(\epsilon_k) = 0$, $\mathbb{E}(\epsilon_{k-1}) = 0$, and $\mathbb{E}(r_{k-1}) = 0$; then $\mathbb{E}(r_k) = 0$. This means that on average the estimate $\mathbf{x}_k$ is equal to the true value of $\mathbf{x}$ for all k, no matter what the choice of gain matrix K_k. The variance of the estimation errors can be minimised at each time k, (a K with minimal variance with be the optimum K).

$$E_k = \mathbb{E}(||\mathbf{x} - \mathbf{x}_k||^2) \tag{6.5}$$

$$= \mathbb{E}(r_k^T r_k) \tag{6.6}$$

$$= \mathbb{E}(Tr(r_k r_k^T)) \tag{6.7}$$

$$= Tr(\mathbb{E}(r_k r_k^T)). \tag{6.8}$$

$$= Tr(C_k). \tag{6.9}$$

Where $C_k = \mathbb{E}(r_k r_k^T)$ is the covariance of the estimation error r_k. Substituting in equation 6.4 to this gives

$$C_k = \mathbb{E}\left(((I - K_k A_k) r_{k-1} - K_k \epsilon_k)((I - K_k A_k) r_{k-1} - K_k \epsilon_k)^T\right) \tag{6.10}$$

Expanding these brackets, taking the expectation and omitting some terms which are equal to zero gives,

$$C_k = (I - K_k A_k) C_{k-1} (I - K_k A_k)^T + K_k R_k K_k^T, \tag{6.11}$$

Where R_k is a matrix of the covariance of the additive noise ϵ. This expression for the covariance of the estimation error is dependent on the gain matrix K. The choice of K that minimises Equation 6.5 is the optimal K. To obtain this, E_k is differentiated with respect to K_k.

$$\frac{\delta E_k}{\delta K_k} = -2 C_{k-1} A_k^T + 2 K_k (A_k P_{k-1} A_k + R_k), \tag{6.12}$$

where p_{k-1} is the covariance matrix of the previous state estimate. To find the K_k in which the derivative is zero, equation 6.12 is set to zero and solved for K_k, which gives:

$$K_k = P_{k-1}A_k^T(A_{k-1}C_{k-1}A_k^T + R_k)^{-1}. \tag{6.13}$$

This can also be substituted into equation 6.11 to obtain another expression for the covariance;

$$C_k = (C_{k-1}^{-1} + A_k^T R_k^{-1} A_k)^{-1}. \tag{6.14}$$

The optimal choice for the matrix K has been derived, which along with Equation 6.2 produces a least squares solution recursively. This has the benefit over regular least squares, that the matrices do not need to be built in full. If memory is restrictive, this means it is possible to treat the matrices line by line only.

The recursive least squares solution is also equivalent to the Kalman Filter, on a system with dynamics

$$x_k = x_{k-1} \tag{6.15}$$

$$y_k = A_k x_k + \epsilon_k. \tag{6.16}$$

This is a particular system of the general dynamical systems that Kalman Filters are commonly used for (ie. $x_k \neq x_{k-1}$). As in [73], the kalman filter is defined as

$$d_{k+1} = y_{k+1} - A_{k+1}x_{k+1} \tag{6.17}$$

$$K_{k+1} = C_{j+1}A_{k+1}^T(A_{k+1}C_{k+1}A_{k+1})^{-1} \tag{6.18}$$

$$x_{k+1} = x_k + K_{j+1}d_{k+1} \tag{6.19}$$

$$C_{j+1} = (I - K_{j+1}A_{k+1})P_{j+1}. \tag{6.20}$$

These equations are identical to the update equations for the recursive least squares solution. There also exists a Bayesian interpretation of the kalman filter solution. Considering an inverse problem, such as Equation 4.10 ($y = Ax + \epsilon$), the solution distribution for this problem using data only up to row $k-1$ is $p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1}) = \mathcal{N}(A_{k-1}\mathbf{x}_{k-1}, R_{k-1})$. This solution can be updated to include the data from the k'th row. Due to the system dynamics in equation 6.15, it is the case that

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1}) = p(\mathbf{x}_k|\mathbf{y}_{k-1}) \tag{6.21}$$

Then in order to update to include the k'th row of y, by Bayes theorem:

$$p(\mathbf{x}_k|\mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_{k-1})} \tag{6.22}$$

And hence taking the MAP of this posterior distribution gives the solution at the next time step k. The solution distribution can be updated this way, with the posterior obtained at step $k-1$ being the prior distribution used in step k.

The recursive/online solution has an advantage in memory, that the full observation matrices do not need to be made all at once, but this does comes at a cost in computation time.

This approach provides a solution to the large scale of the problems, and also retains the covariance matrix which contains any uncertainty quantification information. This achieves uncertainty quantification, and scalability for large problems. There still remains a restriction on the types of priors that can be used.

6.2 Gaussian Approximations

When dealing with Gaussian prior distributions, and assuming Gaussian error, it is possible to obtain the exact posterior in closed form.

When using a non-Gaussian prior, generally the choice is between either efficiently obtaining a MAP estimate which ignores the certainty quantification, or expensively sampling from the posterior distribution using MCMC or others. Neither of these are desirable for the use cases looked at in this thesis.

To tackle this issue, [74] proposes using a Gaussian approximation to a non-Gaussian prior, in order to be able to still retain the UQ information at a lower computational cost. This approach uses Normal scale-mixtures of generalised inverse Gaussian, along with variational Bayesian expectation maximization (VBEM). The Normal scale mixtures are defined by,

$$p(\mathbf{x}|\theta) = \prod_{j=1}^{p} N(x_j; 0, \theta_j), \tag{6.23}$$

$$p(\theta) = \prod_{j=1}^{p} GIG(\theta_j; v, \delta, \lambda), \tag{6.24}$$

where p is the size of $\mathbf{x}$. It can be shown that the marginal prior distribution on x is a Laplace distribution when $v = 1, \delta = 0$.

The proposed VBEM algorithm uses expectation maximization (EM) to minimise the Kullback-Leibler (KL) divergence between the posterior distribution and a Gaussian approximation. It can be written that,

$$\ln p(\mathbf{y}) = \ln \int p(\mathbf{y}, \mathbf{x}, \theta) \mathrm{d}\mathbf{x} \mathrm{d}\theta \geq \int \ln \left(\frac{p(\mathbf{y}, \mathbf{x}, \theta)}{q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x})} \right) q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x}) \mathrm{d}\mathbf{x} \mathrm{d}\theta, \tag{6.25}$$

using Jenson's inequality. Subsequently maximizing the term on the RHS of this inequality is equivalent to minimizing the difference between the two sides of the inequality; hence

$$\ln p(\mathbf{y}) - \int \ln \left(\frac{p(\mathbf{y}, \mathbf{x}, \theta)}{q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x})} \right) q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x}) \mathrm{d}\mathbf{x} \mathrm{d}\theta = \int \ln \left(\frac{p(\mathbf{x}, \theta | \mathbf{y})}{q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x})} \right) q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x}) \mathrm{d}\mathbf{x} \mathrm{d}\theta. \tag{6.26}$$

The term on the RHS is now exactly the definition of the KL divergence,

$$= \mathrm{KL}(q_\theta(\theta) q_{\mathbf{x}}(\mathbf{x}) || p(\mathbf{x}, \theta | \mathbf{y})), \tag{6.27}$$

which is a measure of statistical distance between distributions. Minimising this distance will hence obtain the best Gaussian approximation to the posterior distribution. This leads to objective functions which can be iteratively updated, with the Gaussian approximation of the sparsity priors included in the updates. The algorithm is defined by Algorithm 1, (taken directly from [74]).

Also presented in [74] is a sequential version of this, which can batch the data and forward models into sets of rows, following similarly to the Kalman approach seen earlier. This algorithm is defined by Algorithm 2, (again taken directly from [74]).

Hence this algorithm addresses all of the three aims mentioned earlier: non-Gaussian priors, sequential, and retaining uncertainty quantification information. For this reason, this algorithm has been applied to both the image formation and image improvement applications considered in this thesis.

Algorithm 1 Variational Bayesian N-GIG

Input: matrix A, observed data y, parameters, $\gamma, \lambda, \delta, \nu$, initial guess (μ^0, m^0, C^0), convergence criteria $\epsilon, T > 0$ and distance functional d.

1: Specify functional forms of $\theta_{VBEM}^{t+1}(\mathbf{x}), \theta_{EM}^{t+1}(\mathbf{x})$, depending on δ, ν.
2: Set $t = 0$ and initial guess all zeros.
3: **while** $t \leq T$ and $d((\mu^t, m^t, C^t), (\mu^{t-1}, m^{t-1}, C^{t-1})) > \epsilon$ **do**

(a) Compute $\theta_{VBEM}^{t+1}(\mathbf{x})$ and $\theta_{EM}^{t+1}(\mathbf{x})$

(b) Compute

$$G_{VBEM}^{t+1} = D(\theta_{VBEM}^{t+1})A^T(AD(\theta_{VBEM}^{t+1})A^T + \gamma^2 I_n)^{-1}$$

$$G_{EM}^{t+1} = D(\theta_{EM}^{t+1})A^T(AD(\theta_{EM}^{t+1})A^T + \gamma^2 I_n)^{-1}$$

(c) Compute

$$m^{t+1} = G_{VBEM}^{t+1} Y$$

$$C^{t+1} = (I_n - G_{VBEM}^{t+1})D(\theta_{EM}^{t+1})$$

$$\mu^{t+1} = G_{EM}^{t+1} Y$$

(d) $t = t + 1$

4: **end while**

Output:(μ^*, m^*, C^*)

Algorithm 2 Online Approximate Variational Bayesian N-GIG

Input: matrix A, observed data $\mathbf{y}$, parameters, $\gamma, \lambda, \delta, \nu$, initial guess (μ^0, m^0, C^0), convergence criteria $\epsilon, T > 0$ and distance functional d, batch size M.

1: Set $n = 1, \hat{A}_1 = A_1$. Do Algorithm 1, outputting (μ^*, m^*, C^*) and functional forms of $\theta^{t+1}_{VBEM}(\mathbf{x}), \theta^{t+1}_{EM}(\mathbf{x})$.

2: **for** $n = 2, ...$ **do**

3: Set $t = 0$, initial guess all zeros, and

$$\hat{y}_n^{VBEM} = \begin{pmatrix} \hat{A}_{n-1} m^*_{n-1} \\ \hat{Y}_n \end{pmatrix}, \hat{y}_n^{EM} = \begin{pmatrix} \hat{A}_{n-1} \mu^*_{n-1} \\ \hat{Y}_n \end{pmatrix}, A_n = \begin{pmatrix} \hat{A}_{n-1} \\ \hat{A}_n \end{pmatrix}$$

4: **while** $t \leq T$ and $d((\mu^t, m^t, C^t), (\mu^{t-1}, m^{t-1}, C^{t-1})) > \epsilon$ **do**

(a) Compute $\theta^{t+1}_{VBEM}(\mathbf{x})$ and $\theta^{t+1}_{EM}(\mathbf{x})$

(b) Compute

$$G^{t+1}_{VBEM} = D(\theta^{t+1}_{VBEM}) A^T (AD(\theta^{t+1}_{VBEM}) A^T + \gamma^2 I_n)^{-1}$$

$$G^{t+1}_{EM} = D(\theta^{t+1}_{EM}) A^T (AD(\theta^{t+1}_{EM}) A^T + \gamma^2 I_n)^{-1}$$

(c) Compute

$$m^{t+1} = G^{t+1}_{VBEM} Y$$

$$C^{t+1} = (I_n - G^{t+1}_{VBEM}) D(\theta^{t+1}_{EM})$$

$$\mu^{t+1} = G^{t+1}_{EM} Y$$

(d) $t = t + 1$

5: **end while**

6: Compute rank M approximation $U\Sigma^2 U^T \approx A_n A_n^T$, and set $\hat{A}_n = U^T A_n$

7: **end for**

Output:(μ^*, m^*, C^*)

Chapter 7

Results

This section presents results obtained using the methodolgy described on some real data examples. This demonstrates the effect the prior distribution has on the reconstructed images, (using a realistic testing dataset) for both the image formation, and image improvement problems. This testing was done using two datasets, GOTCHA dataset and Bright Sapphire II dataset.

When testing image formation, the starting data was phase history data, this refers raw data obtained from the radar. It contains multiple discretely sampled pulses which have traveled from the radar, to the scene, and back to the radar again.

When testing the image improvement, the starting data was formed radar images. This means that each pixel value is a complex reflectivity of what can be found at that position in the scene.

7.1 Image Formation

Using the methods described in section 6, images have been formed using the GOTCHA phase history data, which are compared to a back-projected image. Results are presented using this dataset, (figures 7.1-7.3) and a synthetically degraded version of this dataset, (figures 7.5-7.7). For each comparison there is a back-projected image, a maximum likelihood image (produced using the basic kalman filter approach), and a sparse Laplacian prior image (produced using the Gaussian approximation-kalman filter approach from section 6.2). This compare the differences between an image a model based approach obtains rather than back projection, and the impact of using a

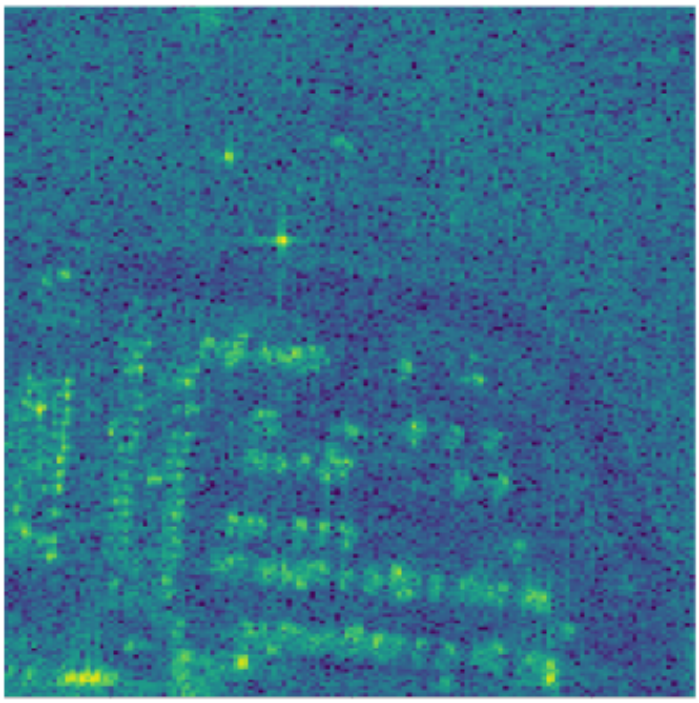

Figure 7.1: Back-projected image from Gotcha dataset.

prior.

Comparing figures 7.1-7.3, Figure 7.2 is formed with a uniform prior distribution (or no prior distribution). This means that no prior knowledge about the scene is used- that all scenes are equally likely. Figure 7.3 is formed using Laplacian distribution priors with zero mean, this type of prior enforces sparsity in the solution. This comparison allows the impact of using the prior to be seen. Specifically here, the road coming down vertically from the top of the scene is a feature which appears different in these images. Both of these images should be compared to figure 7.1, which is a back projection of the same scene, obtained using the same data. The back projected scene appears brighter due to the method of construction, however a more meaningful comparison is comparing the usable information conveyed by the image; due to the lack of ground truth and availability of no reference image quality metrics, a visual comparison was used here. For example, information such as the placement of the roads, and the cars within the scene. It can be seen that the Laplacian image has less noise than the other two images and some features can be identified clearer. This dataset is not particularly noisy or interference ridden, so the same comparison has been done using a synthetically degraded version of this dataset, to compare how well the methods can suppress the presence of extra noise or clutter in the spectrum. Figures 7.5-7.7 shows

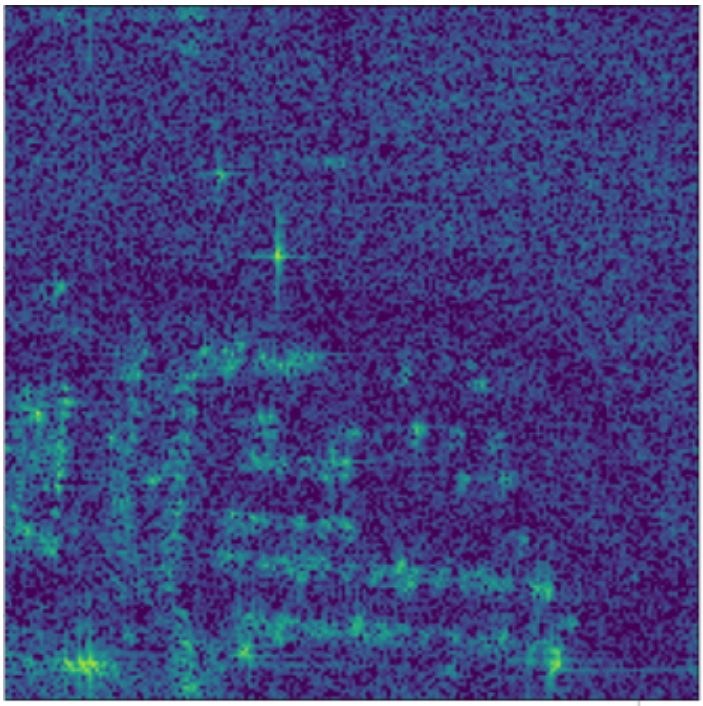

Figure 7.2: Least squares/ Maximum likelihood image from Gotcha dataset, computed with basic kalman filter approach.

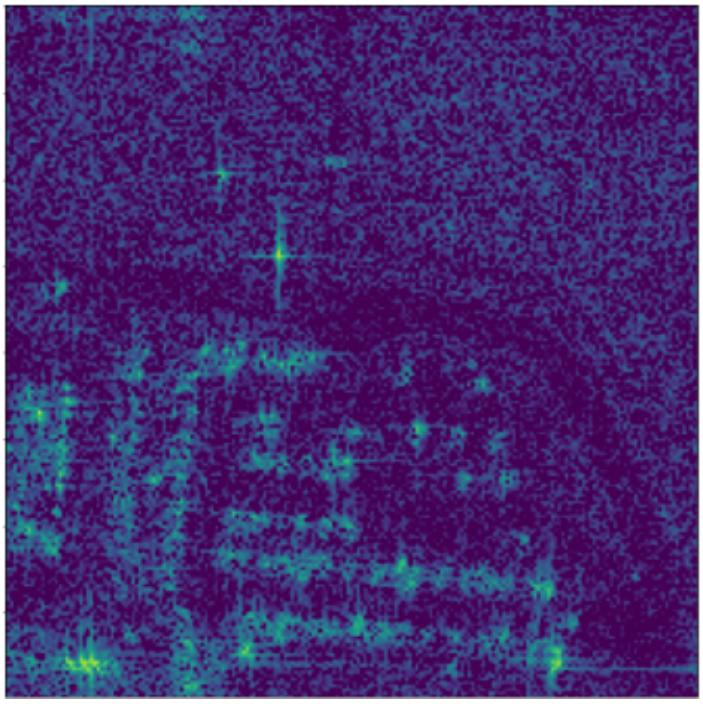

Figure 7.3: MAP estimate image from Gotcha dataset, using a Laplacian sparsity promoting prior, computed with Gaussian approximation kalman filter approach.

the SAR image constructions under the synthetic degradation. Here, it can be seen that although the cars are clearly still present in all of the images, the road going across the scene horizontally is not well identifiable in the back projected image or the Bayesian image with no prior. This displays a benefit of using a sparse Laplacian prior for image formation, as this road is distinguishable in the laplacian MAP estimate.

As well as the MAP estimate, the covariance matrix can also be obtained. This is displayed in figure 7.4, which shows the entries of the diagonal of the covariance matrix. Each entry represents the variance of the corresponding pixel. It can be seen that there is a higher variance on the pixels which have a greater value in the MAP estimate. This is due to these pixels deviating from the sparsity promoting prior, by being further away from zero.

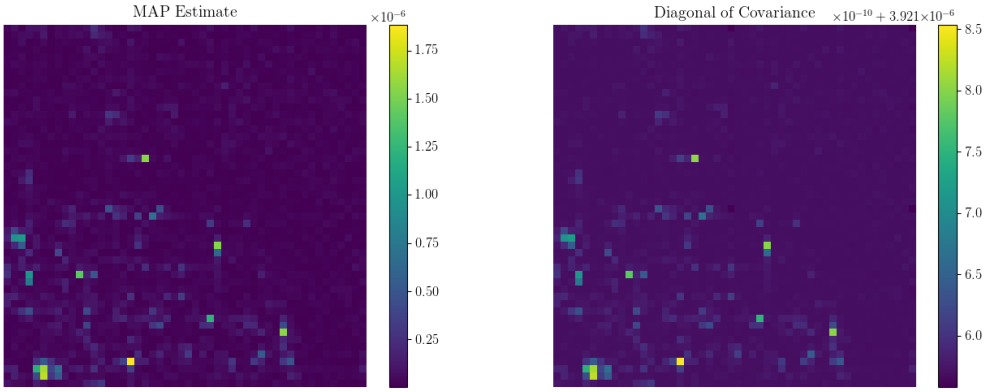

Figure 7.4: MAP estimate alongside the diagonal of the covariance matrix, displaying the UQ in the solution. Obtained using the Gaussian approximation kalman filter approach from section 6.2. The MAP estimate was using a Laplacian prior. Note that the variances are on a much smaller scale than the MAP estimate, with the $\times 10^{-10}$ being the multiplier for the colourbar indices.

Also presented here is an image formed using a stronger Laplacian prior, figure 7.8. This image has an increased level of sparsity and shows only the brightest scatterers. This doesn't display an optimal image of the scene, however it demonstrates the effect of the prior on the image formation.

It is concluded that using a sparse prior when forming SAR images can give the end user a higher level of situational awareness of the scene, particularly if there is a higher level of noise or interference present. It is however noted that using a sparse prior can preclude the use of other conventional techniques such as coherent change detection.

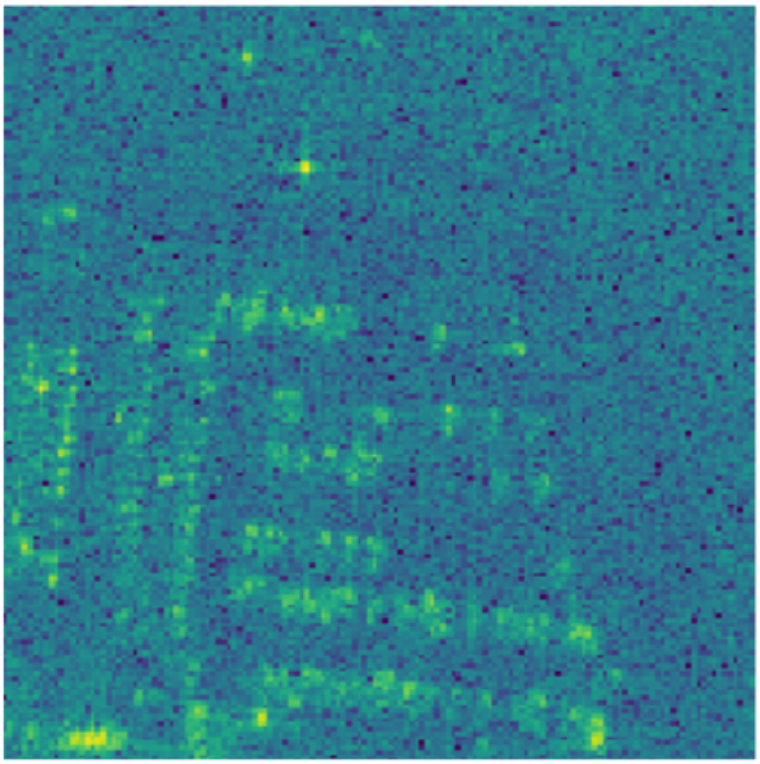

Figure 7.5: Back-projected image from Gotcha dataset with additive noise added to the phase history data.

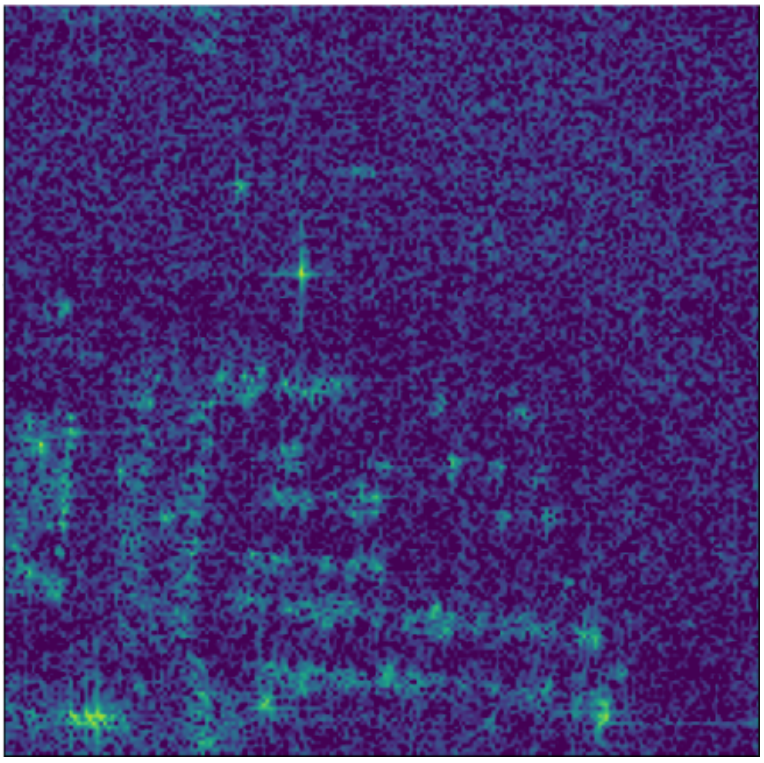

Figure 7.6: Least Squares/ Maximum Likelihood image from Gotcha dataset with additive noise, obtained from the basic kalman filter approach.

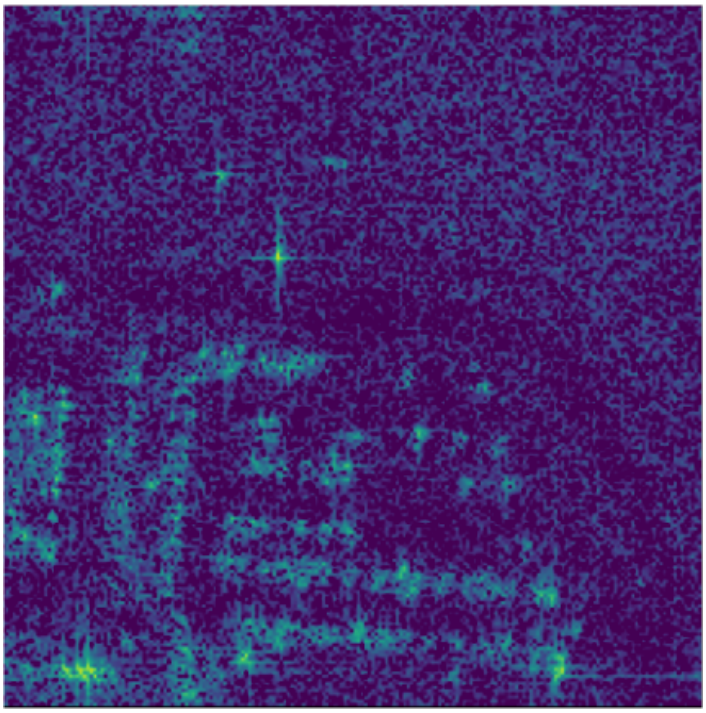

Figure 7.7: MAP estimate image from Gotcha dataset with additive noise, produced using the Gaussian approximation kalman filter approach from section 6.2. This had a sparsity promoting Laplacian prior.

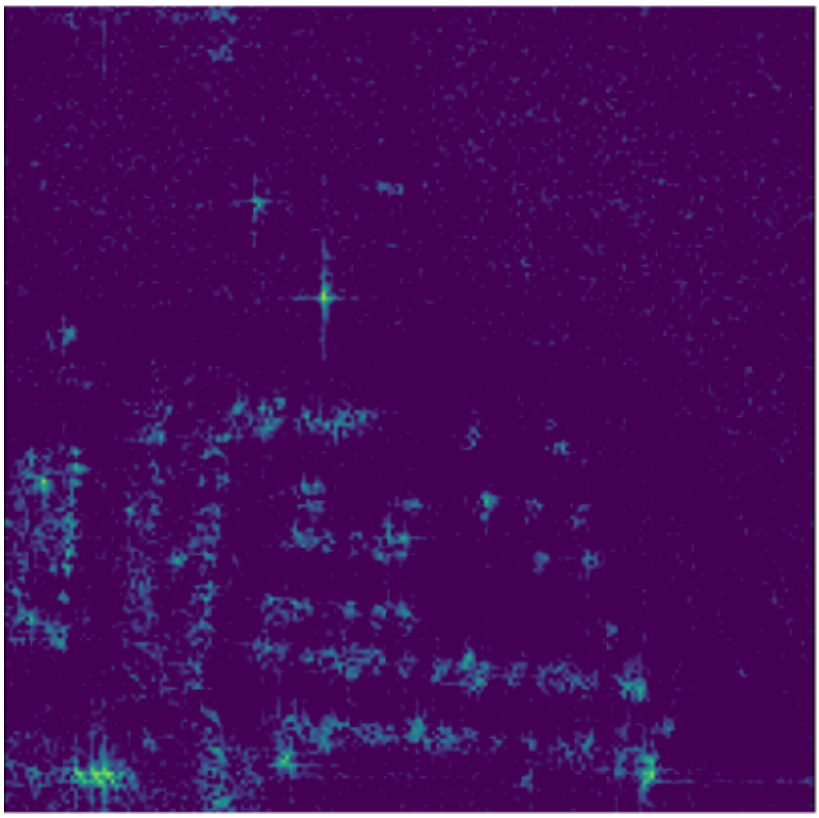

Figure 7.8: Image construction using a stonger sparsity promoting laplacian prior, on data with additive noise, again using the Gaussian approximation kalman filter approach.

7.2 Image Improvement

The results of forming an improved SAR image from a set of poor quality images is tested again against the Bright Sapphire II LF dataset, [75]. This dataset is a low frequency dataset imaged over a scene with fields, roads and houses; with a few through wall targets in the scene. BThe frequency band of this data set overlaps largely with TV broadcasting signals. Inevitably there is a high amount of interference arising in the data, resulting in some unwanted features in the resultant images. Figure 7.9 shows a few aligned images from the Bright Sapphire II dataset, formed via back projection.

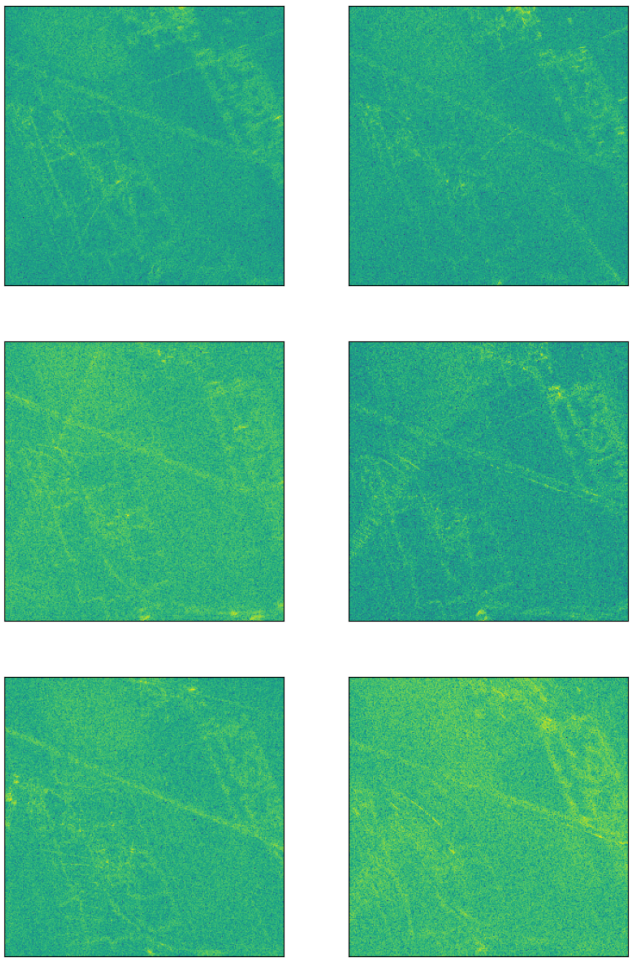

Figure 7.9: Selection of pre-aligned images from the Bright Sapphire II dataset, showing a high level of interference.

These images appear hazy, due to the low signal to noise ratio that is achieved in this part of the spectrum. The Bayesian construction of the true scene is shown in Figure 7.10. This image shows the features of the scene more prominently, and the haze covering the main scene is significantly reduced. For example, the feature at the top of the scene (this is a housing estate) is meaningfully easier to make out the shape.

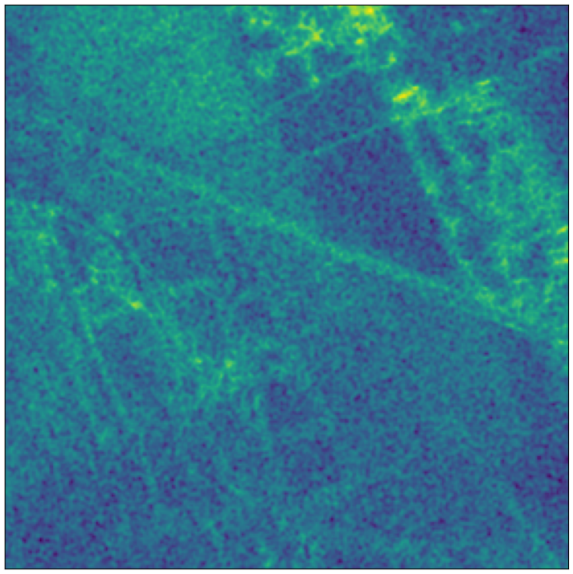

Figure 7.10: Bayesian reconstructed image from the Bright Sapphire II dataset, made using a Laplacian sparsity promoting prior.

Chapter 8

Conclusion

This thesis has described Bayesian techniques which could be applied at various stages of the radar image formation processes. Standard back-projection has been compared to maximum likelihood estimation, and maximum a priori estimation using both Gaussian and Laplacian priors. These techniques have shown promise at reducing the detrimental effects that the presence of congestion in the EM spectrum induces in radar imaging. This work is intended to act as a proof-of-concept for model-based techniques in SAR imaging in these settings, and displays potential benefits such techniques can yield in high congestion scenarios.

Two distinct strands have been researched here, image formation techniques, and image improvement techniques. Both strands have had methodology tested against synthetic datasets and real datasets. It is shown in this thesis that the inclusion of a prior to regularise the solution makes a physically and statistically significant impact on the quality of the reconstructed image.

Continuing this work would include working further on computational challenges, though these were overcome for a proof of concept work, there are still significant challenges in developing to real time. Additionally, significant robustness testing of the techniques, requiring real radar datasets with a strong ground truth. And finally, developing nonlinear (more realistic) modelling, for example modelling strong sources of electromagnetic interferes in the forward modelling..

Appendices

Appendix A

SAR Observation Matrix

This Appendix describes a discretisation of the formulation of the SAR observation matrix.

The echo signal from SAR can be expressed as:

$$S(t_1, t_2) =$$

$$\begin{cases} \sum_{k=1}^{N} \frac{x_k}{d_i^a(t_2)} \exp\left(\pi f_r \left(t_1 - 2\frac{d_i(t_2)}{c}\right)^2 i - 4\pi f_c \frac{d_i(n)}{c} i\right), & \text{if } 0 < t_1 - 2d_i(t_2) < p. \\ 0, & \text{otherwise.} \end{cases} \tag{A.1}$$

$$= \begin{cases} \sum_{k=1}^{N} x_k \phi_k(t_1, t_2), & \text{if } 0 < t_1 - 2d_i(t_2) < p. \\ 0, & \text{otherwise.} \end{cases} \tag{A.2}$$

where

- $S(t_1, t_2)$ is the signal received at slow time (aperture location) sample t_1 and fast time (frequency location) sample t_2.
- f_c is the carrier frequency.
- f_r is the LFM chirp rate.
- c is the speed of light in the medium propagating through.
- t_1 represents fast time sample number.
- t_2 represents slow time sample number.

- $d_i(n)$ is the distance between the scatterer and the plane, which is dependent on the scatter and the slow time sample number, but not the fast time due to the stop-and-go assumption.
- N is the number of scattering coefficients in the scene, as before.
- $d_i^a(t_2)$ is a factor representing the attenuation of the wave, relating directly to the distance between the scatterer and the plane.
- p is the length of the pulse emitted.
- the $0 < t_1 - 2d_i(t_2) < p$ dependence is due to the time it takes for the signal to get to and from the scatterer.

This can be written as a vector of the SAR echo signal which is our observed data $\mathbf{y}$,

$$\mathbf{y} = [S(1,1), S(1,2), ..., S(1,n_f), S(2,1), ..., S(n_s,n_f)] \tag{A.3}$$

Hence the matrix A which relates the scattering coefficients $\mathbf{x}$ and echo signal $\mathbf{y}$ by $\mathbf{y} = A\mathbf{x} + \epsilon$ is,

$$A = \begin{bmatrix} e^{i\phi_1(1,1)} & e^{i\phi_2(1,1)} & \dots & e^{i\phi_N(1,1)} \\ e^{i\phi_1(1,2)} & e^{i\phi_2(1,2)} & \dots & e^{i\phi_N(1,2)} \\ \vdots & \vdots & \dots & \vdots \\ e^{i\phi_1(1,n_f)} & e^{i\phi_2(1,n_f)} & \dots & e^{i\phi_N(1,n_f)} \\ e^{i\phi_1(2,1)} & e^{i\phi_2(2,1)} & \dots & e^{i\phi_N(2,1)} \\ \vdots & \vdots & \dots & \vdots \\ e^{i\phi_1(n_s,n_f)} & e^{i\phi_2(n_s,n_f} & \dots & e^{i\phi_N(n_s,n_f)} \end{bmatrix} \tag{A.4}$$

A is called a SAR measurement or observation matrix.

Bibliography

[1] M. Tao, J. Su, Y. Huang, and L. Wang, "Mitigation of radio frequency interference in synthetic aperture radar data: Current status and future trends," *Remote Sensing*, vol. 11, no. 20, 2019.

[2] A. Repetti, M. Pereyra, and Y. Wiaux, "Scalable bayeian uncertainty quantification in imaging inverse problems via convex optimization," *SIAM Journal on Imaging Sciences*, 2018.

[3] T. Ryberg and C. Haberland, "Bayesian inversion of refraction seismic travel-time data," *Geophysical Journal International*, 2017.

[4] T. Bodin, M. Sambridge, H. Tkalcic, P. Arroucau, K. Gallagher, and N. Rawlinson, "Transdimensional inversion of receiver functions and surface wave dispersion," *Journal of Geophysical Research Solid Earth*, 2012.

[5] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Multi-static passive sar imaging based on bayesian compressive sensing," in *Compressive Sensing III*, vol. 9109, p. 910902, International Society for Optics and Photonics, 2014.

[6] T. A., "Gaussian mixture models in pytorch," *https://angusturner.github.io/generativemodels/2017/11/03/pytorch-gaussian-mixture-model.html*, accessed 20 06 2019.

[7] M. Cetin, *Feature-enhanced synthetic aperture radar imaging*. Boston University, 2001.

[8] M. D. Desai and W. K. Jenkins, "Convolution backprojection image reconstruction for spotlight mode synthetic aperture radar," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. 505–517, 1992.

[9] Y. Yao, S. Raman, M. Schiek, A. Leff, S. Frassle, and K. Stephan, "Variational bayesian inversion for hierachical unsupervised generative embedding (huge)," *Science Direct*, 2018.

[10] S. Baillet and L. Gernero, "A bayesian approach to introducing anatomo-functional priors in the eeg/meg inverse problem," *IEEE*, 1997.

[11] N. Turkkan and T. Pham-Gia, "Algorithm as 308: Highest posterior density credible region and minimum area confidence region: the bivariate case," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 1, pp. 131–140, 1997.

[12] B. Berkin, G. Vivek, and A. Elfar, "Multi-contrast reconstruction with bayesian compressed sensing," *Magnetic Resonance in Medicine*, 2011.

[13] R. Arnold and J. Townend, "A bayesian approach to estimating tectonic stress from seismological data," *Geophysical Journal International*, vol. 170, no. 3, pp. 1336–1356, 2007.

[14] V. Kolehmainen, A. Vanne, S. Siltanen, S. Jarvenpaa, J. Kaipio, M. Lassas, and M. Kalke, "Parallelized bayesian inversion for three-dimensional dental x-ray imaging," *IEEE*, 2006.

[15] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in *2008 15th IEEE International Conference on Image Processing*, pp. 641–644, IEEE, 2008.

[16] L. Potter and J. Parker, "Sparsity and compressed sensing in radar imaging," *IEEE*, 2010.

[17] R. Raj, "Hierachical bayesian approaches to imaging and compressive sensing." ICERM Workshop, 2017.

[18] A. Bouzerdoum, F. Tivive, and V. Tang, "Multi-polarization through-the-wall radar imaging using joint bayesian compressed sensing," *University of Wollongong Australia*, 2014.

[19] Y. Huang, J. Beck, S. Wu, and H. Li, "Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery," *Science Direct*, 2016.

[20] H. C. Thom, "A note on the gamma distribution," *Monthly Weather Review*, vol. 86, no. 4, pp. 117–122, 1958.

[21] V. Roth and T. Lange, "Feature selection in clustering problems," in *Advances in neural information processing systems*, pp. 473–480, 2004.

[22] J.-p. Xu, Y.-n. Pi, and Z.-j. Cao, "Bayesian compressive sensing in synthetic aperture radar imaging," *IET Radar, Sonar & Navigation*, vol. 6, no. 1, pp. 2–8, 2012.

[23] S.-J. Wei, X.-L. Zhang, J. Shi, and G. Xiang, "Sparse reconstruction for sar imaging based on compressed sensing," *Progress In Electromagnetics Research*, vol. 109, pp. 63–81, 2010.

[24] X. X. Zhu and R. Bamler, "Tomographic sar inversion by l1 norm regularization the compressive sensing approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010.

[25] A. Achim, P. Tsakalides, and A. Bezerianos, "Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, 2003.

[26] S. Zhang, Y. Liu, X. Li, and G. Bi, "Logarithmic laplacian prior based bayesian inverse synthetic aperture radar imaging," *MDPI*, 2016.

[27] S. Zhang, Y. Liu, and X. Li, "Bayesian bistatic isar imaging for targets with complex motion under low snr condition," *IEEE*, 2018.

[28] P. Garrigues and B. A. Olshausen, "Group sparse coding with a laplacian scale mixture prior," in *Advances in neural information processing systems*, pp. 676–684, 2010.

[29] V. Tang, A. Bouzerdoum, S. Phung, and F. Tivivie, "Enhanced through-the-wall radar imaging using bayesian compressive sensing," *SPIE - Compressive Sensing II*, 2013.

[30] Q. Wu, Y. Zhang, M. Amin, and F. Ahmad, "Through the wall radar imaging based on modified bayesian compressive sensing," *IEEE*, 2014.

[31] V. Tang, S. Phung, F. Tivive, and A. Bouzerdoum, "A sparse bayesian learning approach for through-wall radar imaging of stationary targets," *IEEE*, 2017.

[32] X. Bai, Y. Zhang, and F. Zhou, "High-resolution radar imaging with unknown noise," *International Conference on Radar*, 2018.

[33] X. Zhou, H. Wang, Y. Cheng, and Y. Qin, "Off-grid radar coincidence imaging based on variational sparse bayesian learning," *Hindawi Publishing Corporation-Mathematical Problems in Engineering*, 2016.

[34] Z. Wang, W. Xie, K. Duan, and Y. Wang, "Clutter suppression algorithm based on fast converging sparse bayesian learning for airborne radar," *Signal Processing*, 2017.

[35] J. Ward, "Space-time adaptive processing for airborne radar," *Lincoln Laboratory*, 1994.

[36] M. Wicks, M. Rangaswamy, R. Adve, and T. Hale, "Space-time adaptive processing: a knowledge-based perspective for airborne radar," *IEEE*, 2006.

[37] L. Yang, L. Zhao, G. Bi, and L. Zhang, "Sar ground moving target imaging algorithm based on parametric and dynamic sparse bayesian learning," *IEEE*, 2016.

[38] S. Luo, G. Bi, X. Lv, and F. Hu, "Performance analysis on lv distribution and its applications," *Digital Signal Processing*, vol. 23, no. 3, pp. 797–807, 2013.

[39] X. Ren and L. Chen, "Four-dimensional sar imaging algorithm using bayesian compressive sensing," *Journal of Electromagnetic waves and applications*, 2014.

[40] J.-p. Xu, Y.-n. Pi, and Z.-j. Cao, "Bayesian compressive sensing in synthetic aperture radar imaging," *IET Radar, Sonar & Navigation*, vol. 6, no. 1, pp. 2–8, 2012.

[41] G. Xu, M. Xing, L. Zhang, Y. Liu, and L. Yachao, "Bayesian inverse synthetic aperture radar imaging," *IEEE*, 2011.

[42] A. P. Tran, F. Andre, and S. Lambot, "Validation of near-field ground-penetrating radar modeling using full-wave inversion for soil moisture estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5483–5497, 2013.

[43] S. Lambot and F. André, "Full-wave modeling of near-field radar data for planar layered media reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2295–2303, 2013.

[44] F. André, M. Jonard, and S. Lambot, "Non-invasive forest litter characterization using full-wave inversion of microwave radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 828–840, 2014.

[45] F. M. Watson, *Better Imaging for Landmine Detection an Exploration of 3D Full-Wave Inversion for Ground-Penetrating Radar*. The University of Manchester (United Kingdom), 2016.

[46] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *2007 IEEE radar conference*, pp. 128–133, IEEE, 2007.

[47] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 244–254, 2010.

[48] G. Rilling, M. Davies, and B. Mulgrew, "Compressed sensing based compression of sar raw data," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[49] M. Cheney and B. Borden, *Fundamentals of Radar Imaging*. Society for Industrial and Applied Mathematics, 2009.

[50] V. Churchill and A. Gelb, "Synthetic aperture radar image formation with uncertainty quantification," *arXiv preprint arXiv:2007.06380*, 2020.

[51] G. Xu, M. Xing, L. Zhang, Y. Liu, and Y. Li, "Bayesian inverse synthetic aperture radar imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 6, pp. 1150–1154, 2011.

[52] S. Zhu, A. Mohammad-Djafari, X. LI, and H. Wang, "A novel hierarchical bayesian method for sar image reconstruction," in *AIP Conference Proceedings 31st*, vol. 1443, pp. 222–229, American Institute of Physics, 2012.

[53] G. Xu, M.-D. Xing, X.-G. Xia, L. Zhang, Y.-Y. Liu, and Z. Bao, "Sparse regularization of interferometric phase and amplitude for insar image formation based on bayesian representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2123–2136, 2014.

[54] E. C. Farnett, G. H. Stevens, and M. Skolnik, "Pulse compression radar," *Radar handbook*, vol. 2, pp. 10–11, 1990.

[55] H. G. Alex Martineza, Mengqi Hua and Z. Qiao, "Born approximation, scattering, and algorithm," 2015.

[56] A. Barla, E. Franceschi, F. Odone, and A. Verri, "Image kernels," in *International Workshop on Support Vector Machines*, pp. 83–96, Springer, 2002.

[57] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton university bulletin*, pp. 49–52, 1902.

[58] D. Margalit and J. Rabinoff, "Interactive linear algebra," in *Available: https://textbooks.math.gatech.edu/ila/least-squares.html, year=Accessed: 2021-09-30,.*

[59] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM journal on matrix analysis and applications*, vol. 21, no. 1, pp. 185–194, 1999.

[60] D. A. Belsley, E. Kuh, and R. E. Welsch, "The condition number," *Regression diagnostics: Identifying influential data and sources of collinearity*, vol. 100, p. 104, 1980.

[61] B. Edde, "Radar-principles, technology, applications," *NASA STI/Recon Technical Report A*, vol. 93, p. 14697, 1993.

[62] S. Ellis, "Instability of least squares, least absolute deviation and least median of squares linear regression," *Statistical Science*, vol. 13, no. 4, 1998.

[63] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for pet," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, 1993.

[64] A. Charnes, E. L. Frome, and P.-L. Yu, "The equivalence of generalized least squares and maximum likelihood estimates in the exponential family," *Journal of the American Statistical Association*, vol. 71, no. 353, pp. 169–171, 1976.

[65] M. T. West D., Gunther J., "Inverse problems-based maximum likelihood estimation of ground refelctivity for selected regions of interest from stripmap sar data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, 2008.

[66] A. Wirgin, "The inverse crime," *arXiv preprint math-ph/0401050*, 2004.

[67] L. Wasserman, *All of statistics : a concise course in statistical inference*. New York: Springer, 2010.

[68] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Oxford University Press*, 1965.

[69] H. Abdi and L. Williams, "Tukey's honestly significance difference (hsd) test," *Encyclopedia of Research Design*, 2010.

[70] AFRL, "Gotcha volumetric sar data set," *https://www.sdms.afrl.af.mil/index.php?collection=gotcha*, accessed 02/02/2021.

[71] D. McDonald, "Ritsar," in *Available: https://github.com/dm6718/RITSAR*, 2016. Accessed 02/02/2021.

[72] E. Girden, *ANOVA: Repeated measures*. No. 84, Sage, 1992.

[73] K. Law, A. Stuart, and K. Zygalakis, "Data assimilation," *Cham, Switzerland: Springer*, vol. 214, 2015.

[74] K. J. Law and V. Zankin, "Sparse online variational bayesian regression," *arXiv preprint arXiv:2102.12261*, 2021.

[75] S. Doody, N. Hughes, L. Ramio-Tomas, and E. Mak, "Bright sapphire- lf sar imagery- a first look," *Airbus Defence and Space Ltf, UK*, 2017.