

Intelligent Imaging of Perfusion Using Arterial Spin Labelling

David Owen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Centre for Medical Image Computing, Department of Medical Physics and
Biomedical Engineering
University College London

February 24, 2020

I, David Owen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Arterial spin labelling (ASL) is a powerful magnetic resonance imaging technique, which can be used to noninvasively measure perfusion in the brain and other organs of the body. Promising research results show how ASL might be used in stroke, tumours, dementia and paediatric medicine, in addition to many other areas. However, significant obstacles remain to prevent widespread use: ASL images have an inherently low signal to noise ratio, and are susceptible to corrupting artifacts from motion and other sources. The objective of the work in this thesis is to move towards an “intelligent imaging” paradigm: one in which the image acquisition, reconstruction and processing are mutually coupled, and tailored to the individual patient.

This thesis explores how ASL images may be improved at several stages of the imaging pipeline. We review the relevant ASL literature, exploring details of ASL acquisitions, parameter inference and artifact post-processing. We subsequently present original work: we use the framework of Bayesian experimental design to generate optimised ASL acquisitions, we present original methods to improve parameter inference through anatomically-driven modelling of spatial correlation, and we describe a novel deep learning approach for simultaneous denoising and artifact filtering. Using a mixture of theoretical derivation, simulation results and imaging experiments, the work in this thesis presents several new approaches for ASL, and hopefully will shape future research and future ASL usage.

Impact Statement

In an academic context, the work presented herein offers three principal benefits within the area of ASL imaging. First, the novel experimental design approach for optimising ASL acquisitions offers state of the art results compared to previous comparable work, and is generally applicable to different ASL acquisitions. Moreover, this work has helped to revive the idea of experimental design for ASL acquisitions, as well as presenting a more extensive validation across different ASL acquisitions. Second, the use of shrinkage priors offers a step forward in ASL image processing generally, and may be applied to any of the popular models used for inference. Third, the use of a convolutional neural network for joint denoising and artifact removal presents an exciting new direction in imaging research, and is among the earliest such work. Much of this work has impact in academia outside the specific area of ASL imaging. The experimental design approach developed herein is generally applicable to imaging modalities with a tractable signal and suitable noise model, and has already been used in abdominal imaging and experimental multi-modal acquisitions in both human and animal experiments. The shrinkage prior approach bridges the gap between the emerging field of probabilistic programming – in which multi-level regression is a canonical technique for data analysis – and more traditional numerical methods for Bayesian inference on otherwise-intractable models. Finally, the convolutional neural network approach to joint denoising and artifact removal is generally ap-

plicable to other imaging modalities, including natural images. Finally, there is potential for impact outside academia. Most obviously, many of the methods developed in this work may, one day, be suitable for clinical use. Optimised acquisitions, improved fitting and improved filtering of artifacts all have a place in clinical ASL (as well as in other imaging modalities). Parts of the work in this thesis contributed to the open-source software, NiftyFit, a package for multi-modal MRI fitting.

Acknowledgements

Karolina, I cannot express how grateful I am for your generosity, patience and love.

My heartfelt thanks go to my wonderful supervisors. Andrew, thank you so much for seeing me through this. Dave, I can't believe in retrospect how lucky I was to end up with you as a supervisor. Jon, thank you so much for always pushing my work forward, getting me data, and keeping the work grounded. Seb, thank you for bringing me to London, persuading me to do a PhD, and creating this wonderful environment for research. I must also give an honorary mention to Enrico, who gave significant input on the optimal design work.

Thanks to my cherished peers, collaborators and friends: Michael, Catherine, Zach, Alex, Carla, Michela, and everyone else at TIG and CMIC.

Thanks to my family, who have supported me through years of conference deadlines, flying visits and thesis writing: Mum, Dad, Alan; Helen, Katy, Louise; Grandad.

Thanks to my non-imaging friends, who bore with me as I attempted this mad task and basically ignored them for the past six months: George, Bob, Jon, Angela, Josh, Merlyn, Michael, Oli, Martin, Mo, Laura, Will, Friend.

I also thank those who, sadly, had to leave: Nan, Grandma, Sid, Joan, Jenny.

Finally, thanks to the Wolfson Foundation and EPSRC for funding my work.

Publications and abstracts

As first author

- D. Owen, A. Melbourne, Z. Eaton-Rosen, D. L. Thomas, N. Marlow, J. Rohrer, and S. Ourselin. Deep convolutional filtering for spatio-temporal denoising and artifact removal in arterial spin labelling MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 21–29. Springer, 2018
- D. Owen, A. Melbourne, Z. Eaton-Rosen, D. L. Thomas, N. Marlow, J. Rohrer, and S. Ourselin. Anatomy-driven modelling of spatial correlation for regularisation of arterial spin labelling images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 190–197. Springer, 2017
- D. Owen, A. Melbourne, D. Thomas, J. Beckmann, J. Rohrer, N. Marlow, and S. Ourselin. ADRIMO: Anatomy-DRiven MOdelling of spatial correlation to improve analysis of arterial spin labelling data. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*. ISMRM, 2017
- D. Owen, A. Melbourne, M. Sokolska, D. Thomas, J. Rohrer, and S. Ourselin. Bayesian experimental design for multi-parametric T1/T2 relaxometry and diffusion. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*. ISMRM, 2017
- D. Owen, A. Melbourne, D. Thomas, E. De Vita, J. Rohrer, and S. Ourselin.

Optimisation of arterial spin labelling using Bayesian experimental design. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*. Springer, 2015

As a coauthor

- A. Melbourne, R. Aughwane, M. Sokolska, D. Owen, G. Kendall, D. Flouri, A. Bainbridge, D Atkinson, J. Deprest, T. Vercauteren, et al. Separating fetal and maternal placenta circulations using multiparametric MRI. *Magnetic resonance in medicine*, 81(1):350–361, 2019
- L. Smith, A. Melbourne, D. Owen, M. J. Cardoso, C. Sudre, T. Tillin, M. Sokolska, D. Atkinson, N. Chaturvedi, S. Ourselin, A. Hughes, F. Barkhof, and R. Jäger. Cortical cerebral blood flow in aging: Effects of haematocrit, sex and ethnicity. *European Radiology*, 2019
- R. Pratt, A. Melbourne, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, J. Deprest, G. Kendall, T. Vercauteren, S. Ourselin, et al. Spatial vascular heterogeneity in the normal placenta assessed with multicompartement placental mri. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France*. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018
- L. Smith, A. Melbourne, D. Owen, MJ. Cardoso, C. Sudre, T. Tillin, M. Sokolska, D. Atkinson, N. Chaturvedi, S. Ourselin, et al. Cortical cerebral blood flow in aging: Effects of haematocrit, sex and ethnicity. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France*. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018
- A. Melbourne, R. Pratt, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, J. Deprest, G. Kendall, T. Vercauteren, A. David, et al. Placental insufficiency investigated with multi-compartment placental MRI. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance*

in Medicine, Paris, France. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018

- R. Pratt, A. Melbourne, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, S. Ourselin, et al. Novel placental evaluation using multimodal MRI. *Ultrasound in Obstetrics & Gynecology*, 50(S1):62–62, 2017
- A. Melbourne, R. Pratt, M. Sokolska, D. Owen, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. Separation of fetal and maternal circulations using multi-modal MRI. *Placenta*, 57:291, 2017
- A. Melbourne, R. Pratt, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. DECIDE: Diffusion-rElaxation combined imaging for detailed placental evaluation. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii. ISMRM, 2017*
- A. Melbourne, R. Pratt, D. Owen, M. Sokloska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. Placental image analysis using coupled diffusion-weighted and multi-echo T2 MRI and a multi-compartment model. In *Medical Image Computing and Computer-Assisted Intervention: Perinatal, Preterm and Paediatric Image Analysis workshop–MICCAI 2016. MICCAI, 2016*
- A. Melbourne, N. Toussaint, D. Owen, I. Simpson, T. Anthopoulos, E. De Vita, D. Atkinson, and S. Ourselin. NiftyFit: a software package for multi-parametric model-fitting of 4D magnetic resonance imaging data. *Neuroinformatics*, 14(3):319–337, 2016
- A. Melbourne, Z. Eaton-Rosen, D. Owen, M.J. Cardoso, J. Beckmann, D. Atkinson, N. Marlow, and S. Ourselin. Measuring cortical neurite-dispersion and perfusion in preterm-born adolescents using multi-modal

MRI. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 378–386. Springer, 2015

Contents

1	Introduction	15
1.1	Context and background	15
1.2	Challenges in ASL	15
1.3	Aims and scope	16
1.4	Structure of the thesis	16
2	Background and literature review	18
2.1	Perfusion and blood flow	18
2.1.1	The role of blood and blood flow in the body	18
2.1.2	Homeostasis, blood flow regulation and cerebral blood flow	19
2.1.3	Blood flow imaging methods	20
2.1.4	Applications of perfusion imaging	24
2.2	ASL: MR physics and acquisitions	26
2.2.1	MR imaging	26
2.2.2	Physics of the ASL signal and the Buxton kinetic model . .	27
2.3	ASL: reconstruction, fitting and post-processing	34
2.3.1	ASL pipeline and fitting	34
2.3.2	Error and bias	34
2.3.3	Partial volume	35
2.3.4	Spatial regularisation	36
2.3.5	Motion correction	36
2.3.6	Artifact filtering	38
2.3.7	Inference and machine learning in ASL	38

- 2.3.8 Beyond the Buxton model 39
- 2.4 Optimal experiment design 40
 - 2.4.1 Locally optimal design 40
 - 2.4.2 Robust/Bayesian optimal design 41
 - 2.4.3 Approximate Bayesian optimal designs for Gaussian ap-
proximations 44
 - 2.4.4 Optimal experiment design for medical imaging 46
 - 2.4.5 Optimal experiment design in other MR contrasts 46
 - 2.4.6 Optimal experiment design for ASL acquisition parameters . 46
- 2.5 Optimal design for ASL: overview and insight 48
 - 2.5.1 Translation of optimal design to ASL 48
 - 2.5.2 Priors 49
 - 2.5.3 Measurement constraints: fixed number of measurements
versus fixed duration 53
 - 2.5.4 Optimisation for experimental design 53
 - 2.5.5 PASL versus PCASL 55
- 2.6 Conclusion 56

3 Bayesian experimental design for multi-parametric optimisation of ASL 57

- 3.1 Introduction 57
- 3.2 Theory 59
 - 3.2.1 Arterial spin labelling 59
 - 3.2.2 Bayesian design theory 59
 - 3.2.3 Computationally-tractable optimal design solutions 61
 - 3.2.4 Constrained optimal design 62
- 3.3 Material and methods 62
 - 3.3.1 Experimental design optimisation 62
 - 3.3.2 Variable TR acquisition 63
 - 3.3.3 Specific absorption rate modelling 64
 - 3.3.4 Synthetic data 66

3.3.5	Experimental data	68
3.4	Results	69
3.4.1	Proposed designs	69
3.4.2	Synthetic results	71
3.4.3	Experimental results – PASL	75
3.4.4	Experimental results – PCASL	76
3.4.5	Robustness to abnormal parameter values	78
3.5	Different prior distributions and their effects on design optimisation	81
3.6	Discussion and conclusions	82
4	ADRIMO: Anatomy-DRiven MOdelling of spatial correlation	89
4.1	Introduction	89
4.2	Denoising ASL	90
4.3	Material and methods	92
4.3.1	Arterial spin labelling	92
4.3.2	Theory and implementation	92
4.3.3	Comparison methods	95
4.3.4	Validation for representative patterns of typical perfusion . . .	97
4.3.5	Validation in the presence of focal perfusion changes	100
4.4	Results	101
4.4.1	Typical perfusion	101
4.4.2	Focal perfusion changes	108
4.5	Discussion	110
4.5.1	Increased accuracy of perfusion estimates in representative simulations	113
4.5.2	Significant experimental evidence of improved accuracy in typical perfusion maps	113
4.5.3	Sustained performance in the presence of focal perfusion changes	114
4.5.4	Limitations and future work	115
4.5.5	Conclusion	116

5	Deep convolutional filtering for spatio-temporal denoising and artifact removal	117
5.1	Introduction	117
5.2	Methods	118
5.2.1	Arterial spin labelling	118
5.2.2	Denoising and artifact removal in natural images and medical images	119
5.2.3	Deep convolutional joint filtering for ASL	122
5.2.4	Comparison to pre-existing methods and validation	124
5.3	Results	126
5.3.1	Example images	126
5.3.2	Quantitative evaluation via PSNR	128
5.4	Discussion and conclusions	129
6	Conclusions	133
6.1	Experimental design	133
6.2	Anatomy-driven modelling for spatial regularisation	136
6.3	Deep learning for joint denoising and artifact filtering	139
6.4	Conclusion	142
	Appendices	143
A	Analytic solutions for the optimal design problem of Xie <i>et al</i>	143
B	Software Packages	145
C	Acronyms and abbreviations	146
	List of figures and tables	149
	Bibliography	156

Chapter 1

Introduction

1.1 Context and background

Arterial spin labelling (ASL) is a powerful magnetic resonance imaging technique, which can be used to noninvasively measure perfusion in the brain and other organs of the body. Promising results from both research and clinical use show how ASL might be used in stroke, tumours, dementia and paediatric medicine, in addition to many other areas. Medical imaging methods to map perfusion have been an active area of research for decades, but there are many practical barriers to the widespread use of perfusion imaging. ASL, because it does not require the injection of an exogenous contrast agent, has the potential to overcome these barriers and possibly even to bring perfusion imaging to common clinical practice, as well as expanding its role in research. At the time of writing, the UK Biobank project – in which 100,000 participants will be scanned over a five year period – is piloting ASL protocols, reflecting the excitement as ASL becomes increasingly reliable and informative [17]. However, significant obstacles remain to prevent widespread use of ASL: images have an inherently low signal to noise ratio (SNR), and are additionally prone to corrupting artifacts from motion and other sources.

1.2 Challenges in ASL

ASL has an inherently low SNR. This is discussed further in Chapter 2, but essentially the reason for this is that only a small amount of blood flows into the brain over the timescale on which ASL images are acquired. Because of this poor SNR,

numerous techniques exist to improve ASL image quality, ranging from modified acquisitions to improved parameter estimation methods. It remains challenging, however, to reliably get a high quality image in a practical scan duration.

In addition, but distinct from the low SNR, ASL is highly prone to corruption by subject motion, scanner coil instability, boundary artifacts, and other difficult to model sources of spurious signal. Consequently, there is another key challenge in ASL: how to reliably get a signal that reflects the underlying perfusion rather than being dominated by artifacts, which may be achieved either through avoiding artifacts in the first place or by using post-processing to remove them.

1.3 Aims and scope

The aims of this thesis are the following: to survey the challenges of ASL in depth, including pre-existing solutions; to present original work that attempts to address these challenges; and to validate, both in simulations and *in vivo* experiments, our techniques. We approach these challenges within the paradigm of “intelligent imaging”; this emphasises that image acquisition, reconstruction and processing are strongly coupled, and tailoring these to the individual patient can allow synergistic improvement across all areas. In particular, we explore three different innovations in ASL: improved acquisitions using optimal experiment design (hence improving the overall measurement SNR for a given scan duration), anatomy-driven modelling for spatial regularisation of the ASL signal (improved parameter estimation to increase the effective SNR), and deep learning based filtering for simultaneous denoising and artifact reduction (improving the effective SNR again, but more importantly addressing the problem of difficult-to-process artifacts).

1.4 Structure of the thesis

Chapter 2 reviews several areas of the literature and provides detailed background information for the rest of this thesis. First, we provide a comprehensive literature review of ASL, with contextual information about perfusion imaging in general. We discuss the motivations and applications behind perfusion imaging, the physical principles behind ASL, and the information processing techniques underpinning

ASL images. In Sections 2.1–2.2 we present the history of ASL, with notable “first” achievements and methodological advances focused on the underlying physics. We then review the ASL processing pipeline in Section 2.3, focusing on spatial regularisation (discussed at length in Chapters 4 and 5) and artifact removal (Chapter 5). In Section 2.4, we proceed to give a brief literature review of optimal experiment design, the field of statistics devoted to providing maximally informative experiments. After this, in Section 2.5, we discuss the history of optimal experiment design in ASL and other MR imaging modalities, setting the stage for our original work on experimental design in Chapter 3.

Chapter 3 contains the main contributions of this thesis within the area of optimal experiment design. We present our optimal design approach, including simulations and *in vivo* experiments. This approach yields statistically and practically significant improvements in image quality, and marks a step forward in ASL optimal design. We explain the motivation for our approach, its origins in preceding work, and how it has influenced subsequent research on optimal design.

Chapter 4 presents original work on image spatial regularisation to improve the effective SNR in ASL. We first derive a hierarchical Bayesian model approach for data-driven spatial regularisation of ASL images. Subsequently we show the benefits of this anatomy-driven hierarchical prior through several simulations and experiments. We show that this approach is robust to extreme focal changes in perfusion, as well as substantially improving the quality of more typical ASL images.

Chapter 5 presents original work on the problem of artifacts in ASL images. We present a novel deep learning approach for simultaneous spatio-temporal denoising and artifact removal, exploiting the synergy in these two related problems. This work is one of the first to use deep learning in ASL image processing, and offers an intriguing contrast to more traditional information processing approaches.

Finally, Chapter 6 offers a conclusion to the thesis, summarising the work, recapping its context in the field, and discussing potential future directions. Appendices B–C contain a presentation of software implementations of our methods and a glossary of acronyms and abbreviations.

Chapter 2

Background and literature review

2.1 Perfusion and blood flow

2.1.1 The role of blood and blood flow in the body

The blood plays several crucial roles in the body: it transports oxygen (bound to haemoglobin), used in aerobic respiration; it carries nutrients around the body; it facilitates signalling in several different systems (for example the immunological and endocrine systems); it allows the removal of waste products from organs and tissues; and its use for heat exchange with the environment is fundamental to the body's temperature regulation [18, 19]. For all of these purposes, in human physiology (and mammalian physiology more generally) the blood is circulated through blood vessels by the movement of powerful muscular tissue in the heart [18]. Networks of blood vessels have a semi-fractal structure [20] in which vessels repeatedly subdivide into smaller vessels – the largest vessels in the body (such as the aorta) are on the order of centimetres [18], while the terminating capillaries are on the order of microns [18].

Fresh blood typically feeds into organs and tissues via the capillary bed, and this is where the aforementioned exchange processes occur [19]. The passage of blood into an organ or tissue is known as perfusion, and its regulation is essential for life [18, 19].

2.1.2 Homeostasis, blood flow regulation and cerebral blood flow

To maintain homeostasis, the various systems of the body regulate their physiological condition in several coupled feedback loops. In the case of blood flow, these are commonly divided into *extrinsic* and *intrinsic* regulatory factors [18]. Extrinsic regulation occurs by such means as sympathetic nervous response and circulating hormonal signals, and has a widespread effect on the vascular tone throughout an entire system or even the entire body [18, 19]. Conversely, intrinsic factors are specific to a given organ, with examples including myogenic regulation (a feedback coupling between mechanical stress from blood flow and vessel dilation/constriction), regulation due to metabolic byproducts (a negative feedback loop wherein the products of metabolism encourage increased blood flow by vasodilation), and neurogenic regulation (control by the sympathetic nervous system) [18, 19]. These changes in blood flow properties – particularly localised changes within a single organ – motivate perfusion imaging [21, 22]. Such changes in blood flow are often of interest because they are linked to pathology, the action of various pharmaceutical drugs, or even different patterns of attention and thought as in neuroscientific experiments [23, 24, 25, 26, 27, 28].

Blood flow to the brain is of particular importance because of the brain's sensitivity to perfusion fluctuations [18, 22]. The brain has a high metabolic demand (approximately 20% of the oxygen used in metabolism is used by the brain, despite the brain typically weighing 2% of total body weight), and is more sensitive than other organs to comparatively small changes in perfusion [29]. The human cerebral blood flow in grey matter is approximately 50–70 ml/100g/min in young healthy adults as measured by both positron emission tomography (PET) and ASL studies, decreasing with age [30, 31]. Remarkably, blood flow changes can be measured on the level of individual brain areas – for example, when experiencing a visual stimulus, blood flow increases to the visual cortex to cope with increased demand [32]. This phenomenon (and imaging processes used to observe it) has been hugely influential in modern neurology and neuroscience, and is discussed further

in Section 2.1.3.

2.1.3 Blood flow imaging methods

Here we briefly examine the main perfusion imaging methods, with a focus on how they compare to ASL. We divide these into magnetic resonance imaging (MRI) methods and non-MRI methods.

2.1.3.1 MRI

More detail about the physics underpinning MRI is described in Section 2.2. Here, we simply note that MRI scans involve applying a strong magnetic field to the subject, then measuring the signal emitted by hydrogen nuclei in response to changes in this magnetic field. These emitted signals can, subject to certain assumptions, be used to reconstruct a per-voxel map of tissue properties. The most common type of MRI examines images of structure (again, see Section 2.2); however, there are a myriad of tissue properties that may be measured, including blood flow. Blood flow imaging using MRI is more challenging than the more common structural imaging: the SNR is lower, longer acquisitions lead to subject motion being more problematic, and estimation/interpretation of the flow properties can be complicated by many kinds of confound or artifact [23].

ASL ASL is the primary focus of this thesis, and provides a quantitative measurement of perfusion. ASL uses the blood water as a tracer by magnetically labelling it – that is, ASL does not require the injection of an exogenous contrast agent [24].

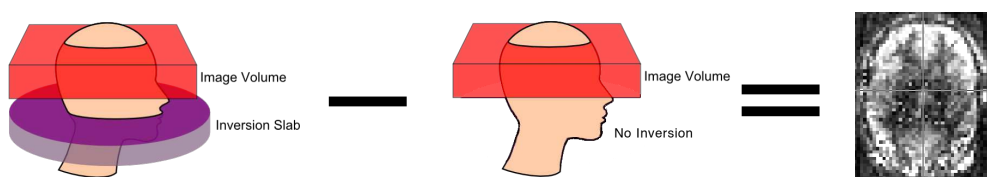


Figure 2.1: A high-level view of ASL measuring perfusion: a perfusion-weighted image (right) is the difference between labelled and control images (left and centre respectively). Figure adapted from [33]

First developed in the late 1980s and early 1990s [22], ASL was slower to achieve popularity than blood oxygenation level dependent imaging (BOLD), its non-quantitative functional MRI contemporary [34]. This was largely due to the

technical challenges involved and the relatively poor quality of ASL images [24, 23]. Over time this situation improved dramatically, and ASL is now available on all major scanners [23].

BOLD Blood-oxygenation level dependent MRI (BOLD) uses the oxygenation-dependent difference in transverse magnetisation (parameterised by observed time constant T_2^* , see Section 2.2) to measure relative changes in blood flow within the brain [34]. Neuronal activation leads to increased oxygen use, which in turn causes an increased supply of blood flow through the haemodynamic coupling as previously discussed. Consequently, the BOLD image is correlated with the cerebral blood flow (CBF). Unlike the other contrasts described here, BOLD does not offer quantitative measurements of perfusion, only relative patterns of activation [23]. However, like ASL, it benefits from having no requirement for an exogenous contrast agent. Its temporal and spatial resolution are typically similar to ASL, although image quality and SNR are typically slightly better [23].

IVIM Intravascular incoherent motion (IVIM) is a contrast based on the popular technique of diffusion imaging [35]. In IVIM, lower-than-normal diffusion gradients are applied, resulting in a signal that is partly a function of water diffusion as usual, and partly a function of microscale perfusion. Like ASL, IVIM benefits from not requiring an exogenous contrast; and like ASL, IVIM offers absolute measurements of blood flow (or pseudo-diffusion). However, unlike ASL, IVIM is mostly used outside the brain: preliminary results show that the original biexponential IVIM model does not describe perfusion well in all vascular geometries, and its usage in the brain is highly experimental, and pseudo-diffusion is not immediately comparable to capillary perfusion as measured using a tracer-based method [36].

When used outside the brain, IVIM has been notably applied in well-perfused abdominal organs such as the liver [37], kidneys [38] and even the placenta [6, 14]. Because of the conceptual similarity to ASL, some of the methods presented in this thesis have subsequently been applied to IVIM experiments in the placenta and the liver [4, 12, 14], although this is largely outside the scope of this thesis, which focuses on ASL.

DSC/DCE Dynamic susceptibility contrast imaging (DSC) uses an exogenous contrast agent (a gadolinium chelate), which changes the T_2^* properties of the MR signal (see Section 2.2) [39, 40]. DSC offers similar information to ASL, albeit at a higher SNR. However, the requirement for an exogenous contrast agent limits its use: in some cases gadolinium chelates are contraindicated (for example when the kidneys are damaged) [41], and in any case it is undesirable to prepare a subject for bolus injection. Notably, there are recent concerns about accumulation of gadolinium-based contrast agents in the brain and other organs, and unnecessary DSC scans are to some extent discouraged [42].

Dynamic contrast enhanced imaging (DCE) is conceptually similar to DSC, except that the images are weighted by longitudinal relaxation (parameterised by time constant T_1) and the perfusion signal comes from gadolinium-driven lowering of T_1 [40]. DCE operates at a longer timescale than DSC or ASL, yielding estimates of permeability properties. As with DSC, DCE benefits from a high SNR relative to ASL, at the cost of using an exogenous contrast agent. Consequently DSC and DCE are often used to generate relatively high quality maps of perfusion properties, but have significant barriers for use in large neuroimaging studies, screening programmes, and even routine clinical use [41, 42].

2.1.3.2 Non-MRI methods

Two of the non-MRI methods described here fall under the category of nuclear imaging, in which radioactive tracers are used to generate the signal [43]: single photon emission computerised tomography (SPECT) and positron emission tomography (PET). Such methods were historically the most reliable and accurate (and first) to provide usable perfusion images, but have significant practical obstacles: intravenous administration of a contrast agent is already a significant barrier to use, as previously discussed; when that contrast agent is a radioactive substance subject to stringent regulation, these barriers become even steeper. The remainder of the methods (CT and ultrasound) are not nuclear imaging although CT, being based on X-rays, does use ionising radiation in the image acquisition phase.

SPECT/PET SPECT for perfusion exclusively uses technetium-based radioisotopes [43]. Most prevalently used for perfusion imaging of the heart, SPECT can also be used in the brain [43]. An advantage of SPECT over PET is that the radioisotopes used are significantly more long-lived: technetium-based radioisotopes have a half-life of six hours, meaning that they can be produced far away from the hospital site and delivered over days [44]. PET for perfusion uses either fluorine-18 deoxyglucose (FDG) or oxygen-15 (O15) as the radiotracer [43]. The former has a half-life of two hours; the latter a half-life of two minutes [44]. In both cases, short half-lives render it necessary to produce radioisotopes close to the scanner site – a significant practical disadvantage over SPECT. Nonetheless, PET's improved spatial resolution, speed, versatility and ease of quantification make it popular enough to overcome this drawback. The fundamental spatial resolution limit of PET in general is on the order of 2mm [45], while in theory significantly smaller resolutions are achievable with MRI – although for ASL as currently measured, this is a comparable resolution.

CT CT for perfusion acts similarly to DSC in that there is an injection of an iodinated contrast agent (or inhalation in the case of a xenon-based contrast), which is tracked over time to provide the signal [21]. CT perfusion imaging is primarily used after ischaemic stroke, to differentiate salvageable tissue from infarct [46]. Diffusion MR imaging is also suitable for this task, but it is not always feasible to get timely access to an MR scanner [46]. Although perfusion CT has the convenience of not requiring a radioactive contrast agent, it nonetheless needs an exogenous contrast agent. It also has the significant disadvantage of requiring a CT scan – an X-ray method with a large (for imaging) dose of ionising radiation, on the order of 5 mSv, reasonably comparable to FDG-PET [46]. As such, while CT is not a form of nuclear imaging, it nonetheless has some of the same disadvantages for widespread applicability, and is mostly limited to urgent use for stroke diagnostics.

Ultrasound There is precedent of using ultrasound to assess blood flow – either using Doppler-based measurements or an injected contrast agent [21, 47]. Doppler-based imaging can only image bulk flow in large vessels, whereas microbubble-

based contrast agents provide more detail at the cost of convenience [47]. In the former case, imaging is not truly comparable to the other approaches discussed here; in the latter case results are preliminary, and the skeleton provides a significant barrier to use throughout the body (for instance in the brain) [47].

2.1.4 Applications of perfusion imaging

Many imaging studies of perfusion focus on the brain, as does most of the original work in this thesis. Consequently we give the brain its own section before discussing applications in other organs of the body.

2.1.4.1 Clinical applications in the brain

Clinically, the area in which perfusion imaging is best established is arguably ischaemia (stroke). There are significant localised blood flow changes in ischaemia by definition, meaning this condition makes a large difference to perfusion images [48, 25]. Several perfusion imaging techniques have shown their capability to diagnose stroke with no other diagnostic information; subsequent research has focused on triage/guidance for thrombolytic therapy, prognosis, and to a lesser extent on guidance for non-thrombolytic therapies. Perfusion imaging has also been used in other vascular disorders such as chronic cerebrovascular disease, moyamoya disease and arteriovenous malformation [26].

In neuro-oncology, perfusion imaging has a history of use for differential diagnosis (tumour or abscess, tumour type), tumour grading (for instance in glioma) and monitoring of therapy [49]. Tumours are characterised by extreme changes in angiogenesis (and are often described as “hot and bloody” [50]), which leads to striking changes in perfusion images.

Perfusion imaging has also shown promise for assessing epilepsy, developmental disorders and dementias such as Alzheimer’s disease or fronto-temporal dementia [25] – as well as cardiovascular diseases in their own right and as comorbidities of other diseases such as diabetes [51]. These applications are less well established, but cerebral perfusion has shown promise as a biomarker for severity of disease, and even has hope of being used prognostically. There is a widespread hope that

dementias might be responsive to currently ineffective treatments if caught earlier, and perfusion imaging might be a way of achieving this [52]. Perfusion imaging of various kinds is also capable of distinguishing dementia subtypes, such as fronto-temporal dementia versus Alzheimer's disease [53].

2.1.4.2 Clinical applications in other organs

One of the most important niches for perfusion imaging is perfusion imaging of myocardial ischaemia, using gamma cameras (the two-dimensional predecessor of SPECT) or CT perfusion [54]. As in the case of stroke, previously discussed, ischaemia by definition produces immediate and prominent changes in perfusion, and hence perfusion imaging is very effective for diagnosis, prognosis and guidance.

Applications in oncology are similar to those described in the case of neuro-oncology: tumours exhibit a different blood flow regime to healthy tissue, meaning that perfusion imaging has high utility for diagnosis and grading [49].

Perfusion imaging has also shown promise in highly vascularised organs of the abdomen, such as the liver, kidney and placenta [37, 55, 56]. Particularly for the placenta, invasive imaging modalities using exogenous contrast agents are discouraged, which encourages the use of noninvasive techniques such as ASL or IVIM.

2.1.4.3 Applications in neuroscience and pharmacological studies

As discussed in Section 2.1, there are localised changes in brain perfusion according to different cognitive demands placed on the brain. For example, when a subject is involved in a visual task or observing visual stimuli, blood flow to the visual cortex increases measurably [32]. This phenomenon has led to widespread use of blood flow imaging for neuroscientific research [57]. Although any suitably sensitive perfusion imaging technique could be used for this purpose, BOLD MRI dominates this application. BOLD does not provide absolute measurements of blood flow, but rather measures changes in T_2^* due to a change in blood oxygenation (which in turn leads to increased local CBF due to the cerebrovascular coupling). BOLD imaging has the advantage of superior time resolution compared to ASL and (like ASL) does not require the injection of an exogenous contrast agent [23].

Perfusion imaging of various kinds has also been used for pharmacological

studies, ranging from neuroscientific research on everyday substances such as caffeine to investigatory research for novel therapeutic drugs [25]. The latter case has mostly been limited to using established clinical applications of perfusion imaging, but given the increasing usage of perfusion imaging in a greater variety of conditions, this area has significant potential for future development.

2.2 ASL: MR physics and acquisitions

2.2.1 MR imaging

MRI exploits the effect of different chemical environments on the magnetic resonance response, producing images of the brain or other organs. We will not discuss the fundamentals of MR imaging in great depth, but will instead present a brief conceptual summary to aid in subsequently understanding ASL MRI and the contributions in this thesis.

Hydrogen nuclei, or protons, are abundant in the body's tissues, due to the ubiquity of water within blood, proteins, fats and other biological materials [58]. When a magnetic field (B_0) is applied to protons, their magnetic moment aligns either parallel or anti-parallel to the direction of this field [58]. Aggregated over protons in a volume, there is a net magnetisation that reflects the difference between these two states. In order to measure the net magnetisation, an orthogonal radiofrequency pulse is used to tip the alignment of these protons, which now have a precession frequency (Larmor frequency) related to both the applied field and the chemical environment. As the magnetisation recovers from this displacement it induces an environment-dependent electromotive force (EMF) in nearby coils, and this is the basis of the MR signal [58]. The recovery is characterised by two time constants: longitudinal relaxation (T_1), determining the time for the magnetisation to recover; and transverse relaxation (T_2), determining the time for excited protons to go out of phase with one another. In practice, the measured transverse relaxation is reduced by inhomogeneities in the magnetic field, and is denoted by T_2^* .

To spatially localise chemical properties – that is, to measure an image rather than a signal aggregated over the entire volume – gradient coils superimpose a

small magnetic gradient on the original field, B_0 , such that $B = B_0 + xG$ where x is the distance along the gradient direction and G is the strength of the gradient magnetic field. Because the precession frequency depends upon B , this has the effect of frequency encoding the signal by spatial position [58]. Consequently, after measurement the signal can be transformed from frequency space (the famous “k-space” formalism [59]) to image space. The choice of gradient frequencies can be expressed as a k-space trajectory (a readout), with the most common choice being a Cartesian readout in which k-space is acquired in lines [58].

There is, of course, much further detail to MR imaging – in particular, there exist a plethora of details considering the acquisition and readout, categorised broadly into gradient-echo and spin-echo acquisitions. The image may be acquired in a single excitation and acquisition, or several acquisitions segmented over k-space. The image can also be acquired slice-by-slice, which is common in ASL [23].

2.2.2 Physics of the ASL signal and the Buxton kinetic model

ASL is based upon the use of magnetically-tagged blood as a tracer. Consequently, much of the resulting signal model (and tracer kinetic analysis) is comparable to that seen in the use of exogenous tracers, such as in PET. A single ASL image can be derived from *two* MR images: a magnetically-tagged image, and a control image. In the tagged image, the blood undergoes adiabatic inversion at labelling location below the brain. In the control image, this inversion pulse is not present, although all other conditions are as in the tagged image, meaning they are otherwise comparable. A fairly large number of these images must be acquired to provide a high-quality image, as the signal difference due to the magnetic inversion is on the order of 1% of the normal MR signal. This is a physiological limit on the signal, as only approximately 1% of the cerebral blood can be replaced on a timescale of seconds [60]. Labelling longer than a timescale of seconds has exponentially diminishing returns in the signal strength, as T_1 decay reduces the usable signal from the labelled blood.

2.2.2.1 Pulsed and continuous ASL

There are, broadly, two common categories of ASL: pulsed ASL (PASL) and (pseudo-)continuous ASL (PCASL). Figure 2.2 illustrates the difference between these two approaches. In the former, the magnetic tagging is performed with a small number of pulses (often a single pulse), inverting a large slab of blood near the imaging region. In the latter, the inversion is performed continuously (or pseudo-continuously, as a shaped train of many individual inversion pulses) for a longer time, at a plane near the imaging region [61]. Blood is tagged as it flows through this plane, in a process known as flow-driven adiabatic inversion. Because of the larger tagged bolus in PCASL, and the larger amount of transferred magnetisation in a continuous label, it has a larger SNR, and is consequently recommended over PASL for standardisation in the recent ASL implementation recommendations produced by the ISMRM perfusion study group [23]. However, there are many circumstances in which a PASL sequence must be used – for example due to restrictions on available scanner hardware, or lower radiofrequency power deposition in a multiple inflow time study. Another subtlety is that PCASL is less sensitive to arterial transit time – it is harder to measure the early arrival of labelled blood at a given voxel with the longer inflow times of PCASL, and hence PCASL has taken some time to become popular for the research-focused multiple inflow time ASL measurements [23, 62].

2.2.2.2 Buxton kinetic model

Neglecting error, the difference between tag and control images results solely from the perfusion of the tagged blood into the brain. A key result came from the analysis of Buxton *et al*, who derived a commonly-used approximate model to relate the signal to the perfusion [60].

Fundamentally, the ASL signal difference is given by the amount of magnetisation that is carried into a voxel by the magnetised blood. The insight of the Buxton model is that it is then possible to express the changing magnetic environment as a convolution of three functions: a delivery function, $c(t)$, which specifies how much magnetised blood arrives at time t ; a residue function, $r(t, t')$, which describes what

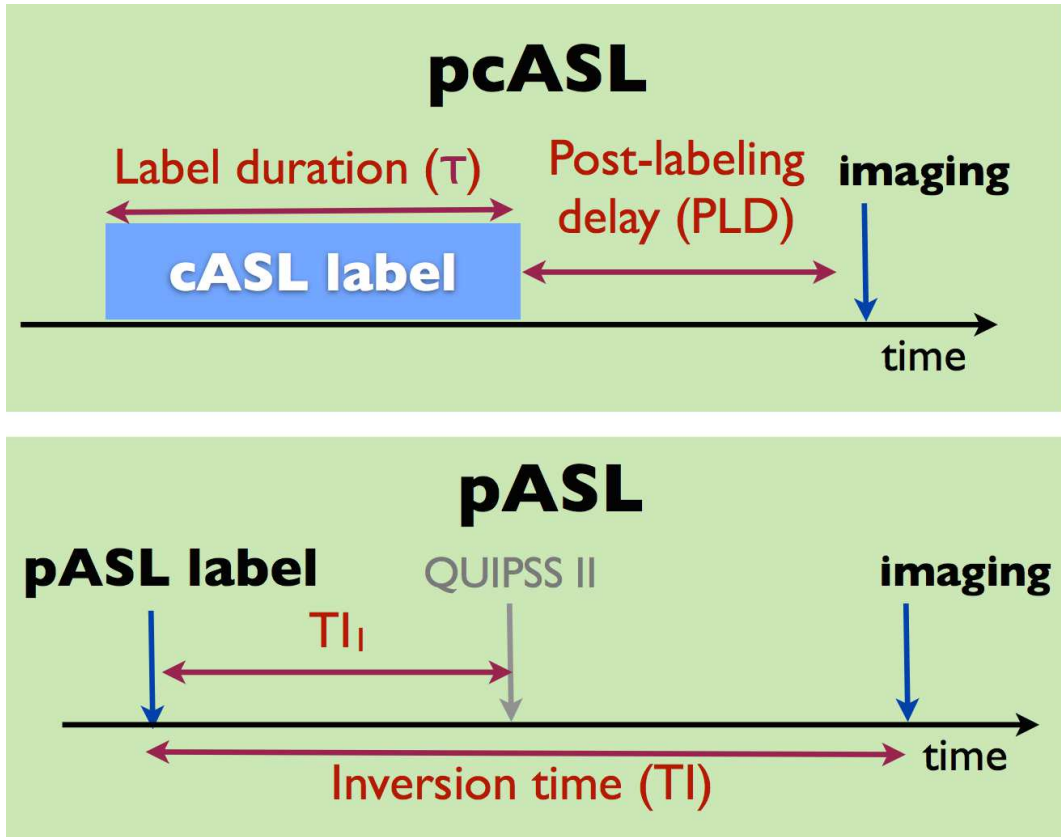


Figure 2.2: Overview of ASL timing, and how it differs between PASL and PCASL. First a label is applied before the imaging region, typically at the neck for neuroimaging. Subsequently, after waiting to allow labelled blood to reach the organ, imaging occurs. In PASL (bottom) the label duration is not necessarily known, although it can be fixed with a quantitative imaging of perfusion with single subtraction (QUIPSS-II) pulse. PASL is usually specified in terms of TI, whereas PCASL is usually specified in terms of PLD. Figure adapted from [63]

fraction of tagged molecules that arrived at t will remain at t' ; and a magnetisation relaxation function, $m(t, t')$, which is the remaining fraction, at t' , of longitudinal magnetisation delivered at t . A common assumption is that the flow is at steady-state, so $r(t, t') = r(t - t')$ and $c(t, t') = c(t - t')$, and they may be respectively written as $r(t)$ and $c(t)$. This yields the following expression for the ASL signal:

$$\Delta M(t) = 2M_0 b f(c(t) * [r(t)m(t)]) \quad (2.1)$$

where $*$ represents convolution, $M_0 b$ is the equilibrium magnetisation of the blood and f is the perfusion (also known as the cerebral blood flow, CBF).

A common assumption is that of “plug flow” – that is, a boxcar-shaped input

function. Another useful assumption is that the exchange kinetics can be described by a single-compartment model – that is, $r(t) = \exp -ft/\lambda$. Finally, it is assumed that labelled blood does not remain in the vessels after reaching the tissue, so magnetisation decays with relaxation time T_{1b} (T_1 for blood) until it reaches the tissue, at which point it abruptly switches to decaying with relaxation time T_1 . This sets $m(t) = \exp -t/T_1$. Applying this modelling to PASL yields a signal model relating perfusion to the measured control-label difference:

$$\Delta M(t) = \begin{cases} 0, & 0 < t < \Delta t \\ 2M_{0b}f(t - \Delta t) \exp(-t/T_{1b})q_p(t) & \Delta t < t < \tau + \Delta t \\ 2M_{0b}f\tau \exp(-t/T_{1b})q_p(t) & \tau + \Delta t < t \end{cases} \quad (2.2)$$

$$q_p(t) = \begin{cases} \frac{\exp(kt)(\exp(-k\Delta t) - \exp(-kt))}{k(t - \Delta t)} & \Delta t < t < \tau + \Delta t \\ \frac{\exp(-k\Delta t) - \exp(-k(\tau + \Delta t))}{k\tau} & \tau + \Delta t < t \end{cases}$$

$$k = \frac{1}{T_{1b}} - \frac{1}{T_1'}, \frac{1}{T_1'} = \frac{1}{T_1} + \frac{f}{\lambda}$$

where $\Delta M(t)$ is the demagnetisation response, t is time, T_1 and T_{1b} are decay constants for magnetisation of water in tissue and blood respectively, f is perfusion magnitude, Δt is the transit delay from the labelling plane to the voxel of interest, τ is the bolus temporal length and λ is the blood-tissue partition coefficient. The assumptions can similarly be applied to create a model for CASL, as done in [60], yielding the following:

$$\Delta M(t) = \begin{cases} 0, & 0 < t < \Delta t \\ 2M_{0b}fT_1' \exp(-\Delta t/T_{1b})q_{ss}(t) & \Delta t < t < \tau + \Delta t \\ 2M_{0b}fT_1' \exp(-\Delta t/T_{1b}) \exp(-(t - \tau - \Delta t)/T_1')q_{ss}(t) & \tau + \Delta t < t \end{cases} \quad (2.3)$$

$$q_{ss}(t) = \begin{cases} 1 - \exp(-(t - \Delta t)/T_1'), \Delta t < t < \tau + \Delta t \\ 1 - \exp(-\tau/T_1'), \tau + \Delta t < t \end{cases}$$

Evidently, if all parameters except for f were known, it would be trivial to invert the model to infer f . This is conceptually the simplest type of ASL measurement: a single inflow time (TI) is used, and then parameters are either measured elsewhere or assumed, allowing f to be calculated by a simple division.

In practice, Δt is not likely to be known *a priori*, and differs from voxel to voxel. For single-TI ASL to work, then, the effects of Δt upon the signal at the measured TI must be minimised. For both PASL and PCASL, this can be fairly workable without further modification [60]: when $t > \Delta t + \tau$, and $T_{1b} = T_1'$, the dependence on Δt is cancelled out. We note that this is easier to see in Equation 2.3. In Equation 2.2, the presence of k in the denominator makes it seem as if $q_p(t)$ would become undefined. However, $q_p(t)$ can be re-expressed as $q_p(t) = \frac{\exp(-k\Delta t)(1-\exp(-k(\tau)))}{k\tau}$, which converges towards unity as $k \rightarrow 0$. Even when T_{1b} and T_1' are not equal but only similar, the dependence on Δt is weak for $t > \Delta t + \tau$, and so a single measurement can be taken at some fairly long delay after the labelling. This has the disadvantage of reducing the signal for almost all voxels, which is especially undesirable in a low-SNR modality such as ASL. However, on balance, this is the recommended implementation of ASL [23].

A notable disadvantage of PASL, compared to PCASL, is that the labelled bolus does not have a clearly-defined width. A common modification – necessary to provide a quantitative CBF measurement with a single inversion time similar to that which PCASL can provide – is to use a second pulse to remove the distal end of the bolus, hence restricting the bolus width to a known time. The main way in which this is done is known as QUIPSS [64, 65] (quantitative imaging of perfusion using a single subtraction), although an improved scheme known as Q2TIPS [66] (QUIPSS-II with thin-slice TI1 periodic saturation) is also common.

2.2.2.3 Multiple inflow time ASL

A more advanced implementation of ASL, which is more common in research, uses multiple inflow times in order to estimate the arterial transit time, Δt (also referred to as ATT). This may be desirable for several reasons. Δt is, itself, a parameter of interest, and is often increased in several diseases [67]. Moreover, by allowing for

measurement of the signal at higher-intensity times (that is, near the peak of the ASL response curve), and accounting for Δt in the model inversion, multi-TI ASL can provide even better estimates of the perfusion [68] [69]. There are other methods that attempt to measure Δt , for example using bipolar gradients to differentially encode blood flow [70], but these remain experimental and are not commonly used.

The principal disadvantage of using multiple inversion times in ASL is that it can increase the time taken to acquire sufficient data. There have been many attempts to reduce the multi-TI scan duration [62, 71], and a large part of the original work in this thesis (Chapter 3) concerns using optimal experiment design to this end.

Multi-TI ASL also has the disadvantage that it complicates the model fit – rather than simple division, as in the single TI case, now two parameters must be simultaneously inferred. Indeed, sometimes an even greater number of parameters are fitted – as in PASL without QUIPSS-II/similar, which requires τ to be estimated [72]; or simultaneous estimation of parameters such as T_1 or the arterial blood volume, ABV [73, 74]; or simultaneous estimation of partial volume effects [75]. The problem of simultaneous inference for multiple parameters is tractable using several different approaches. Solutions can be provided most simply through non-linear least squares fitting [15]; and more advanced Bayesian inversion/inference methods [74, 76] improve on this and render this problem solved, for practical ASL analyses. More daunting is the requirement for a larger scan time to estimate these parameters simultaneously – even when only fitting two parameters (f and Δt), the required scan duration is considerably longer than typical single inversion time CBF quantification [23].

2.2.2.4 Readout

The readout (see Section 2.2) is the k-space trajectory used for acquisition of the image. In ASL, as in MR imaging generally, there is an important distinction between 2D readouts such as echo-planar imaging (EPI), in which slices are acquired separately, and 3D readouts such as gradient spin-echo (GRASE), in which the z dimension is also included within the k-space trajectory. 3D readouts are generally preferred for ASL: the SNR is intrinsically higher due to exciting the whole volume

during each readout, and the entire volume is excited at a given time, rather than at separate per-slice times [77]. This is useful both for quantification and background suppression, discussed in Sections 2.2.2.3 and 2.3.5 respectively.

3D readouts generally lead to fewer ASL volumes, but each at a higher SNR; 2D readouts give more volumes at a lower SNR. EPI is more prone to geometric distortion, while GRASE suffers from worse in-plane blurring [78, 23]. Despite the advantages of 3D readouts, there is no practical consensus on readout – the recent ASL white paper recommends 3D-GRASE, but acknowledges EPI as acceptable, and different scanner vendors offer different readouts as their recommended implementation of ASL [23].

2.2.2.5 Novel acquisitions

There are many different ASL-based acquisitions, in addition to the basic single-TI and multi-TI acquisitions set out above. Quantitative STAR labelling of Arterial Regions (QUASAR) is a model-free ASL technique, inspired by DSC MRI, in which crusher gradients allow the estimation of an arterial input function and deconvolution allows estimates of perfusion and arterial blood volume [79]. Vessel-selective ASL is a technique in which labelling is modulated across the different feeding arteries, allowing inference of vascular territories [80]. Hadamard ASL (or time-encoded ASL) is a time-saving technique in which labelling is time-encoded, typically using a Hadamard sequence for maximum distinguishability [62, 81]. ASL fingerprinting is a recent attempt to use similar compressed sensing methods to magnetic resonance fingerprinting for ASL to reduce the time taken while measuring several parameters simultaneously [82].

There have also been several uses of ASL simultaneously with other modalities, such as calibrated BOLD fMRI [83], joint T_2 and ASL imaging to measure blood-tissue water transfer [84, 85], and joint diffusion and ASL imaging to estimate capillary permeability [86].

2.3 ASL: reconstruction, fitting and post-processing

2.3.1 ASL pipeline and fitting

The reconstruction of the MR signal transforms frequency-encoded signals into a three-dimensional image, where voxel values are a function of the per-voxel chemical environment. However, another level of processing exists for ASL: given a series of control and label images, we have yet to estimate the per-voxel perfusion properties. The measured signal is a function of the perfusion as set out in Section 2.2, which means to estimate perfusion we must invert that model. However, there are many different steps to consider in an ASL processing pipeline: in addition to the fundamental model fitting, common steps include motion correction, partial volume correction, artifact/outlier filtering and spatial regularisation/denoising.

2.3.2 Error and bias

There are several sources of corruption in ASL experiments. Fundamentally, measurement accuracy is limited by thermal noise: random thermodynamic fluctuations in the MRI scanner's measurement coils place a limit on how close a measurement will be to an idealised ground truth [58]. In MRI in general, noise is well-modelled by a Rician distribution [58]. In ASL, because the signal is the difference of two MR magnitude images, the noise is modelled by the difference of two Rician distributions, which is well-modelled by a Gaussian distribution [87]. (An MR magnitude image has Rician noise because both real and imaginary parts have separate Gaussian noise.) In addition to this white noise, there is the difficulty of motion: subjects are prone to move within the scanner, which ranges from small breathing-related movements to larger movements in the case of patient discomfort. Even if subjects were to lie perfectly still, there would be motion within the body due to the pulsation of moving blood from the pumping of the heart – and such pulsation can make a large difference to measured blood flow [88].

A further difficulty is presented by partial volume effects [89, 90, 75, 91]: a given voxel contains a mixture of tissues, such as grey matter, white matter, cerebrospinal fluid, blood vessels, fat, etc. Grey and white matter have different perfu-

sion properties, blood vessels may contain labelled blood that has not yet perfused, while other tissues tend to provide no signal. When fitting does not model the individual contributions of different tissues within a voxel, parameter estimates are incorrect – typically perfusion is underestimated. Finally there are other sources in which the signal is corrupted by artifact: for example signal dropout from metal such as in dental fillings or braces, or other causes [23].

2.3.3 Partial volume

Partial volume effects are a common issue in medical imaging. Voxels are likely to contain several different types of tissue, particularly when the voxel size is coarse as in ASL. Consequently, the signal is a mixture of signals from each tissue type – grey matter perfusion is approximately four times larger than white matter perfusion, with other tissues (for example cerebrospinal fluid or dura) having no perfusion at all. Thus, when attempting to quantify grey matter perfusion, uncorrected partial volume effects lead to underestimates.

There are essentially two ways to correct for partial volume: incorporate volume effects into the ASL signal model, or increase the spatial resolution of ASL images to reduce PV effects. The latter approach remains less popular, as the low SNR of ASL means that it benefits from a larger voxel size; however, attempts at better image fitting can to some extent improve this trade-off [92]. The former approach has been implemented in several different ways: Asllani *et al* used regression to implicitly spatially regularise the signal (and hence make the inverse problem well-conditioned); Chappell *et al* used a data-driven spatial prior to explicitly encode spatial regularisation for the same purpose. Approaches such as those of Asllani or Chappell have been applied in several different configurations of partial volume fraction and ASL data, variously using resampled anatomical images [91, 75]; T_1 images specifically acquired in ASL space [93]; or performing the analysis in high-resolution T_1 space by resampling the ASL data [94].

2.3.4 Spatial regularisation

Spatial regularisation provides a powerful tool for ASL processing, not only for denoising, but also for assisting in partial volume correction, which is an ill-posed problem when there is no assumption of parameter spatial correlation. Most spatial regularisation techniques are intended solely as a way to denoise the ASL signal, but the difference between these applications is often ambiguous.

There are many different approaches to spatial regularisation, with most focused on denoising. Wells *et al* prepared a fairly comprehensive comparison examining five such methods in 2010 [95]. To provide a non-exhaustive list, several methods use a more traditional approach to filtering: for example a Wiener filter, wavelet filter, anisotropic diffusion filtering, and independent components analysis. Subsequent work has introduced methods based on convolutional neural networks [96, 97], nonlocal means filtering [98], and other image processing techniques.

In practice, the most widely-used regularisation technique (other than simple Gaussian smoothing) is the approach pioneered by Groves and Chappell using a spatial prior [99]. This is also one of the few methods used for other purposes in addition to denoising: the spatial prior technique is also used for conditioning the inverse problem of partial volume estimation [75]. Another more recent technique uses total generalised variation (TGV), a generalisation of (familiar from natural image processing) total variation [92]. Like the spatial prior approach, TGV is used both for denoising and for another purpose: by providing a high-quality denoising, Spann *et al* argue that they can acquire ASL at a finer spatial resolution, and hence can minimise partial volume effects.

2.3.5 Motion correction

Motion correction commonly uses registration between images to minimise the effect of inter-scan motion. There are two types of scheme for this: registration of all images to one reference image, and registration of images with other images adjacent in time. The former is conceptually simpler, while the latter is more similar to the functional registration performed in the popular neuroimaging software

library SPM [100]. Registration, in general, is an ill-posed problem; it is difficult to ascertain how well motion correction performed without a ground truth. Motion during an acquisition is especially difficult to solve, as it introduces greater ambiguity into the raw acquired data, meaning motion would have to be estimated during reconstruction.

When images are registered with adjacent images, this may be done in a control/label/control/... pattern [63], or a control to control and label to label pattern [100]. The former is predicated on the assumption that it is better to attempt registration between similarly-oriented images, meaning it is preferable to register images to the closest images in time. The latter attempts to account for the systematic (but small) difference between tagged and control images, in case this affects registration. In practice, all of these modifications have a small effect [100], and there is no clearly superior option. A simple but effective method of motion correction, especially for scans corrupted by intra-acquisition motion, is to discard corrupted data. This can be done either automatically or manually, and is arguably the most practical means of correcting for motion.

The previous methods of motion correction apply only to inter-scan motion – when the patient moves between individual acquisitions. Unfortunately, the problem becomes significantly more difficult when the patient moves during acquisition (intra-scan motion). Recent advances in intra-motion correction for ASL have been made using specialist cameras or low-resolution navigator MR sequences to provide a ground truth for *prospective* motion correction [101, 102]. Here, the pulse sequence is modified to change the scan’s coordinate system to compensate for the motion. This appears to yield small but significant improvements, although it remains developmental technology, and needs further research.

A complementary approach to reduce the effects of inter- and intra-scan motion artifacts also exists at the acquisition level: background suppression (BS). This can be used in conjunction with the aforementioned methods or in their place. In background suppression, the background non-perfusion signal is reduced through careful timing of inversion pulses. When motion occurs, its effect on the esti-

mated ASL signal is smaller, consequently, as the suppressed background signal is smaller relative to the ASL signal. The drawbacks of background suppression are twofold. First, removing structural information makes registration less successful, so although the effect of motion is smaller, it is also harder to correct. Second, the choice of BS pulses now affects the scan result – which T_1 values should the BS be focused on, and for which parts of the image? In the 2D case, this can lead to BS applying more in some slices than others; in the 3D case, this is less clearly defined. In both cases, some tissues are more strongly suppressed than others.

2.3.6 Artifact filtering

ASL is prone to artifacts which corrupt the image. Distinct from thermal noise, such artifacts are not well-modelled by a Gaussian distribution. These can come from uncorrected motion (even after motion correction, artifacts may remain), coil instability, susceptibility artifacts due to implants, chemical shift artifacts and ringing artifacts near air-tissue interfaces, among other causes.

There are several ways to handle artifacts. The simplest way is manual inspection: one can attempt to remove volumes that appear suspect while leaving unaffected volumes. This has problems for reproducibility and can require significant effort in large datasets. Consequently, many authors have developed automated methods. Among the simplest and most widely used method is z-score filtering, developed for ASL by Tan *et al*, in which volumes are classed as outliers (or not) according to whether the volume's signal is far from the mean signal[103]. Other approaches use robust estimators in place of the temporal mean [104] to reduce the influence of outliers. Finally, there are more algorithmic methods [105, 106, 107] – in these, volumes are iteratively ranked for their similarity to other volumes, allowing a prioritisation of volumes by how likely they are to be artifact.

2.3.7 Inference and machine learning in ASL

Many of the methods used in spatial regularisation and artifact filtering enter the territory of machine learning, although such methods can differ drastically in their aims and approaches. The inference methods pioneered in ASL by Chappell *et*

al [74, 99, 75] use numerical methods to provide fast, high quality solutions to the fundamental problem of Bayesian inference for the ASL model – that is, inferring the posterior distribution of the perfusion parameters given the observed data. Spatial priors can enforce regularisation and allow for the simultaneous inference of partial volume parameters, arterial blood volume, and other additional parameters.

Very different to this well-established (and data-parsimonious) way of approaching the problem, modern machine learning methods may be used for similar aims [108]. Rather than explicitly creating models of the spatial correlation or signal regularity, and then performing inference on the data subject to these, data-driven approaches can be used to automatically learn these relationships and apply them [108]. This can similarly be used for spatial regularisation, outlier rejection and parameter estimation. Another valuable use of data-driven machine learning is to use the information in ASL images (and other modalities) to classify or score subjects based on disease status [109].

It must be noted, however, that the use of these methods comes at a significant cost: data-driven machine learning approaches not only require large amounts of high quality data, but also suffer from being less interpretable than classical modelling procedures, and moreover they often need to be retrained whenever the data are subject to small changes in acquisition [108, 110]. As such, these automated machine learning approaches are on the cutting edge of information processing in ASL and medical imaging generally. We discuss some of these issues in more depth in Chapter 5, in which we present an original approach to simultaneous denoising and artifact filtering via a convolutional neural network.

2.3.8 Beyond the Buxton model

Several researchers have explored the use of alternative models to the simplified Buxton model. These tend to explore the use of additional phases [111], more realistic input functions, or additional compartments [112]. Although the use of these is supported by the evidence in some cases, their benefit is marginal. Usually, the data are dominated by error and external biases to such an extent that the bias introduced by model inaccuracies is negligible.

2.4 Optimal experiment design

This section examines the mathematical principles that underpin optimal experiment design, and discusses the background of optimal design approaches in medical imaging. This gives a context to the original work in Chapter 3, where we present Bayesian experimental design approaches in multi-TI ASL.

2.4.1 Locally optimal design

An experiment may be described, for the purposes of this work, as a process that gathers data for parameter inference. Such an experiment consists of a forward model, g ; a noise model, e ; parameters to be estimated, θ ; independent variables (also known as design variables), η ; and output data, y .

Assuming additive noise (a common and reasonable assumption for ASL is additive Gaussian noise – see Section 2.3.2), this may be expressed thus:

$$y = g(\theta, \eta) + e \quad (2.4)$$

If the noise model is known, one may formulate an expression for the posterior distribution of the parameters, $p(\theta|y, \eta)$, using Bayes' rule:

$$p(\theta|y, \eta) = \frac{p(y|\theta, \eta)p(\theta|\eta)}{p(y|\eta)} \quad (2.5)$$

Note that the design variables, η , have been explicitly included above, because they will become the free variables in the optimal design problem. Experimental design consists in choosing η to maximise some metric of experimental usefulness – historically, in the literature for problems with approximately Gaussian errors, these metrics have been based upon functions of the Fisher information matrix, H , elements of which are defined by $H_{i,j} = \frac{\partial^2 y}{\partial \theta_i \partial \theta_j}$ [113]. The most commonly used such function for parameter estimation has been the determinant (D-optimality), although several other functions have been used, such as trace (G-optimality) or the minimum eigenvalue (E-optimality). Theoretical justification for these metrics was initially non-rigorous, but later work showed that they naturally arise from information-theoretic arguments.

When one wishes to maximise information gain between a prior belief and the posterior (post-experiment) belief, the difference between prior and posterior distributions must be maximised. Different measurements of the difference between distributions give rise to different utility functions. A common choice of utility function is the Kullback-Leibler divergence [113] between the prior and the posterior distributions, which gives rise to the D-optimality criterion (determinant of the Fisher information matrix). This divergence quantifies the information difference between two distributions, F and G , and is given thus where Θ is the set of possible θ and $F||G$ means divergence of F relative to G :

$$D_{KL}(F||G) = \int_{\Theta} p_F(\theta) \log \frac{p_F(\theta)}{p_G(\theta)} d\theta \quad (2.6)$$

In locally optimal design, which was the form of experimental design most common before computational approaches, the optimal design was selected by setting θ to a representative point value of parameters, and then optimising some experimental design utility function as previously described. The ensuing design is locally optimal in the sense that, for this specific value of θ , it provides the maximum possible information gain [113, 114].

2.4.2 Robust/Bayesian optimal design

A significant problem with locally optimal design, however, is that the resulting experimental designs are often strongly dependent upon θ [113]. Indeed, θ and y are unknown before the experiment by definition – the purpose of the experiment is to estimate θ , and an outcome cannot be known until the experiment has been performed. To find a design that is optimal given a prior distribution of the parameters, rather than using a utility function evaluated at some point estimate of θ , one should use $U(\eta)$: the expectation of the utility, which is evaluated over the support of the parameters, Θ , and the support of the outcomes, Y . Evaluating this expectation allows one to compare the merit of different designs, η , despite θ and y being unknown.

Hence,

$$u(\eta, y, \theta) = D_{KL}(p(\theta|y, \eta) || p(\theta)) = \int_{\Theta} p(\theta|y, \eta) \log \frac{p(\theta|y, \eta)}{p(\theta)} d\theta \quad (2.7)$$

Integration over the entire parameter space, Θ , removes the dependence upon the parameters, θ . Hence,

$$\begin{aligned} U(\eta) &= \int_Y \int_{\Theta} u(\eta, y, \theta) p(\theta, y | \eta) d\theta dy \\ &= \int_Y \int_{\Theta} p(\theta|y, \eta) \log \frac{p(\theta|y, \eta)}{p(\theta)} d\theta p(y|\eta) dy \\ &= \langle D_{KL}(p(\theta|y, \eta) || p(\theta)) \rangle_{y|\eta} \end{aligned} \quad (2.8)$$

where $\langle A \rangle_B$ represents the expectation of A with respect to B .

This definition of the utility makes intuitive sense: the expected utility is equal to the expected gain of information in θ . That is, a larger divergence from posterior to prior implies that the posterior is more different to the prior – and hence offers a greater increase in information.

The optimal experimental design, then, is the η that maximises $U(\eta)$:

$$\eta^* = \arg \max_{\eta} U(\eta) \quad (2.9)$$

Use of this utility function is typically called “robust” or “Bayesian” D-optimality to distinguish it from the locally optimal case. This is a generalisation of the special case where the distribution is Gaussian [113]. Robust D-optimality – and other “alphabet optimality” designs have a long and rich history in experimental design [114, 113], although the problem as posed here is often analytically intractable and numerically challenging. This motivates the use of several different approximate techniques, as discussed in subsequent sections.

2.4.2.1 Common Approximations and Approximate Solution Methods

In order to derive solutions to the optimal design problem, there are many approximate methods that may be used, depending upon the nature of the problem. The less computationally-demanding method is to approximate the distribution involved by some surrogate function that is easy to evaluate. Depending upon the complexity of the surrogate function, this may be simple, or may require a great deal of effort. The alternative method is to evaluate the distribution numerically, for example using Monte Carlo simulations. This has the advantage of being asymptotically exact, for all practical cases [115], but has the disadvantage of bearing a high computational cost.

2.4.2.2 Numerical Simulation

Numerical simulation is straightforward in principle, although many techniques have been developed for reducing its computational cost. Essentially, one uses a two-stage sampling procedure to approximate $U(\eta)$ and $p(y|\eta)$.

Drawing N samples indexed by i , y^i , from $p(y|\theta = \theta^i, \eta)$,

$$U(\eta) \approx \frac{1}{N} \sum_i^N \log p(y^i|\theta^i) - \log p(y^i|\eta) \quad (2.10)$$

As $p(y|\eta)$ does not, in general, have an analytic form, it is similarly approximated using importance sampling, here using index j for those samples and drawing M of them:

$$p(y^i|\eta) \approx \frac{1}{M} \sum_j^M p(y^i|\theta^{i,j}, \eta) \quad (2.11)$$

One common method to speed up this approximate (but asymptotically correct) method is to reduce the number of new θ samples used. Because a large part of the computational cost is sampling of the likelihood for each new $\theta^{i,j}$, it is beneficial to reuse samples θ^i , such that $N = M$. The complexity is reduced (for a fixed value of η) from $O(NM)$ to $O(N)$. This sample reuse does increase the estimator's bias, but

the effect is negligible for all practical implementations [115].

If this numerical method were used to find the optimal design, it would likely be necessary to develop a faster method to search and hence optimise over the space of the design variables, as discussed in [115]. However, in this work, the sampling method is used only a small number of times, to validate the results of the approximate solution method discussed below. Consequently, a grid search is acceptable, as it need only be performed a small number of times.

2.4.3 Approximate Bayesian optimal designs for Gaussian approximations

Many different approximations exist, again depending upon the nature of the problem, ranging in difficulty from simple Gaussian approximations [113] to complex, tailored solutions (namely, Gaussian process emulators [116] and polynomial chaos surrogates [115]). In this work, when taking the surrogate approach, only the Gaussian linearisation is considered, because the distributions involved in the ASL design problem (see Section 2.2.2) are comparatively well-modelled by Gaussians for this task.

When the problem is approximated by a Gaussian, the situation is in some senses analogous to approximating a function in analysis by its linearisation. Indeed, linearisation is used to achieve the Gaussian approximation. One considers the Hessian (the inverse of the Fisher information matrix) – the matrix of second derivatives of the data residuals during parameter fitting. Then, through approximations, one reaches the final cost function, $\det(H)$. Denoting the data and parameter residuals (i.e. the log-probability for Gaussian noise model and Gaussian distributed parameters) as Λ , noise for measurement i as σ , parameter prior standard deviation as ϕ and mean as θ_0 for parameter l :

$$\Lambda^2 = \sum_i^N \left(\frac{y_i - g(\eta; \theta)}{\sigma_i} \right)^2 + \sum_l^{|\theta|} \left(\frac{\theta_l - \theta_{l,0}}{\phi_l} \right)^2 \quad (2.12)$$

$$\frac{\partial \Lambda^2}{\partial \theta_j} = -2 \sum_i \frac{[y_i - g(\eta; \theta)]}{\sigma_i^2} \frac{\partial g(\eta; \theta)}{\partial \theta_j} + \frac{1}{\phi_j^2} (\theta_j - \theta_{j,0}) \quad (2.13)$$

$$\frac{\partial^2 \Lambda^2}{\partial \theta_j \partial \theta_k} = 2 \sum_i^N \frac{1}{\sigma_i^2} \left(\frac{g(\eta; \theta)}{\theta_j} \frac{g(\eta; \theta)}{\theta_k} - (y_i - g(\eta; \theta)) \frac{\partial^2 g(\eta; \theta)}{\partial \theta_j \partial \theta_k} \right) \quad (2.14)$$

Ignoring the second order terms (which are close to zero near the minimum) allows this to be reformulated to remove the data, y :

$$H_{jk} = \sum_i \frac{1}{\sigma_i^2} \left(\frac{\partial g(\eta; \theta)}{\partial \theta_j} \frac{\partial g(\eta; \theta)}{\partial \theta_k} \right) \quad (2.15)$$

This is familiar as the Fisher information matrix, which is the inverse of the parameters' covariance matrix. Minimising the determinant of this matrix corresponds to minimising the volume of the confidence interval ellipsoid for the parameter estimates. Numerical solution of this problem depends upon the distributions involved, and is discussed further in subsequent sections.

2.4.3.1 Exact Results for Gaussian-Distributed Data

One notable result relates the D-optimal design, under the Gaussian approximation, to an important special case that has an analytic solution. This is the situation when the distributions of interest are themselves Gaussian and linear, and Bayesian D-optimality maximises the determinant of information matrix (plus the prior covariance).

This gives insight into the Gaussian approximation discussed above. So long as the approximation is good, the proposed design, based upon minimising the determinant of the approximate information matrix, will be close to the truly optimal design. In the simple case that the covariance matrix is diagonal, this will clearly minimise the product of covariances. More generally, this optimisation problem will reflect a trade-off between measurement precision and the associated covariances.

2.4.4 Optimal experiment design for medical imaging

Making use of the theoretical background of optimal experimental design as discussed above, several researchers have applied experimental design to medical imaging. In medical imaging, there is typically a high cost associated with scanner/equipment time, meaning that there are large potential benefits to improved experimental designs. This section reviews previous uses of experimental design in other MR imaging methods before discussing pre-existing uses of experimental design in arterial spin labelling.

2.4.5 Optimal experiment design in other MR contrasts

Optimal experimental design has long been used in biology [117, 118] and nuclear imaging [119], but its use in MR imaging was comparatively recent. Experimental design approaches to magnetisation transfer imaging were used by Cercignani [120] and others [121], significantly improving the quality of magnetisation transfer estimates. A similar approach was also successfully used for Dixon imaging to estimate water and fat [122]. Diffusion measurements have been similarly optimised, with different design optimisations performed for many of the different models by which diffusion MRI is quantified [123, 124]. Again, these optimised designs produce superior parameter estimates, at least within a certain region of validity – when the true underlying parameters are drastically different from the targeted design values, the optimised designs can perform worse.

2.4.6 Optimal experiment design for ASL acquisition parameters

The most fruitful ASL optimal design approach taken in previous work made use of the Gaussian approximation method discussed in Section 2.4.3. This was the work of Xie *et al* [71], who created an iterative exchange algorithm to compute the optimal inflow times for a multi-TI ASL experiment, yielding improvements in Δt estimation at a small cost to f estimation, or improvements in f estimation with no cost to Δt . In follow-up work, Xie *et al* explored the use of adaptive experimental design [125]: iterating the design optimisation as data was acquired, rather than op-

timising the design based only upon the initial prior. This was especially beneficial for subjects with abnormal parameter values due to pathology (in this work, moyamoya disease). A similar approach to ASL optimal design was taken by Santos *et al* [126], who applied a very similar technique to Xie *et al*, yielding comparable improvements, and showed that these improvements were sustained after simplification of the design into a smaller number of clustered measurements. This work also used additional simulations to evaluate the proposed experimental designs at different values for the perfusion and transit time. This approach to design optimisation has the additional benefit of easily supporting optimisation for a subset of parameters: Xie *et al* show that designs can be generated for perfusion and transit time, or can be generated to prioritise perfusion.

Kramme *et al* took a different approach to acquisition optimisation, some years after Xie *et al*: rather than making any claim of optimality, they presented three different designs intended to compensate the T_1 decay over time – that is, to provide more measurements at longer TIs. They proceeded to compare the performance of these designs in simulations and experiments, both for normal transit times and prolonged transit times. They concluded that a quadratic distribution of inversion times outperforms equally-spaced and linearly-spaced inversion times, particularly for prolonged transit times. The Kramme approach has the benefit of simplicity compared to the optimal design approaches of Xie and others, although in theory a parameter-driven optimisation should yield a better result.

Relatedly, while not being strictly optimal design, optimisation approaches have also been used to select background suppression pulses [127]. By using a utility function based on simulation of the magnetisation signal with different values of T_1 , Maleki *et al* found optimal timings for background suppression, reducing background signal to 1% across a broad range of T_1 values. This utility function is more straightforward than the Fisher information as used in D-optimal design, but the overall concept and application are very similar: optimisation of a signal metric marginalised approximately over some parameter prior (in this case T_1).

2.5 Optimal design for ASL: overview and insight

This section provides detailed background information about optimal design in the context of ASL. In particular, this section relates approaches within the same mathematical treatment, and shows the benefits and drawbacks of previous such methods in practical use. This section also sets out the different trade-offs and model/optimisation choices in the literature on experimental design for ASL. These differences lie in four different areas: treatment of priors, measurement constraints, optimisation tractability, and pulsed versus continuous ASL.

In Section 2.5.3, on measurement constraints, we examine how optimal design procedures limit themselves by scan time – clarifying a key limitation of the original work by Xie *et al*, and how it differs from subsequent optimal design work by ourselves and others. In Section 2.5.4, on optimisation tractability, we delve into the practical details of optimisation within these design problems – describing how these problems can be made tractable through approximate methods, simplification based on symmetry arguments, or heuristic global optimisation techniques. Finally, Section 2.5.5 compares PASL and PCASL, and we set out several relevant but potentially non-obvious differences that labelling type makes on optimal design – in particular: ranges of values, differences in models, and key differences in how specific absorption varies with label duration.

2.5.1 Translation of optimal design to ASL

In multi-TI ASL, the measured signal is related to the underlying parameters that characterise blood flow: the perfusion, f , and the transit time, Δt . This is discussed at greater length in Section 2.4. Practically, this means that the forward model for optimal design, $g(\theta)$, is equal to $\Delta M(f, \Delta t)$. Different values of f and Δt produce different kinetic curves, as shown in Figure 2.3. Noisy measurements are taken at some measurement times, t , after having a label applied for label duration, τ . The purpose of optimal design, then, is to select the optimal values of t – or, even more challenging, the optimal values of t and τ .

Intuitively, one wishes to measure near the high-SNR peak at $t = \Delta t + \tau$, to estimate f . However, one also wishes to measure near $t = \Delta t$, to estimate Δt . Thus

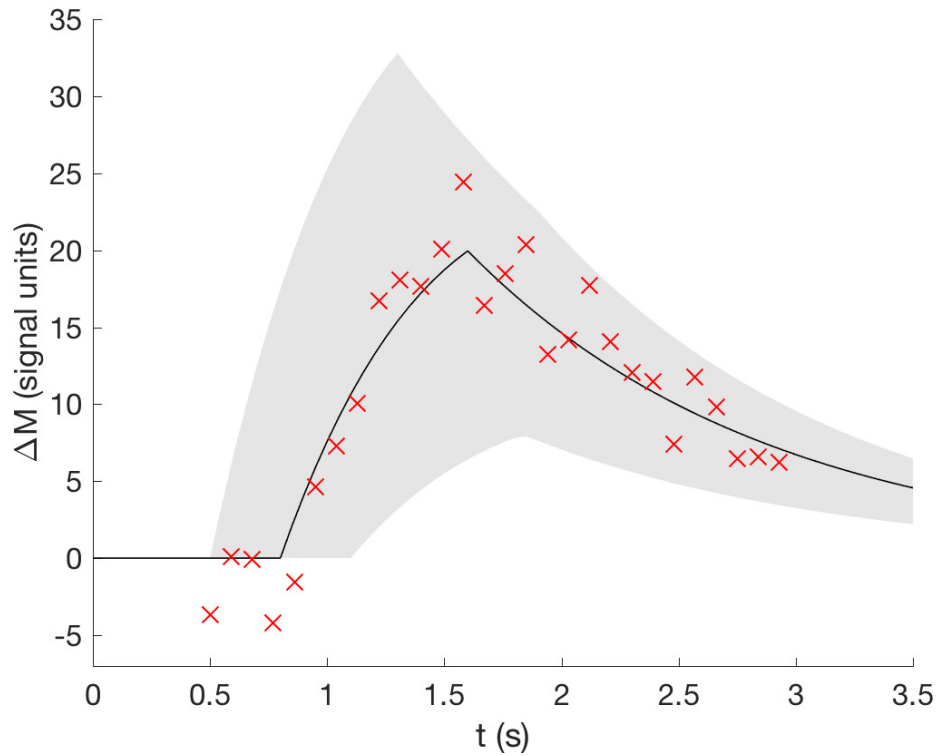


Figure 2.3: The black curve shows the ASL kinetic model (PASL), with f and Δt set to typical values of 40 ml/100g/min and 0.8 s respectively. The shaded region shows how the curve varies as f and Δt vary over common values (20–60 ml/100g/min for f and 0.5–1.0 s for Δt). Blood arrives at $t = \Delta t$, the first turning point of the curve; then collects, increasing the measured signal, until the second turning point, $t = \Delta t + \tau$, when there is no more inflowing blood and the signal begins to decay. Red crosses show example noisy measurements from this forward model: the task in model fitting is to infer f and Δt from these noisy measurements.

the optimal design will involve some trade-off between these parameters – albeit made under uncertainty as to their values, represented by a prior.

2.5.2 Priors

Optimal design problems make use of priors over the parameters, as set out in Section 2.4. These are necessary to marginalise the local utility, $u(\eta; \theta)$, over the entire space of parameters, Θ . That is, $U(\eta) = \int_{\Theta} u(\eta; \theta) p(\theta) d\theta$. In practical optimal design problems, this prior will be at least weakly informative intuitively, to improve one’s measurements it is necessary to have an idea of what needs to be measured.

2.5.2.1 Prior distributions

The earlier ASL optimal design methods [71, 125, 126] used Gaussian priors, consistent with common optimal design practice and with the inference tools in FSL [76]. These were truncated to be non-negative, because perfusion and transit time are non-negative. Many physiological properties, including perfusion, are normally distributed over a population, to a first approximation, as a consequence of the Central Limit Theorem [71]. As such they were a natural starting point, although even when the parametric form of the distribution has been fixed, there remain difficulties in choosing its hyperparameters. Should the Gaussian be chosen to be very broad, for instance covering all physiologically plausible values relatively equally? (This might be, for example, $f \sim \mathcal{N}(\mu = 70, \sigma = 30)$ ml/100g/min and $\Delta t \sim \mathcal{N}(\mu = 1.5, \sigma = 1.0)$ s, where \mathcal{N} denotes a normal distribution.) Or should it be narrower, perhaps reflecting prior knowledge of a particular population, or even of a particular individual? In their initial work, Xie *et al* set these hyperparameters for a young and healthy population [71] ($f \sim \mathcal{N}(\mu = 70, \sigma = 24)$ ml/100g/min, $\Delta t \sim \mathcal{N}(\mu = 0.7, \sigma = 0.3)$ s); subsequently they explored per-individual priors using a system of real-time feedback and optimisation in a closed loop [125].

Priors may be found from previous knowledge of physiologically plausible values, or may be made more informative if data already exists for a suitable population. One innovative way around the chicken-and-egg problem of wanting better priors to measure, but needing measurements to create priors, is to use a rapid low resolution prescan to generate an approximate estimate of transit time. Another solution to this problem is the use of a feedback loop in which priors are initially broad, but parameter estimates are continuously updated as data is acquired – with priors updated based on these estimates, and used to select subsequent measurements.

2.5.2.2 Global optima versus ensemble of local optima

As previously mentioned, the most fruitful ASL optimal design approach taken in previous work made use of the Gaussian approximation discussed in Section 2.4.3, which leads to a utility function based on the Fisher information matrix. This was

the work of Xie *et al* [71], who used a coordinate exchange algorithm to compute the optimal inversion times for a multi-TI ASL experiment. Deferring the discussion of coordinate exchange optimisation until Section 2.5.4, we note that Xie *et al* made a strong assumption about the priors in their optimisation procedure, which to some extent weakened their method.

Xie *et al* assumed that the ensemble of locally optimal designs gave a solution equivalent to the globally optimal design. They sampled from the parameter prior, evaluated the optimal design for this combination of parameters, and then constructed a distribution of optimal points, using this to select their final design points. In the best case, at each f and Δt , the optimisation finds the global optimum for that combination of parameters. Essentially, this answers the question: “if $f = f_i$ and $\Delta t = \Delta t_i$, what would be the best set of inversion times to measure them?”. This optimisation is performed over a range of f_i and Δt_i sampled from the f and Δt priors, $p(f)$ and $p(\Delta t)$. They assumed that the ensemble of results also answers the question: “if $f \sim p(f)$ and $\Delta t \sim p(\Delta t)$, what would be the best set of inversion times to measure them?”. However, these are not equal in general, or even often [113]. In the ASL model they are not equal, and this leads to appreciable differences between the designs proposed by Xie and Santos and the true optimal designs, as is discussed in Chapter 3.

2.5.2.3 Implicit priors and experimental design

One point, which might not be immediately clear, is that one may view *any* ASL experiment as coming from an optimal design process. This is the case even when a radiographer makes *ad hoc* choices about acquisition parameters to use: effectively, in this example, the radiographer is using their prior knowledge of physiology (particularly transit times) and the ASL signal (particularly the location of its high-SNR peaks) to estimate a set of measurements that give high quality parameter estimates. This latter estimation step, of course, is not formalised as in conventional experimental design methods. Nevertheless, the underlying procedure is the same.

Recognising that we can view any process to generate acquisitions in this way, we can examine the method of Kramme *et al* to reveal the implicit assumptions

made concerning priors.

Kramme *et al* assume an acquisition has a given number of acquisitions, N_{tot} . They define a search space of N_{TI} equally-spaced inversion times, $t_i = t_{min} + i * \Delta$. They impose a design in advance where $N_i = \frac{(N_{tot} - N_{TI}) * f_i}{\sum_j^{N_{TI}} f_j} + 1$ with $f_i = i$ (linear), $f_i = i^2$ (quadratic), or $f_i = e^i$. Each of these reflects an assumed prior on the signal parameters and forward model.

The intuition that guided Kramme *et al* to these designs was that T_1 decay causes lower SNR at longer TIs – subsequently, more TIs are necessary at longer TIs to achieve a comparable SNR at all TIs. This is equivalent to a utility function based on the T_1 decay model rather than on the perfusion parameters. Moreover, the attempt to equalise SNR across these TIs is equivalent to different assumptions concerning the signal decay – this is, of course, an exponential decay in T_1 , but the imposition of the ASL kinetic model is treated here as an unmodelled confound, handled heuristically by using a linear or quadratic weighting scheme. They conclude that a quadratic weighting scheme, as shown in Figure 2.4, copes best with the dual demands of equalising SNR while improving (rather than hurting) ASL parameter estimation.

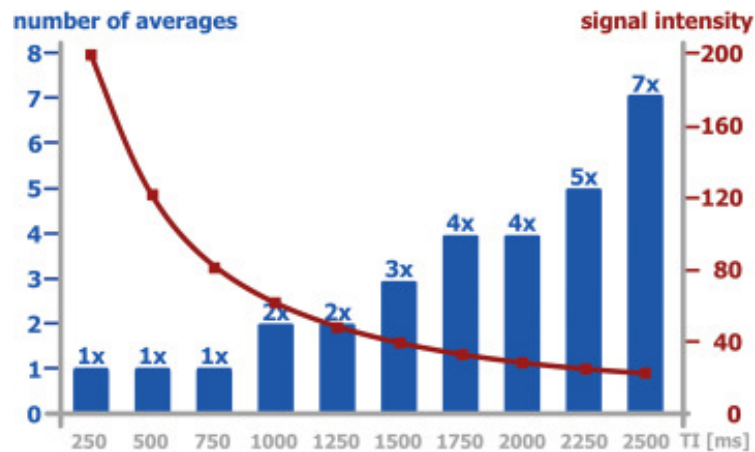


Figure 2.4: MR signal decay due to T_1 relaxation (red curve) was used as an argument for weighting the number of averages (blue bars) at different TIs differently, with quadratically more averages at the lower intensity regions. The figure is reproduced from Kramme *et al* [128].

2.5.3 Measurement constraints: fixed number of measurements versus fixed duration

The typical approach for optimal design in general is to assert that there are a given number of measurements to be taken. This matches well with the classic experimental design problems, and was the approach taken by Xie *et al* and Kramme *et al*. Clearly, a constraint on measurements is necessary: the entire motivation of optimal design for this problem is that scan time should be reduced or scan quality improved within the same scan time. However, the assumption of a fixed number of measurements leads to shortcomings in the optimal design process, as not all ASL measurements need to take the same amount of time. Our work (set out in Chapter 3) was the first to account for this during optimisation.

The fact that ASL measurements may not need to take the same amount of time, although a fundamental consequence of MR physics, is easily missed. Historically, most ASL acquisitions have used a single repetition time (TR) for all measurements. This TR has been set by constraints on basic feasibility (TR must be greater than TI plus a readout duration, by definition) as well as constraints on SAR. There is a complex trade-off between TI and TR, and one of the first attempts to explore this systematically with MR experiments came from Johnston *et al* [129], who showed that TR could be reduced for shorter TIs and hence acquisition times could be substantially shortened.

2.5.4 Optimisation for experimental design

Optimisation is necessary in the optimal design process to find a design giving the best value of the utility function described in Section 2.4. This optimisation is complementary to the normal use of optimisation in quantitative imaging methods: typically one has unknown parameters, θ , and known measurement parameters, η , and the task is to estimate $\theta|y \sim g(\theta; \eta)$. In experimental design, conversely, one has priors $\theta \sim p(\theta)$, and the task is to choose η to optimise $\int_{\theta} u(\theta; \eta)p(\theta)d\theta$.

Unfortunately, the optimisation of these design parameters is often more difficult than the counterpart optimisation to estimate parameters such as perfu-

sion [113]. This is because the presence of a (typically nonlinear) utility function, and marginalisation over parameter priors, can make optimisation drastically harder, particularly global optimisation.

There are several approaches to solve this problem. In certain special cases, analytic solutions may exist [113]. Failing this, one can resort to any of the global nonlinear optimisation approaches in the literature – variants on grid search, evolutionary search, multi-start linear search, etc [113, 116]. Often the loss function is expensive to evaluate, so a cheap-to-evaluate surrogate function may be used to estimate the loss and hence speed up optimisation [116].

It is very common, in experimental design problems, to use coordinate exchange as the optimisation procedure. In coordinate exchange, the curse of dimensionality is overcome by optimising design points separately in turn [113, 71], assuming this will ultimately lead to the globally optimal design. Various means exist to select which design point to exchange at a given time, and how to optimise a design point [116]. Global optimality in the nonlinear case is not generally guaranteed, but in practice for most realistic problems, performance of these approaches is very good [113]. In the approach of Xie *et al*, the exchange algorithm iterated over each design point (inversion time) in ascending order, performing line search between the neighbouring design points. This was iterated to convergence. Their approach was actually local design optimisation, as discussed in Section 2.5.2.2, meaning that this was computationally feasible. However, similar approaches are often used in global design optimisation, and as will be seen in Chapter 3, this underpinned our approach.

We note, in passing, that the design problem as posed by Xie *et al* can be solved analytically, and that their use of separate coordinate exchange optimisation and ensembling steps is valid, but unnecessary. We discuss this, and present a proof, in Appendix A. Unfortunately, in the problem as posed in our work, no such solution presents itself.

2.5.5 PASL versus PCASL

The choice of pulsed versus pseudo-continuous ASL might appear, at first, to make little difference to the optimal design methods considered here. However there are three key areas in which the choice can make a difference: the range of values for TIs and LDs, the sensitivity of the model to Δt , and the relationship between SAR and LD.

Although optimal design was first applied to PASL by Xie *et al* in 2008, and they noted then that their method was equally applicable to PCASL, optimal design has only been used for PCASL very recently – both by ourselves and by Wood *et al*. This is likely due to the ambivalent results achieved by Xie *et al* with their weighted ensemble approach: improved f estimation at the cost of Δt estimation is an unappealing prospect for PCASL, which already improves f estimation (compared to PASL) while being relatively insensitive to Δt . The move towards true globally optimal design somewhat alleviated this, and this is explored further in Chapter 3.

2.5.5.1 Ranges of values for parameters

The range of values for TI/LD will tend to be longer in PCASL. This is a fundamental consequence of the physics underpinning the ASL signal: Q2TIPS implementations typically have an upper limit for LD – PCASL acquisitions are much more flexible to use a long bolus, and when a longer bolus is used, the TI is correspondingly longer for the same PLD. Conversely, because of the improved SNR in PCASL, fewer acquisitions are typically needed to produce a usable image, which will in turn affect the optimisation results.

2.5.5.2 Sensitivity to transit time

When all other factors are the same, it is generally harder to measure Δt with PCASL as opposed to PASL, whereas PCASL offers more accurate and reliable f measurements. This is discussed in Section 2.2.2, and originates from the reduced sensitivity to the ATT. In our own work, this can be seen in the trade-off between f and Δt in one contrast versus the other. The worsening in Δt performance is significantly worse in PCASL.

2.5.5.3 SAR

PASL and PCASL have different relationships between LR and SAR. In PASL with a defined bolus (e.g. Q2TIPS, see Section 2.2.2.2), the RF energy deposited is proportional to $t - \tau$, whereas in PCASL it is proportional to τ [66, 23]. This can influence the allowable values of t and τ when designs are optimised. We discuss this in greater detail in Chapter 3.

2.6 Conclusion

We have presented a detailed survey of the ASL literature, covering MR physics, the development of ASL acquisitions, applications and processing pipelines. This provides background information for the thesis. In particular, we have conducted an in depth review of optimal design work in ASL, focusing on the differences between how the problem is posed, methods of solution, and the designs obtained. This section has linked the specific problem of ASL experimental design to the more general discussion of optimal design in Section 2.4. We have shown, most importantly, the motivation for a globally optimal design approach, and how this can make a difference compared to the method of Xie *et al.* This discussion sets the scene for Chapter 3 in which we present our original optimal design approach, including its applications to both PCASL and PASL, as well as simultaneous optimisation of label durations with inflow times.

Chapter 3

Bayesian experimental design for multi-parametric optimisation of ASL

This chapter presents work on optimal design for ASL, and is based on an original publication and subsequent work. Parts of this chapter were also showcased informally at the European Cooperation in Science and Technology on ASL in Dementia [130]. The publication focused on imposing more useful scan time constraints in a general optimisation framework which found true globally optimal and duration-constrained designs [5] for PASL experiments, while subsequent work in this chapter applied this to PCASL and examined *multi-parametric* optimisation of both label durations and inflow times. Chapter 2 gives a detailed exploration of the context in which this work was done, and compares the different ASL optimal design methods in the literature before our method.

3.1 Introduction

Here, we develop a Bayesian experimental design approach to optimise multiple delay ASL scans of any chosen duration, applying our method in pulsed and pseudo-continuous ASL, and showing that the same method can be used to jointly optimise several parameters to achieve an even greater improvement. This is one of the first applications of optimal design to pseudo-continuous ASL [131], and is the first use

of optimal design to simultaneously optimise inversion times and label durations. Moreover, our approach is one of the first such design methods to optimise the acquisition for a given scan time rather than a fixed number of measurements [131], allowing for a superior design and a more practically useful optimisation constraint.

In ASL measurements in the brain, blood is magnetically tagged at the neck by an inversion pulse, and then allowed to flow into the brain before the MR image is acquired (see Chapter 2 for more detail). ASL acquisitions have several choices of acquisition parameters, in particular the label durations (LDs) – the temporal width of the labelled blood water bolus; and inflow times (TIs) – the time between when the labelling begins and the image is acquired. These choices can have a significant effect on the accuracy and precision of the perfusion and arterial transit time estimates [60]. In previous work there have been attempts to optimise TI selection using experimental design [71, 128], which has led to small but statistically significant improvements in parameter estimation. This previous work, however, assumed a fixed number of TIs in the acquisition, and so did not account for the possibility of using a larger number of shorter TIs to achieve the total scan duration. The relaxation of this assumption, while offered as potential future work, has only recently begun to be explored [5, 131].

In this chapter we address this more realistic situation in which there is a fixed amount of scanner time available, and the task of experimental design is to select the best possible set of ASL measurements that fill this time. Here, such measurements are jointly optimised within a novel Bayesian framework for constrained experimental design of imaging experiments. Subsequently, we apply this same framework to the problem of simultaneously optimising TIs and LDs for the first time, yielding even further improvement in parameter estimates. We show that our optimisation framework is equally applicable to pulsed ASL (PASL) and pseudo-continuous ASL (PCASL), and to TI-only or joint TI-LD optimisation.

Using results from numerical simulations and experimental results from nine healthy volunteers, we demonstrate that our framework improves parameter estimation in ASL compared to a more conventional multi-TI acquisition with evenly-

spaced TIs. We also use simulations to investigate how the optimised design performs when f or Δt take abnormal values, as is the case in some pathologies. When optimising TIs for a five-minute PASL acquisition we achieve significant improvements in f estimation (approximately a 18% reduction in error), with no practical change in Δt estimation. When optimising TIs for a five-minute PCASL acquisition we achieve smaller improvements at a cost to Δt . However, when we jointly optimise TIs and LDs we achieve significant further improvements in f estimation (27% reduction in error) while boosting Δt estimation back to be comparable to the conventional design. Finally, this chapter briefly compares how the different treatment of priors in different methods leads to different proposed designs for the same set of priors, illustrating the effects of different methods side by side.

3.2 Theory

3.2.1 Arterial spin labelling

In ASL, blood water is magnetically tagged by spin inversion before flowing into the brain. Images are repeatedly acquired, with and without tagging, and the differences between these images are used to fit to a kinetic model. When images are acquired at several inflow times, one can simultaneously estimate perfusion (the amount of blood perfusing through the tissue) and arterial transit time (the time taken for blood to travel from the labelling plane to the image voxels) [60]. In this work, the single compartment kinetic model of Buxton *et al* [60] is used throughout to model the ASL signal, both for PASL and PCASL. All constants, where not stated, use the recommended values given by Alsop *et al*[23].

3.2.2 Bayesian design theory

The objective of Bayesian experimental design is to maximise the information gained from a set of experiments. Experiments yield a set of measured data points, y , which are related to the parameters of interest, θ , and the design parameters, η , by a forward model, $y_i = g(\theta; \eta_i) + e$. e here is additive Gaussian noise, the most commonly used noise model for ASL data. In multi-TI ASL, $\theta = (f \ \Delta t)'$ where f is perfusion and Δt is transit time, and $g(\theta)$ is the signal model of Buxton *et al* [60],

with $g(\theta; \eta) = \Delta M(\eta; f, \Delta t)$. η here is the set of inflow times, t_i . If one were to simultaneously adjust the LDs, τ_i , then η would contain both t_i and τ_i . We chose not to optimise over τ_i in our PASL experiments, as our PASL implementation was physically unable to use LD greater than 0.8s due to the size of the RF transmitter coil limiting the spatial extent over which tagging could occur. In our PCASL experiments, however, we optimised over t_i on their own, and t_i and τ_i jointly.

Because the noise model is Gaussian, maximisation of the information gain for a given θ is approximately equivalent to maximisation of the Fisher information matrix's determinant, $u(\theta, \eta)$ [113], where

$$u(\theta, \eta) = \det \sum_i \begin{pmatrix} \frac{\partial^2 g_i}{\partial \theta_1^2} & \frac{\partial^2 g_i}{\partial \theta_1 \theta_2} & \cdots \\ \frac{\partial^2 g_i}{\partial \theta_2 \theta_1} & \frac{\partial^2 g_i}{\partial \theta_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (3.1)$$

Initially, it is unclear what value of θ to use when evaluating this utility function. θ is not known ahead of time – indeed, it is θ that we seek to estimate. In a Bayesian approach, the utility is marginalised over a prior for θ , $p(\theta)$ [113]:

$$U(\eta) = \int_{\theta} \log u(\theta, \eta) p(\theta) d\theta = \int_f \int_{\Delta t} \log u(f, \Delta t) p(f) p(\Delta t) d\Delta t df \quad (3.2)$$

In the early history of Bayesian experimental design, experimenters sought to avoid the computationally demanding step of evaluating the *expected* Fisher information by evaluating it once at a representative point estimate of parameter values [113]. Subsequent work improved on this by sampling from the θ prior, and then optimising for each sample, making the assumption that the distribution of point-wise optimal designs reflects the optimal design for that prior [71]. This assumption is only approximately true, however, and cannot be used when there are certain constraints (in this work, scan duration) on η . Consequently, we use a numerical approach to approximate Equation 3.2, allowing us to find the true solution and respect feasibility constraints on η .

3.2.3 Computationally-tractable optimal design solutions

To find the expected utility for a given design, an adaptive quadrature technique [132] is used to approximate Equation 3.2. Using a high-performance implementation of an approximate global optimisation algorithm [133], the parameter space is iteratively divided into subregions, over which the integral is approximated. Subregions are refined preferentially when they have larger error, leading to highly accurate approximations of the overall integral. This estimate of the expected utility is then used as the utility function by which η is selected.

Throughout this work, $p(\theta) = p(f, \Delta t)$ is assumed to have f and Δt independently distributed, such that $p(f, \Delta t) = p(f)p(\Delta t)$. $p(\Delta t)$ is a normal distribution, with $\Delta t \sim \mathcal{N}(0.8, 0.3)$ s for PASL and $\Delta t \sim \mathcal{N}(1.2, 0.3)$ for PCASL. $p(f)$ is a log-normal distribution, with $f \sim \log \mathcal{N}(3.5, 0.5)$ (i.e. the mean value of f , as opposed to $\log f$, is approximately 40 ml/100g/min, and similarly $\sigma_f \approx 20$ ml/100g/min). These distributions were chosen to be broadly representative of physiologically-plausible f and Δt across the whole population [23], ensuring the optimised design works over a wide range of values. The f prior uses a somewhat low mean value to ensure that a broad range of f values can be included in the optimisation, even low f values due to partial volume effects or pathology. If more information were known ahead of time, such as reference values for a specific clinical population [71] or pre-existing measurements from a given patient, this could be used instead to further improve the design optimisation.

Performing an exhaustive search for the optimal solution is impractical, as there are many inflow times in a typical scan duration – in this work, approximately 30 such inflow times. In a naive exhaustive search, each inflow time is an additional dimension over which to search, meaning that this search cannot be performed on a realistic timescale. Fortunately, there is a simplifying symmetry in the utility function: when η is restricted to the inflow times, $U(\eta)$ does not depend upon the order of elements in η . This follows from Equation 3.1: overall utility is a function of the sum of individual utilities, making it commutative under reordering of inflow times. Thus, with no loss of generality, t can be constrained to be in increasing order.

Such a constraint lends itself to solution by a joint coordinate-exchange/line-search algorithm [71] [116], in which each inflow time is optimised separately, bounded between its neighbouring inflow times. Such an optimisation scheme can easily extend to joint optimisation of TIs and LDs by changing each line-search to a 2D grid-search, with each LD constrained to be less than its corresponding TI. Although there is no theoretical guarantee of global optimality, the results show good agreement with more time-consuming solutions from heuristic global optimisation methods such as controlled random search with local mutation [133, 134].

3.2.4 Constrained optimal design

Much of the experimental design literature concerns experiments with a fixed number of measurements. In ASL, and medical imaging more generally, this often is not the case. Instead, there is a fixed amount of time available in which to acquire data. Different acquisition parameters will result in a given measurement taking more or less time, and this constraint changes the optimal solution. Hence, in addition to the constraint that TIs are ordered, we impose a duration constraint, for our experiments here requiring that the whole ASL acquisition last no longer than five minutes. To calculate the experiment's duration, we set an experimentally-determined inter-acquisition delay (0.5 s in PASL experiments and 2 s in PCASL experiments, discussed in Methods) to wait between every TI, which allows the experiment to comply with MR Specific Absorption Rate (SAR) limits. We also enforce that Δt must be positive – effectively truncating its normal prior. The optimisation is performed, in parallel, over a range of TI list lengths. Finally, the resulting design with the highest utility is selected.

3.3 Material and methods

3.3.1 Experimental design optimisation

Computation of the optimal design was performed with a C++ implementation of the constrained optimisation approach discussed in Theory, using libcuba [132] for quadrature and NLOpt [133] for optimisation. Evaluation of the optimal design took approximately three hours on a modern computer using a 2.5 GHz In-

tel Xeon processor. In practice, computational time could be reduced through a more efficient optimisation scheme [116] and suitable approximation of the utility function [125]. With such approaches even real-time optimisation should be feasible [125], although these numerical improvements are beyond the scope of this work.

3.3.2 Variable TR acquisition

Typically, ASL images are acquired using a constant relaxation time (TR), regardless of the TIs used. In this work we use a variable TR, in order to take advantage of the time savings available from shorter TIs. Using a constant TR is a less efficient use of scanner time, but has remained standard due to being conceptually and practically easier. When varying the TR between measurements, several authors have shown the importance of pre-saturation pulses to remove residual signal [129]. Simulations and experiments have shown that, after appropriate pre-saturation, the effect of residual signal is small even at short TIs [129]. We follow this practice and use FOCI pre-saturation to remove residual signal from the previous measurement [135]. We set TR proportional to TI: for each measurement, $TR = TI + t_{pause}$. t_{pause} is fixed to a constant value across measurements, for the sake of simplicity in acquisition design. t_{pause} is set to 0.5 s for PASL and 2 s for PCASL, based on preliminary experiments and the simulation results discussed in Section 3.3.3.

Background suppression can be challenging in a variable TR acquisition and earlier work on variable TR ASL did not use background suppression as a result [129]. Other work has used background suppression, taking care to verify that the variable TR's affect on background suppression did not greatly affect results [131]. In this work, we use background suppression pulses *during* labelling, in a similar fashion to Dai *et al* [136] Most background suppression techniques in ASL are applied between labelling and acquisition, so when there is little time between these, the background suppression timing deviates further from its optimal values. We use two hyperbolic secant background suppression pulses, one during labelling and one after labelling, optimally timed [127]. If the quality of background suppression varies substantially between measurements, this will result in variable noise

magnitude – or worse yet, subtraction artifacts in the images. We guard against this by inspecting images manually, finding no visible sign of subtraction artifact. We also find our perfusion parameters are inline with literature values, with no sign of bias between different methods (see Section 3.4).

3.3.3 Specific absorption rate modelling

SAR can become a limiting factor in duration-optimised acquisitions. Using a variable TR acquisition reduces “dead” time in the experiment, but this also worsens the problem of SAR, as the RF energy is applied over a shorter time than in a typical sequence with constant TR. This is not a safety-critical concern in practice, as modern scanners have a safety interlock that aborts a scan when the SAR limit is reached. However, if optimisation yielded designs that could not be run due to SAR limits, it would clearly not be useful. Consequently, it is important that these optimised designs fit within SAR constraints. We achieve this through the use of a fixed pause, t_{pause} , after every measurement in our protocol.

PASL and PCASL have opposite relationships between LD and SAR; for PASL, $E_{PASL} \propto (t - \tau)$ whereas for PCASL, $E_{PCASL} \propto \tau$ [66, 23], where E is the RF energy deposited in each case. For the PASL experiments, in which τ is fixed to the reasonably common value of 0.8 s, t_{pause} was simply set to 0.5 s by early trial and error [5]. Because the optimised PASL design for $t_{pause} = 0.5$ s did not produce particularly long TIs compared to the reference design, and had the same LD, there was no reason to expect it to do any worse than the reference design in terms of SAR.

For the PCASL experiments, SAR was more of a concern, particularly given the long values of τ produced by initial runs of our optimisation procedure. We were hesitant to simply set t_{pause} by trial and error for PCASL, because this might have led to time-consuming experiments for each subject. Worse, this might have led to some values of t_{pause} working in only some subjects, making fair comparison of results difficult. Instead, we used pulse sequence simulations from the Siemens Healthcare IDEA platform to estimate the SAR. This allowed us to ensure feasibility of our optimised designs.

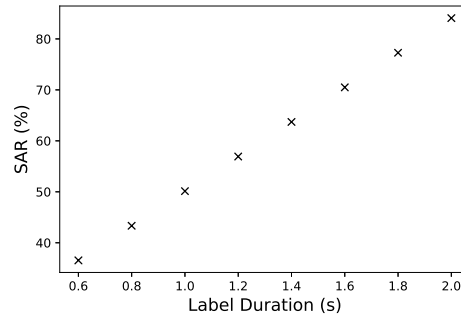


Figure 3.1: Scanner estimates of SAR, as a percentage of the allowable limit, versus label duration for PCASL. Energy deposition is linear with τ , matching published theoretical models.

First, we ran simulations for our sequence at several different values of LD. This allowed us to set the parameters E_0 and m in a linear model: $SAR_{PCASL} = \frac{E_0 + m \times \tau}{T}$, where $T = \sum_i t_i + t_{pause}$ is the total time over which the sequence is applied. We show this in Figure 3.1. This model could be used to estimate the SAR for any design, as defined by its TIs and LDs. This simulation was intentionally run assuming a heavier subject mass (90kg), as SAR issues are known to be worse in larger subjects. By choosing a heavier mass, we hoped to enforce a pessimistic estimate of the relationship between sequence and SAR.

Due to constraints in the platform, these simulations used the same pulses (labelling, readout, background suppression, etc) as our experimental acquisitions, but SAR was estimated for a given label duration repeated 10 times at $TR = 6$ s. This in turn allowed us to take SAR estimates for a given label duration and scale them to a given acquisition duration ($E = SAR \times T$). The “perfect” model fit of Figure 3.1 confirms that the scanner simulation matches the linear models of RF deposition in PCASL. We emphasise this fit cannot be taken as evidence of the simulation’s accuracy, and many sources have noted the relative *inaccuracy* of scanner SAR simulations [137]. However, it does show that our model matches the scanner’s own estimate of SAR, and hence that a design satisfying the SAR constraint from this model would avoid triggering the safety interlock.

We subsequently used this SAR relationship to ensure feasibility of our generated PCASL designs, requiring that the SAR was below the allowable limit. We

ultimately set t_{pause} to 2 s for the PCASL experiments, based on design feasibility and previous experience with the reference acquisition. Using the relationship of Figure 3.1, the SAR was estimated at approximately 90% of the limit for the optimised designs and 80% of the limit for the reference design. We did not search extensively over t_{pause} , although this might be a valuable addition in future work. SAR safety interlocks did not activate during our acquisitions, providing some evidence of our model’s suitability. Moreover, qualitatively, the scanner’s real-time estimate of SAR during acquisition matched the pattern of our calculated whole sequence SAR. The reference design had lower SAR in this real-time estimate than the two optimised designs, which were broadly similar to one another and were closer to the allowable limit.

3.3.4 Synthetic data

Synthetic data were generated from the appropriate forward model (PASL or PCASL) with additive Gaussian noise, chosen such that the SNR was broadly representative of real ASL data at $\sigma \approx M_0/100$ [71]. An example PASL time course is shown in Figure 3.2. Simulations, to assess performance across parameter space, were implemented by dividing the parameter space into a grid of 150 by 150 evenly spaced points and simulating 1000 noisy ASL signals at each point. In both PASL and PCASL simulations, the range of values was 0–150 ml/100g/min for f and 0.1–5.0 s for Δt , both chosen to cover the full range of plausible values relative to the prior distribution. This was done for both optimised and reference designs. Least-squares fitting was subsequently performed on each of these datasets to estimate parameters for both designs. All these simulations were performed in MATLAB, including implementation of the ASL forward model. In addition to assessing expected performance of optimal and reference designs, we used a similar approach to explore how the optimised design copes with abnormal parameter values, in Section 3.4.5.

Performance in the synthetic data was measured through the root mean square error (RMSE) and the expected coefficient of variation (CoV), which were both evaluated at each pair of parameter values based upon the 1000 estimates. By

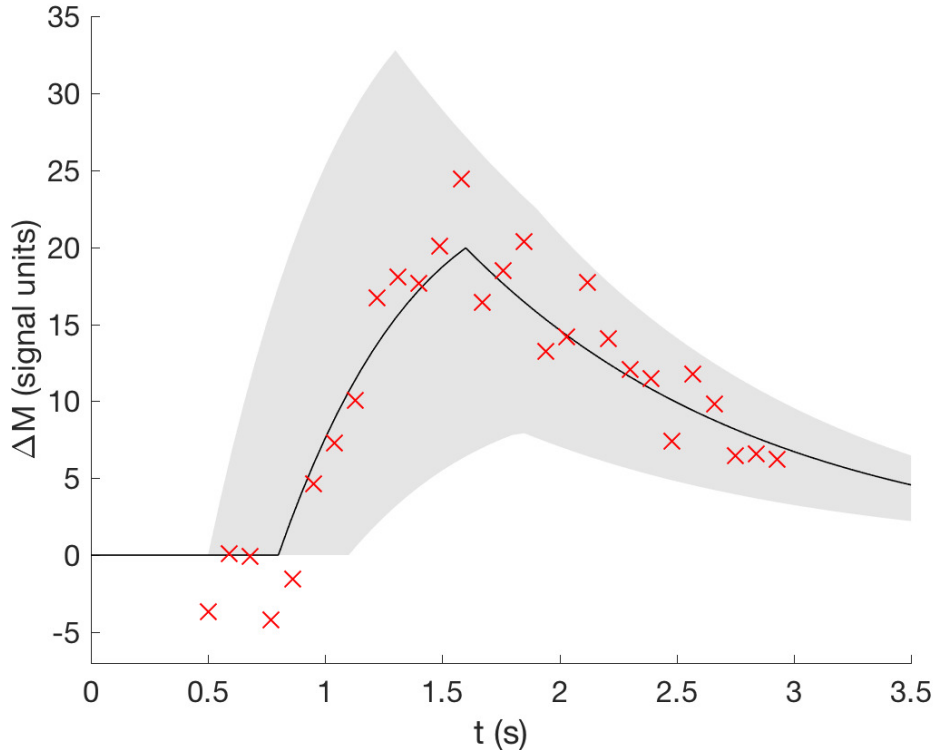


Figure 3.2: The black curve shows the PASL kinetic model, without noise, simulated at $f = 40$ ml/100g/min and $\Delta t = 0.8$ s. Labelled blood first arrives at $t = \Delta t$, and the signal magnitude peaks at $t = \Delta t + \tau$. The shaded region shows the variation in the curve within one standard deviation of the parameter means ($20 < f < 40$ ml/100g/min, $0.5 < \Delta t < 1.1$ s). \times shows example noisy measurements, which can be used to invert the model to estimate f and Δt .

definition, a better design produces estimates of the parameters that are closer to the true values. Because estimation in this problem is approximately unbiased, RMSE and CoV show good agreement, and the CoV is a more intuitive way to express results as a dimensionless indicator of the standard deviation: $CoV = \frac{\sigma}{\mu}$. The difference in the CoV, $\Delta CoV = CoV^R - CoV^O$, should be positive where the optimised design outperforms the reference. It expresses how much of a reduction in estimates' variance comes about from the new design, compared to the reference. The optimal design is that which minimises the generalised variance of the posterior, so assuming posterior independence of f and Δt , any design with $\sigma_f^O \sigma_{\Delta t}^O < \sigma_f^R \sigma_{\Delta t}^R \implies CoV_f^O CoV_{\Delta t}^O < CoV_f^R CoV_{\Delta t}^R$ is considered an improvement on the reference design.

3.3.5 Experimental data

For the PASL experiments, data were acquired from nine healthy subjects (seven male, two female; ages 22-34) using a 3T Tim Trio scanner (Siemens Healthcare, Erlangen, Germany). Q2TIPS was used to fix the bolus length to 0.8s [138]. For the PCASL experiments, data were acquired from five healthy subjects (three male, two female; ages 20–31). In both experiments a two-segment 3D-GRASE readout with background suppression was used, although the optimal design approach is applicable to any readout. f and Δt were estimated using FSL BASIL for ASL parameter inference [74]. Scan duration was fixed at 5 minutes for both optimal and reference scans. The specific TI and LD values used for the optimised and reference scans are described in the Results section.

In the absence of ground truth data, each of the optimised and reference scans was acquired twice, allowing the measurement of test-retest reproducibility, which is commonly used to judge potential acquisition improvements. To minimise the effects of subject motion and small drifts in perfusion values, measurements were acquired in an interleaved fashion, alternating between optimised and reference TIs. To assess the experimental results in the absence of ground truth values for f and Δt , we calculated whole grey matter test-retest coefficients for both parameters in all subjects, for both reference designs and optimised designs [71]. These were subsequently compared using a repeated measures ANOVA to test for significant differences [139], allowing us to show whether perceived changes in performance were statistically significant. When post hoc testing, to see which methods were significantly different to which other methods, we used Bonferroni correction – and p values are stated after adjustment. We subsequently examined parameter maps to confirm these results visually, as well as checking for any introduction of systematic bias.

In order to generate grey matter masks, MPRAGE T_1 -weighted structural scans were also acquired. The grey matter was segmented using pre-existing atlas-based segmentation software [140], thresholded at grey matter volume fraction ≥ 0.8 , before being registered to the ASL images using FLIRT [141, 142]. Finally, to allow

for estimation of M_0 , and hence absolute quantification of f , saturation-recovery images were acquired at three recovery times (1s, 2s, 5s).

3.4 Results

3.4.1 Proposed designs

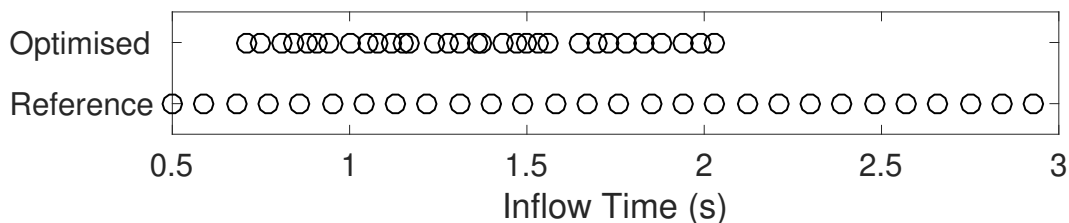


Figure 3.3: PASL optimised design (top) and reference design (bottom), defined by their TIs. Shifting TIs to more informative values near Δt and $\Delta t + \tau$ makes the acquisition more efficient, and avoiding longer TIs allows more measurements to be made in the same scan time.

The optimised design for PASL is shown in Figure 3.3, with the reference design also shown for comparison. (TI and TR values are also listed in Tables 3.4-3.5). The more conventional reference design used 28 TIs, equally spaced between 0.5s and 3s, chosen to be similar to a representative reference scan used in previous optimal design work [71, 5]. The optimised design used 32 TIs, which tend to cluster between 0.8s and 1.5s. This makes intuitive sense, as the forward model predicts higher signal magnitudes near $t \approx \Delta t + \tau$ – while also encouraging the use of some shorter measurements to estimate Δt . However, accounting for the effect of TI choice on scan duration, as done here, discourages longer, time-consuming TIs. This trade-off explains why the longest TIs are shorter than those in the reference scan. It also illustrates the value of this approach: the optimised scan not only chooses more informative TIs, but was able to fit in more TIs than the reference scan. It shows that it is preferable to use a larger number of shorter TIs rather than a smaller number of longer TIs, within reason, and this is reflected in the optimised design.

Although not shown here, the pattern of selected TIs remains broadly the same when the total scan duration is changed. Shorter duration constraints tend to reduce

the number of TIs in the tails of the distribution, especially longer TIs – that is, when there is less time available in which to scan, TIs near $t \approx \Delta t + \tau$ are prioritised for near-average values of Δt , with costly longer TIs tending to be discarded. Conversely, when there is more time available in which to scan, more points are placed at longer TIs, reflecting their value for measurements with long Δt .

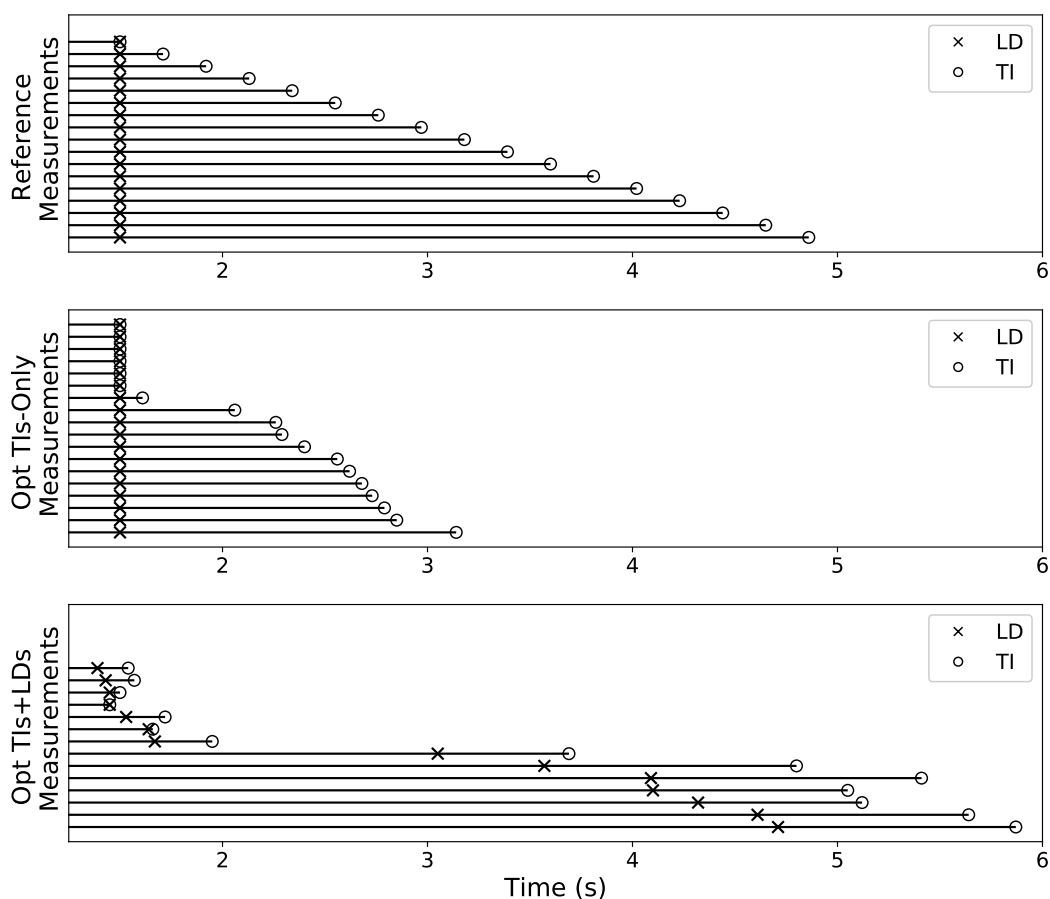


Figure 3.4: PCASL reference design (top), TI-optimised design (middle) and TI+LD-optimised design (bottom). \circ marks TI and \times marks LD. Note that we are specifying TIs rather than PLDs, for consistency with the PASL results.

The optimised design for PCASL is shown in Figure 3.4, with the reference design shown for comparison. (TI, LD and TR values are also listed in Tables 3.6–3.8). The reference design uses 17 TIs, the TI-optimised design uses 18 TIs and the TI+LD-optimised design uses 14 TIs. When only TIs are optimised, the result is conceptually similar to PASL: longer TIs are deprioritised, and TIs are clustered around the expected peak of the kinetic model. When TIs and LDs are optimised

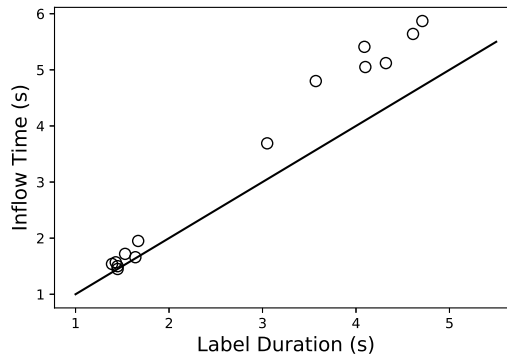


Figure 3.5: TI against LD for the the PCASL TI+LD optimised design. The line shows identity (TI=LD), highlighting how the gap between LD and TI increases at longer TIs. Long LDs are generally favoured (near as possible to the identity line) at lower TI values, but at higher values, this is deprioritised. Somewhat similarly to the PASL optimised design, measurements cluster roughly around Δt and $\Delta t + \tau$.

jointly, however, the result is quite different: the optimisation favours a cluster of shorter TIs (presumably to measure Δt) and very long TIs with long LDs (to measure f with a very high SNR, as the magnitude of the kinetic model increases with LD). Figure 3.5 shows the relationship between TI and LD in the joint optimisation in more detail. It shows that although LD increases with TI, it does not remain a constant offset below TI, but instead the gap between LD and TI grows with TI. This reflects the diminishing returns from longer LD as the signal reaches saturation.

3.4.2 Synthetic results

Table 3.1 summarises the expected improvements from the optimised design for PASL, compared to the reference. As expected, the performance is best near the prior's mean, because the design optimisation gives more weight to the high probability mass near here. Within the first standard deviation, f estimation improves on the reference design by approximately 27%, as measured by both RMSE and CoV ($P < 0.05$). Δt estimation, conversely, shows no significant difference. Improvements from the optimised design then decrease as they are evaluated over more of the parameter space. When evaluated over the whole space, f estimation improves on the reference design by approximately 17% ($P < 0.05$). Δt estimation is slightly impaired by the optimal design – decreasing by 7% ($P < 0.05$). However, this de-

crease is less than half of the increase in f performance, meaning the optimised design performs better overall. Over the entire parameter space, then, these results suggest there should be a large improvement in f estimation, and a smaller worsening of Δt estimation.

Table 3.1: Synthetic results for PASL, evaluated within one and two standard deviations of the prior, and over the whole space. Root mean square error (RMSE) is measured in the same units as the parameter, i.e. ml/100g/min for f and seconds for Δt . Δ RMSE and Δ CoV are the percentage changes between reference and optimised acquisitions, where positive change indicate better performance in the optimised design.

	f (ml/100g/min)						
	RMSE ^R	RMSE ^O	Δ	CoV ^R	CoV ^O	Δ	P
1SD: $20 < f < 60, 0.5 < \Delta t < 1.1$	0.27	0.20	27.2	0.761	0.552	27.4	0.0015
2SD: $0 < f < 80, 0.2 < \Delta t < 1.4$	1.24	0.99	19.9	3.24	2.61	19.5	0.0095
Whole space: $0 < f < 150, 0.1 < \Delta t < 5.0$	3.34	2.75	17.7	7.67	6.39	16.7	0.0124

	Δt (s)						
	RMSE ^R	RMSE ^O	Δ	CoV ^R	CoV ^O	Δ	P
1SD: $20 < f < 60, 0.5 < \Delta t < 1.1$	0.0051	0.0046	9.7	0.653	0.590	9.6	0.4966
2SD: $0 < f < 80, 0.2 < \Delta t < 1.4$	0.0216	0.0218	-0.1	3.02	3.02	-0.3	0.0861
Whole space: $0 < f < 150, 0.1 < \Delta t < 5.0$	0.0516	0.0553	-7.3	9.25	9.91	-7.1	0.0080

Table 3.2: Synthetic results for PCASL, evaluated within one and two standard deviations of the prior, and over the whole space. Root mean square error (RMSE) is measured in the same units as the parameter, i.e. ml/100g/min for f and seconds for Δt . Δ RMSE and Δ CoV are the percentage changes between reference and optimised acquisitions, where positive change indicate better performance in the optimised design. $p < 0.05$ for all.

	f (ml/100g/min)									
	RMSE ^R	RMSE ^{TI}	Δ	RMSE ^{TI+LD}	Δ	CoV ^R	CoV ^{TI}	Δ	CoV ^{TI+LD}	Δ
1SD: $20 < f < 60, 0.9 < \Delta t < 1.5$	0.624	0.503	19.4	0.415	33.5	0.92	0.75	18.4	0.62	32.6
2SD: $0 < f < 80, 0.6 < \Delta t < 1.8$	0.622	0.559	10.1	0.455	26.8	0.99	0.89	10.1	0.73	26.2
Whole space: $0 < f < 150, 0.1 < \Delta t < 5.0$	0.649	0.610	6.0	0.472	27.0	1.09	1.04	4.59	0.80	26.6

	Δt (s)									
	RMSE ^R	RMSE ^{TI}	Δ	RMSE ^{TI+LD}	Δ	CoV ^R	CoV ^{TI}	Δ	CoV ^{TI+LD}	Δ
1SD: $20 < f < 60, 0.9 < \Delta t < 1.5$	0.0094	0.0096	-2.13	0.0094	0.00	0.87	0.86	1.15	0.85	2.30
2SD: $0 < f < 80, 0.6 < \Delta t < 1.8$	0.0101	0.0109	-7.34	0.0099	1.98	1.03	1.07	-3.88	1.00	2.91
Whole space: $0 < f < 150, 0.1 < \Delta t < 5.0$	0.0106	0.0119	-12.3	0.0107	0.934	1.07	1.13	-5.61	1.09	-1.87

Table 3.2 summarises the expected improvements from the optimised designs for PCASL. As previously, the performance is best near the prior’s mean, because the design optimisation gives more weight to the high probability mass near here. Within the first standard deviation, f estimation in the TI-optimised design improves on the reference design by approximately 20%, as measured by both RMSE and CoV ($P < 0.05$). Δt estimation, conversely, shows no significant difference. Improvements from the TI-optimised design then decrease as they are evaluated over more of the parameter space. When evaluated over the whole space, f estimation in the TI-optimised design improves on the reference design by approximately 6% ($P < 0.05$). Here, optimisation of only the TIs leads to a practically significant deterioration in Δt estimation (10%), whereas joint optimisation of TIs and LDs gives no practical difference.

The effect of joint LD+TI optimisation is to significantly increase the improvements achieved in TI optimisation in f estimation, while removing the negative effects on Δt estimation. Joint optimisation improves f estimation by 45% within the first standard deviation, with this improvement decreasing gradually to 30% over the entire parameter space. Δt estimation also degrades further from the prior mean, but remains comparable to the reference design throughout.

3.4.3 Experimental results – PASL

Whole grey matter test-retest correlation coefficients are shown for f and Δt , in all subjects, in Figure 3.6. For f , these results show good agreement with the simulations: test-retest correlation coefficients are significantly higher for f in the optimised experiment ($p = 0.0087$). In the reference design, the mean f test-retest coefficient is 0.79, as opposed to 0.84 in the optimised design – an improvement approximately corresponding to the 20% improvement in RMSE and CoV predicted by simulations. For Δt , there is no clear difference, with the mean coefficient roughly equal to 0.70 in both designs, and tests do not suggest significant differences in the correlations ($p = 0.80$). Although simulations suggest that Δt estimation is slightly worse in the optimised experiment, the difference is small, and it is unsurprising that it cannot be detected in these experimental results.

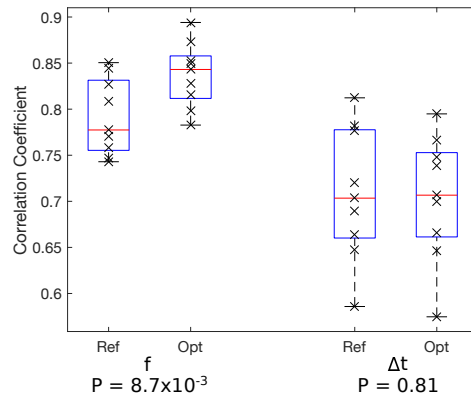


Figure 3.6: Boxplots of whole-GM test-retest coefficients for f and Δt in the PASL experiments. Boxes show lower and upper quartiles of the data, and the line within each box shows the sample median. Crosses mark individual subjects' coefficients. When there is no increase in bias, a higher coefficient is preferable, as it reflects a lower variance between test and retest.

Example parameter estimates from a representative subject (Subject 2, showing improvement in f estimation and no apparent difference in Δt estimation) are shown in Figure 3.7, along with the differences between test and retest images. By visual inspection, the f test-retest difference image shows slightly smaller differences for the optimised design, suggesting a greater consistency in its estimated results. This interpretation is supported by the higher test-retest correlation coefficient. There is no appreciable difference in the Δt images, which similarly agrees with simulation-based predictions and test-retest statistics. In both cases, optimised and reference images are visually similar, confirming that they are providing similar measurements of the same underlying distribution of parameter values.

3.4.4 Experimental results – PCASL

Example parameter estimates are shown in axial slices in Figure 3.9, along with the differences between test and retest images. Again, the f test-retest image shows smaller error for the optimised design, suggesting a greater consistency in its estimated results, supported by the higher test-retest correlation coefficient. Moreover, the TI+LD image has visibly smaller test-retest error than the TI-optimised image, showing the additional improvement from joint optimisation. These results are consistent with simulations and are significant in the test-retest correlation anal-

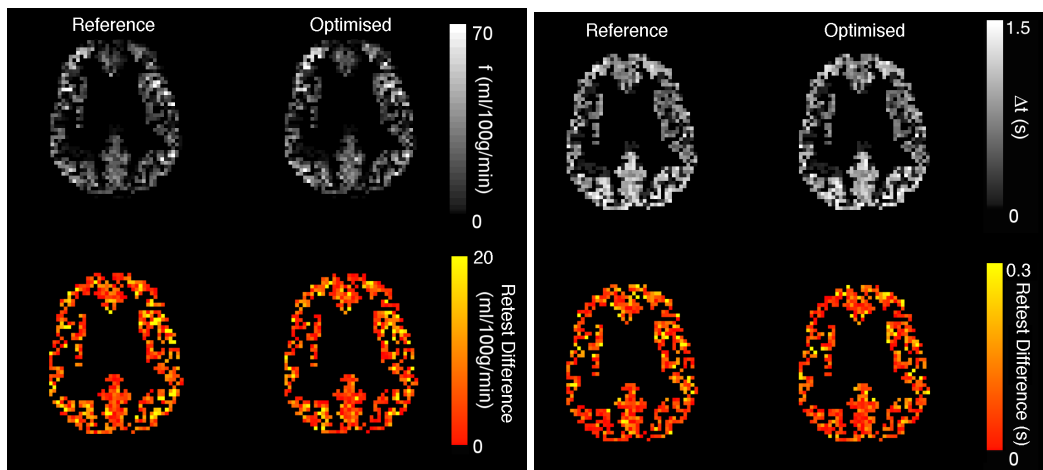


Figure 3.7: Example grey matter maps of f (left) and Δt (right) for Subject 4 from the PASL experiments. Absolute test-retest difference maps are also shown for each parameter. The same axial slice is shown throughout for comparison. Estimated maps are similar for both parameters, indicating no great change in bias, whereas test-retest difference maps are smaller for f and similar for Δt , indicating lower variability and hence improved performance.

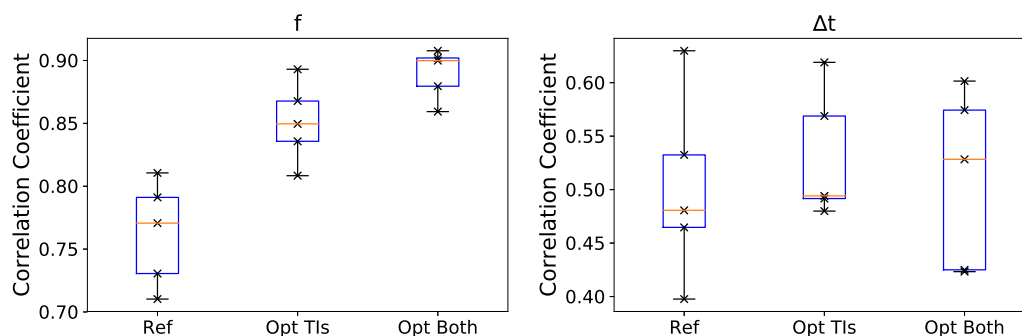


Figure 3.8: Boxplots of whole-GM test-retest coefficients for f and Δt from the PCASL experiments. Boxes show lower and upper quartiles of the data, and the line within each box shows the sample median. Crosses mark individual subjects' coefficients. When there is no increase in bias, a higher coefficient is preferable, as it reflects a lower variance between test and retest.

ysis (Figure 3.8). Running post hoc comparisons using repeated measures t-tests (p values stated after Bonferroni adjustment), optimisation improves f estimation ($p < 10^{-5}$ for each design versus reference), and joint TI+LD optimisation is better than TI optimisation alone ($p < 0.01$). Meanwhile, the ANOVA for Δt estimation does not show a significant difference between methods ($p = 0.275$).

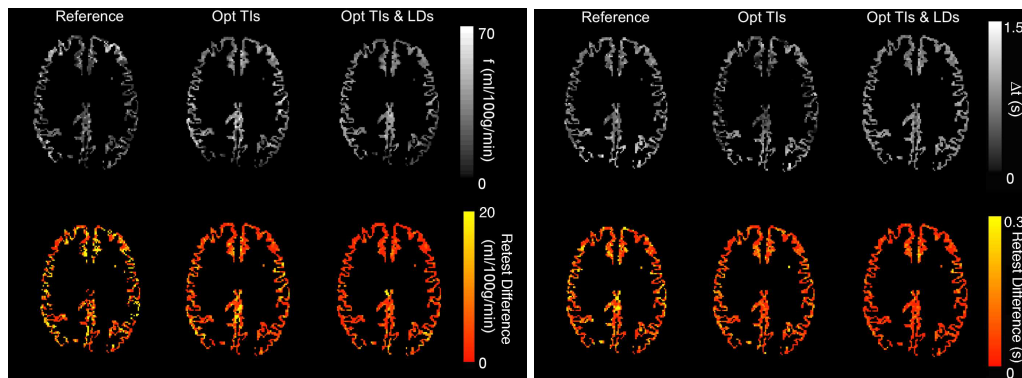


Figure 3.9: Example grey matter maps of f (left) and Δt (right) from the PCASL experiments. Absolute test-retest error maps are also shown. The same axial slice is shown for each parameter, for comparison. Estimated parameter maps are broadly similar, indicating no change in bias, whereas the maps of absolute test-retest error are clearly smaller in the optimised acquisitions, and lowest of all in the joint optimisation.

3.4.5 Robustness to abnormal parameter values

A summary of simulation results under extreme values of f or Δt , for the PASL design, is shown in Table 3.3. Generally, the optimised design performs better than the reference design, although it performs significantly worse when the transit time is prolonged. In each case, results are discussed further within their own section. To allow visualisation of the results and comparison of the designs, Figures 3.10 and 3.11 show 95% confidence intervals for the parameter residuals. Where these confidence intervals are reduced by a design, the CoV is correspondingly smaller, indicating improved estimation.

3.4.5.1 Abnormal perfusion

Results from abnormal perfusion simulations are shown in Figure 3.10. We simulate hypoperfusion by setting the mean perfusion lower than is typical, with f ranging from 20 ml/100g/min to a healthy value of 40 ml/100g/min, all with $\Delta t = 0.8$ s. Simulation shows that the optimised design continues to outperform the reference design (CoV for f improves by 10%, CoV for Δt improves by 38%), even in these challenging conditions.

We simulate hyperperfusion by setting perfusion higher than is typical, with f ranging from a healthy value of 40 ml/100g/min to 100 ml/100g/min, all with

Table 3.3: Simulation results for optimised vs reference designs under various combinations of abnormal parameter values. Positive change in RMSE and CoV indicate better performance in the optimised design.

	f						
	f	Δt	RMSE ^O	RMSE ^R	CoV ^O	CoV ^R	Δ RMSE (%)
Hypoperfusion	20	0.8	0.6384	0.7027	0.0319	0.0351	10.0664
Hypoperfusion	30	0.8	0.6341	0.7165	0.0211	0.0239	12.9827
Normal Perfusion	40	0.8	0.6296	0.7176	0.0157	0.0179	13.9682
Hyperperfusion	70	0.8	0.6323	0.7221	0.0090	0.0103	14.2094
Hyperperfusion	100	0.8	0.6315	0.7249	0.0063	0.0072	14.7793
Reduced Transit Time	40	0.4	0.4332	0.5449	0.0108	0.0136	25.7923
Reduced Transit Time	40	0.6	0.5076	0.6254	0.0127	0.0156	23.1968
Normal Transit Time	40	0.8	0.6296	0.7176	0.0157	0.0179	13.9682
Prolonged Transit Time	40	1.0	0.8392	0.8108	0.0210	0.0203	-3.3835
Prolonged Transit Time	40	1.2	1.3380	0.9488	0.0335	0.0237	-29.0844

	Δt						
	f	Δt	RMSE ^O	RMSE ^R	CoV ^O	CoV ^R	Δ RMSE (%)
Hypoperfusion	20	0.8	0.0167	0.0231	0.0208	0.0288	38.3774
Hypoperfusion	30	0.8	0.0109	0.0155	0.0137	0.0194	42.2734
Normal Perfusion	40	0.8	0.0082	0.0117	0.0103	0.0146	42.9196
Hyperperfusion	70	0.8	0.0047	0.0067	0.0058	0.0084	43.8514
Hyperperfusion	100	0.8	0.0032	0.0047	0.0040	0.0058	44.0825
Reduced Transit Time	40	0.4	0.0091	0.0108	0.0228	0.0269	17.8053
Reduced Transit Time	40	0.6	0.0080	0.0104	0.0134	0.0173	29.2129
Normal Transit Time	40	0.8	0.0082	0.0117	0.0103	0.0146	42.9196
Prolonged Transit Time	40	1.0	0.0100	0.0129	0.0100	0.0129	28.3804
Prolonged Transit Time	40	1.2	0.0140	0.0145	0.0117	0.0121	3.4890

$\Delta t = 0.8$ s. Simulations show, again, that the improved performance of the optimal design is relatively insensitive to the value of f . CoV for f estimation is improved by 14%, and CoV for Δt estimation is improved by 43%.

3.4.5.2 Abnormal transit time

Results from abnormal transit times simulations are shown in Figure 3.11. We simulate a prolonged transit time by setting Δt to be longer than is typical, with Δt ranging from a healthy value of 0.8 s to 1.2 s, all with $f = 40$ ml/100g/min. These are the only conditions under which the optimised strategy performs significantly worse than the reference strategy: on average, the CoV is 29% worse in the optimised design for f , and 3.5% better for Δt .

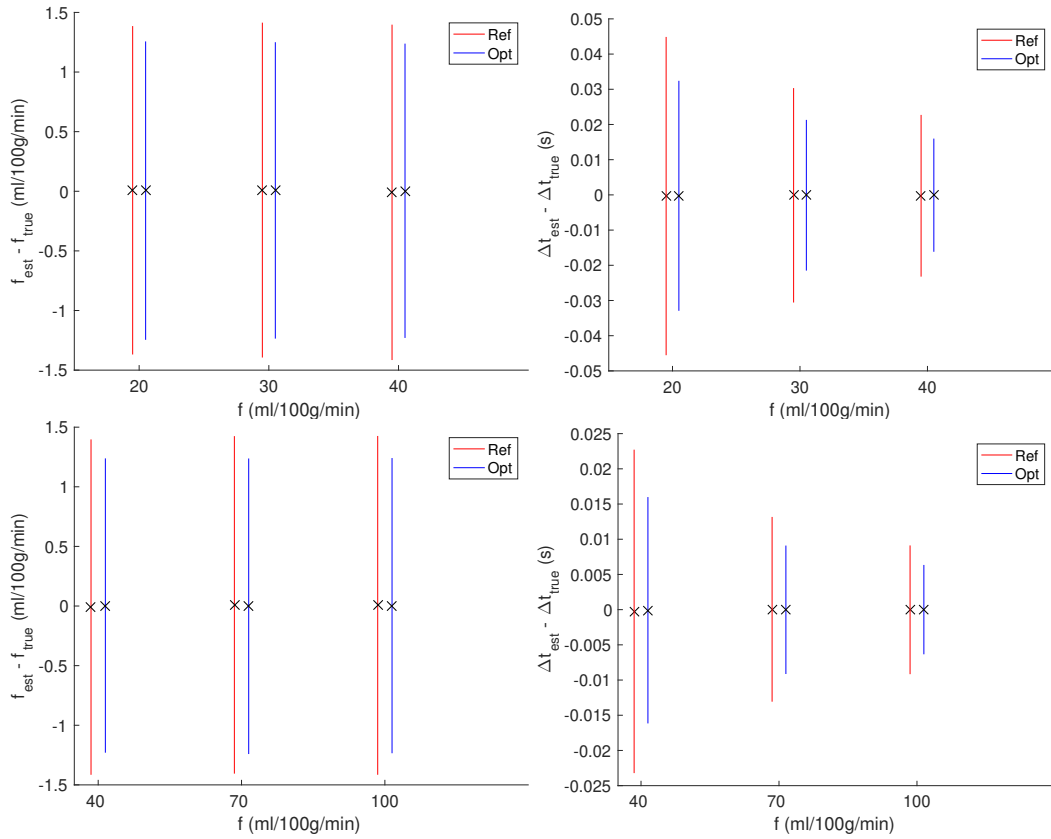


Figure 3.10: Simulation results for abnormal perfusion values. The top row shows results for hypoperfusion and the bottom row shows results for hyperperfusion. Bars show the 95% confidence intervals for f and Δt residuals at each f value, calculated over 10,000 noisy simulations, with reference design on the left (red) and optimised design on the right (blue) in each case. In each case, \times marks the mean residual.

We simulate a reduced transit time by setting Δt to be shorter than is typical, ranging from 0.4 s to a healthy value of 0.8 s, all with $f = 40$ ml/100g/min. Simulations show that the optimised design outperforms the reference design in this case, with the f CoV decreased by 25% and the Δt CoV decreased by 17%. f estimation is improved more than Δt estimation, as both the optimised and reference designs have a similar number of measurements near the curve discontinuity ($t = \Delta t$), which matters more for Δt estimation. Conversely, the optimised design has proportionally more measurements near $\Delta t + \tau$ here, making it significantly more useful for f estimation.

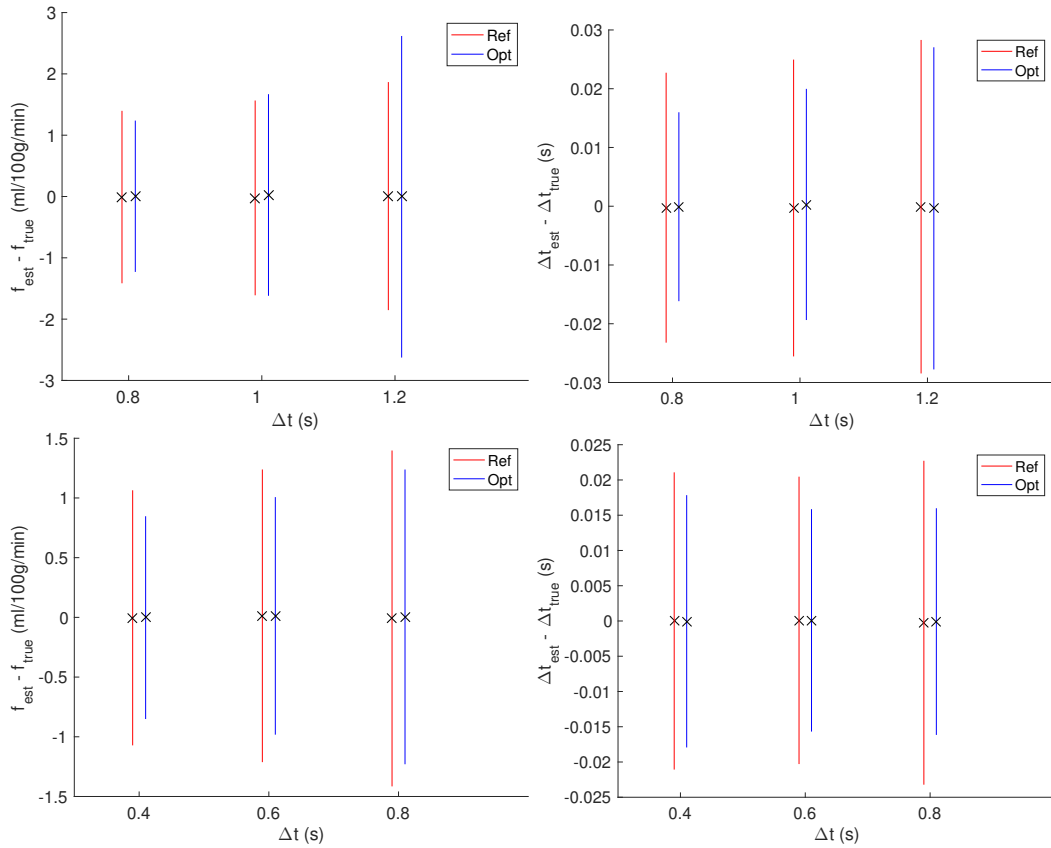


Figure 3.11: Simulation results for abnormal transit time. The top row shows results for prolonged Δt and the bottom row shows results for shortened Δt . Bars show the 95% confidence intervals for f and Δt residuals at each Δt value, calculated over 10,000 noisy simulations, with reference design on the left (red) and optimised design on the right (blue) in each case. In each case, \times marks the mean residual.

3.5 Different prior distributions and their effects on design optimisation

We show in Figure 3.12 how different priors affect the distribution of measurements in the different methods of optimisation. Our work and that of Xie *et al* has focused on truncated Gaussian and log-normal priors, to enforce positivity while getting a broader coverage of perfusion values, closer to covering an entire physiologically plausible range [5]. In both cases, the same techniques could have been used with any prior, however. Recent work has used uniform priors ($0.5 < \Delta t < 1.8$ s), arguing that these are fairest to use in the absence of knowledge of an individual's parameter distribution [131].

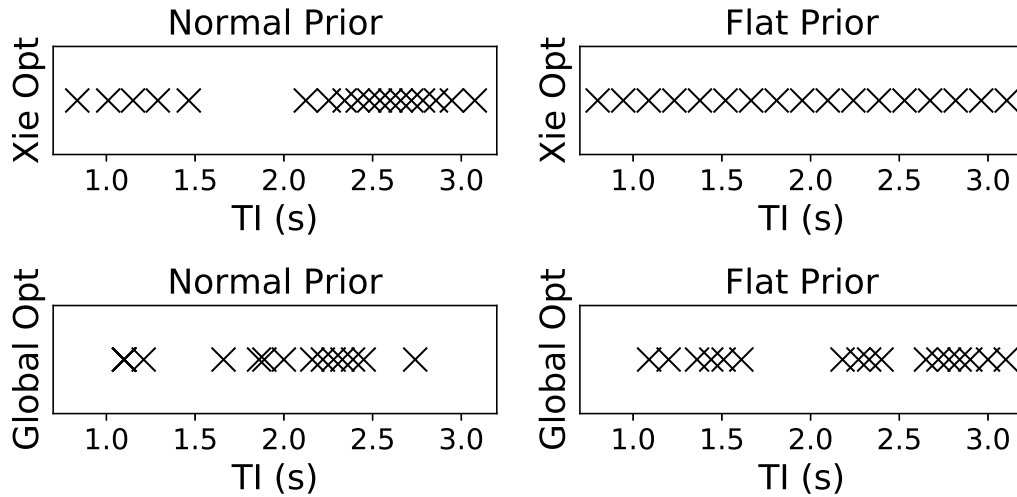


Figure 3.12: Optimised designs, Xie method versus truly global method, flat prior $\Delta t \sim U(0.5, 1.8)$ versus normal prior $\Delta t \sim \mathcal{N}(\mu = 1.15, \sigma = 0.325)$ such that two standard deviations of the normal prior reflect the entire range of the flat prior.

The Xie approach tends to undervalue shorter TIs in the case of normal priors, and in the case of uniform priors simply places TIs uniformly across that range. Conversely, in the truly global approach used by ourselves and subsequent work [131], we can see that flat priors and normal priors produce more conceptually similar results, clustering around key values. The main difference is that for the normal prior case, optimisation favours lower TIs because it is more willing to discount less likely, higher values of Δt .

3.6 Discussion and conclusions

The optimal design approach in this work has demonstrated its effectiveness in significantly improving ASL-based estimation of perfusion without greatly affecting arterial transit time estimation. The optimisation, for this five-minute ASL experiment, leads to a coefficient of variation improvement of approximately 20% as predicted by simulations, with a corresponding improvement in experimentally determined test-retest correlation coefficients. Simultaneous optimisation of TIs and LDs leads to further improvements in f estimation, and show the effectiveness of the optimal design paradigm. There is good agreement between simulations and experimental results, which demonstrates the validity of this model-based optimisa-

tion approach. Moreover, optimisation with a constrained scan duration allows for additional TIs to be used in the acquisition, which can improve robustness of the experiment.

Further simulations show the optimised acquisition is fairly robust to abnormal values of f or Δt in PASL, continuing to outperform the reference design under conditions of hypoperfusion, hyperperfusion and shortened transit time, performing worse only in the case of prolonged transit time. It may seem surprising that Δt estimation is improved in simulations of hypoperfusion and hyperperfusion, especially because it is not practically changed under normal perfusion conditions. There is an intuitive explanation, however: much of a design's ability to measure Δt comes from measuring the very start of the labelled blood's arrival, and the optimised design used $\Delta t = 0.8$ s as the mean value. Because that was the value of Δt in the abnormal perfusion simulations, the optimal design performs significantly better than reference for Δt estimation. Similarly, it is intuitively sensible that changes in Δt affect the optimised design more, which in turn describes the significant impairment under prolonged transit times: the free variables in the design optimisation *are* the inflow times, so a change in the transit time will drastically affect the design performance. Performance for Δt is impaired less, because Δt is easier to estimate from the low-magnitude signal near the discontinuity in the Buxton curve, when labelled blood just begins to arrive. Many of the time points from the optimised design are placed near this discontinuity, so Δt measurement is relatively unaffected. f estimation is impaired more because f is most effectively estimated from the higher-magnitude (and, more importantly, *later*) peak of the Buxton curve, near $t = \Delta t + \tau$. When the transit time is prolonged, the optimised design measures far less of the f -sensitive part of the curve, and hence performs worse at f estimation. The same logic explains why the design performs well for shorter transit times: more of the optimised design's inflow times are placed at shorter inflow times, meaning more of the curve is usefully measured in this case.

A challenge in this work is the problem of determining appropriate prior distributions for design optimisation. This is something of a double bind: if the parame-

ters were known ahead of time, it would be easy to optimise for them – but of course, there would then be no need to measure them. In this work we have mostly taken the approach of using broad priors based on the assumption of physiologically normal values in a young, healthy population. This would be less effective were we to use the same design for an older population. A simple solution would be to derive priors based on previous studies of similar populations, should they be available. Another solution might be to take even broader, flatter priors – for instance a uniform distribution extending to long Δt values. Yet another solution might be to use a rapid low-resolution pre-scan [136] or a closed-loop feedback system in which, within a given scan, intermediate fitting results inform optimisation to select subsequent acquisitions.[125]. All these approaches would be supported by our general method, although the latter would require numerical modifications for speed, as mentioned in Section 3.3.1. We note that the problem of determining an appropriate prior is present in some form however one selects TIs for an ASL acquisition: our optimisation approach makes this explicit, but the prior selection problem is present when one simply chooses a set of TIs – this always depends on selecting TIs in a useful range to measure the signal.

Although not examined in this work, there is an implicit trade-off between f and Δt estimation. This work improves f estimation (approximately 20% improvement in results' consistency) with no practical impairment of Δt estimation – this is the opposite case to previous optimal design work without duration constraints [71], which slightly improved Δt estimation while slightly impairing f estimation. We believe the reasons for Δt estimation remaining unimpaired are twofold. Firstly, the differences predicted from simulations are very small and difficult to measure. Secondly, the simulations do not account for how the increased number of inflow times in the optimised acquisition may improve robustness against outliers due to motion or hardware instability. The optimised acquisition has more TIs, so it is more robust against outliers, and might be expected to perform even better than simulations suggest.

Future work could examine the trade-off between parameters in the design op-

timisation, exploring how designs can be targeted towards particular parameters of interest. This might be useful where a cohort exhibits changes in one parameter (e.g. f or Δt), and researchers wish to study these changes in detail while still gathering information to estimate the other parameter. Future work might also explore duration-constrained optimal design using population-specific or subject-specific priors, adapting the ASL acquisition for maximum experimental efficiency when there are known abnormal parameter values associated with a particular pathology, or even more strongly informative subject-specific priors derived from previous measurements [125]. This should lead to a performance improvement even in subjects with prolonged transit time, where the optimised design currently performs worse than the reference. This would be valuable for imaging elderly subjects and patients with neurodegenerative diseases, who often present such prolonged transit times. The general optimisation framework in this work is applicable to many other imaging modalities, and further future work will also examine how it may be applied to other quantitative MR acquisitions, such as relaxometry or diffusion imaging. Finally, a promising direction for future work might explore how the optimal design approach may be applied to time-encoded ASL – an ASL technique that uses time encoding in the labelling phase to shorten scan durations when acquiring multi-TI data.

The choice of a five-minute window for scanning is comparatively brief for multi-TI ASL, but still provides reasonably consistent measurements, as shown in the experimental results. It is also useful for allowing test-retest measurements on both the optimised and the reference strategy within a 20-minute period. Moreover, such a time period is broadly representative of the realistic time available for ASL scans in multi-parametric research studies and in clinical practice. It would be valuable future work to explore real-time optimisation of the experimental design in this situation, allowing for the design to be optimised at the time of acquisition. Ideally this would also involve prioritisation of TI-ordering, to ensure viability of parameter estimates within a short time before refining parameter estimates' accuracies. This would be similar to adaptive optimal design approaches [125], but using the more

realistic duration-constrained optimisation presented in our work.

Our work, set out in this chapter, argued that previous work on optimal design in ASL was not finding true globally optimal design, but rather a prior-weighted ensemble of locally optimal designs. Our work was the first to find a true globally optimal design for an ASL experiment [5], with subsequent work by ourselves and others exploring this in PCASL [131]. Our work also explored multi-parametric optimal designs – designs in which inflow times and label durations were optimised simultaneously, to further boost acquisition efficiency. In addition, our approach has been used in other modalities [4]. The subsequent work of Woods *et al* strikingly uses a further approximation to ease the optimisation in ASL specifically [131]. They note that the utility function’s optimum is insensitive to the perfusion, because the perfusion’s contribution to the utility function comes primarily from a multiplication that can be factored out. Most of our work was done before this, and attempting to retain generality across many kinetic models relevant to MR imaging studies. As a result, we did not take advantage of this simplifying approximation. Our approach is hence significantly slower, albeit offering greater generality – hence its successful application to other imaging contrasts where such an assumption may not be available [4].

In conclusion, our work reduces the time needed for ASL experiments – or, equivalently, obtains better perfusion estimates from ASL experiments of a given duration. Poor image quality, only somewhat alleviated by lengthy scan durations, is often given as a major weakness of ASL [23], and our work directly improves this. Simulations show that the optimised design approach is robust to hypoperfusion, hyperperfusion and shortened transit times. Experimental validation confirms the improvement in f estimation, with no practical impact on Δt estimation. This work enables multi-TI ASL experiments to be performed in the most time-efficient way possible in a chosen duration, making the best possible use of scanner time to achieve a maximally-accurate perfusion image. We hope that this direction of research, and the subsequent work it has inspired, may eventually help increase the use of ASL in research studies and in clinical trials.

Table 3.4: PASL inversion times (ms) for reference and optimised acquisitions. Label duration is 0.8s for all measurements.

Reference	500	590	680	770	860	950	1040	1130	1220	1310	1400	1490	1580	1670		
	1760	1850	1940	2030	2120	2210	2300	2390	2480	2570	2660	2750	2840	2930		
Optimised	710	750	810	840	880	910	940	1000	1050	1080	1120	1150	1170	1240	1280	1310
	1360	1370	1430	1470	1500	1530	1560	1650	1700	1730	1780	1830	1880	1940	1990	2030

Table 3.5: PASL relaxation times (ms) for reference and optimised acquisitions.

Reference	1000	1090	1180	1270	1360	1450	1540	1630	1720	1810	1900	1990	2080	2170		
	2260	2350	2440	2530	2620	2710	2800	2890	2980	3070	3160	3250	3340	3430		
Optimised	1210	1250	1310	1340	1380	1410	1440	1500	1550	1580	1620	1650	1670	1740	1780	1810
	1860	1870	1930	1970	2000	2030	2060	2150	2200	2230	2280	2330	2380	2440	2490	2530

Table 3.6: PCASL inversion times (ms) for reference, TIs-optimised and joint-optimised acquisitions.

Reference	1500	1710	1920	2130	2340	2550	2760	2970	3180
	3390	3600	3810	4020	4230	4440	4650	4860	
Opt TIs-only	1500	1500	1500	1500	1500	1500	1610	2060	2260
	2290	2400	2560	2620	2680	2730	2790	2850	3140
Opt TIs+taus	1450	1500	1540	1570	1660	1720	1950	3690	4800
	5050	5120	5410	5640	5870				

Table 3.7: PCASL label durations (ms) for reference, TIs-optimised and joint-optimised acquisitions.

Reference	1500 (all)													
Opt TIs-only	1500 (all)													
Opt TIs+taus	1450	1450	1390	1430	1640	1530	1670	3050	3570	4100	4320	4090	4610	4710

Table 3.8: PCASL relaxation times (ms) for reference, TIs-optimised and joint-optimised acquisitions.

Reference	3500	3710	3920	4130	4340	4550	4760	4970	5180
	5390	6600	5810	6020	6230	6440	6650	6860	
Opt TIs-only	3500	3500	3500	3500	3500	3500	3610	4060	4260
	4290	4400	4560	4620	4680	4730	4790	4850	5140
Opt TIs+taus	2450	3500	3540	3570	3660	3720	3950	5690	6800
	7050	7120	7410	7640	7870				

Chapter 4

ADRIMO: Anatomy-DRiven

MOdelling of spatial correlation

This chapter presents work on spatial regularisation of ASL images, primarily in the context of denoising. The chapter is based on the original publication of this work [3]. We thank Stefan Spann and his coauthors for providing an implementation of the TGV denoising method to compare against.

4.1 Introduction

As discussed at length in Chapter 2, ASL images have an intrinsically low SNR. In addition to addressing these issues through acquisition improvements such as that discussed in Chapter 3, post-processing methods offer the opportunity of improving data quality by increasing the effective SNR, and hence the accuracy of the derived perfusion estimates [23]. Many methods exist to denoise ASL images, either through filtering [95] or explicit regularisation using a spatial prior during estimation of perfusion [99]. However, filtering approaches are often sub-optimal, relying on a trade-off between smoothing details in the data and failing to provide the maximum possible amount of regularisation. Moreover, filtering methods tend to be burdened with a myriad of tuning parameters, the choice of which can affect the quality of the results. These parameters make it hard to compare results from different experiments, and also complicate the ASL post-processing pipeline through their need for optimisation. Spatial priors do not generally suffer this burden, but

their performance suffers for ASL data acquired with a single post-labelling delay (the most common type of ASL acquisition) compared to multiple-delay ASL [90].

In this chapter, we present a technique, Anatomy-DRiven MOdelling (ADRIMO), for Bayesian data-driven spatial regularisation in single-delay ASL data. By using cross-modality spatial information from the available structural images, ADRIMO improves parameter estimate accuracy significantly more than existing regularisation techniques. At the same time, ADRIMO removes the need to set tuning parameters, allowing for more reproducible results with less effort to set up pipelines for ASL analysis. Building on our previous presentation of the theoretical underpinnings of ADRIMO and early experimental evidence of its efficacy [3], here we validate the approach through extensive ASL simulations, as well as comparing it to state of the art filtering methods. We show ADRIMO’s improved performance first through simulations of typical perfusion distributions, where mean absolute error is reduced by 20–80% relative to the state of the art; and then in experiments using an extensive (N=130) ASL dataset, where ADRIMO significantly improves test-retest repeatability and statistical power for detecting group differences in perfusion. Finally, in rigorous simulations and experimental data we show that performance is sustained even in the presence of highly focal perfusion changes, justifying the use of ADRIMO on potentially heterogeneous real-world data.

4.2 Denoising ASL

Image denoising is a fundamental problem in computer vision, and the methods that have been applied to ASL images have been inspired by predecessors in computer vision more broadly [143]. Initial approaches used local smoothing, typically with a Gaussian kernel [144]. The motivation for this is that spatial correlation between nearby voxels can be used to improve the effective SNR – and this same logic underpins practically all denoising methods, which differ in how they model that spatial (or spatio-temporal) correlation.

Subsequent authors explored more advanced means of spatial regularisation, including anisotropic diffusion [95, 145], wavelet-based and total variation regu-

larisation methods [144, 146]. Each of these was a breakthrough in the computer vision literature, and each led to many subsequent publications on application in different problems, MR imaging and ASL among them. These methods were applied as-is to ASL difference images, as well as combined with other steps in the ASL processing pipeline, such as temporal filtering [98] and partial volume correction [147]. One of the more thorough head to head comparisons was conducted by Wells *et al*, in the same work that introduced their own denoising approach [95]. Generally, subsequent methods have often been compared in a more ad hoc way, if at all.

Most recently, total generalised variation (TGV) stands out as the front runner for these “tunable” methods [92]. Such methods have several tuning parameters that must be set by the user. This is in many senses a disadvantage as it complicates the processing pipeline and makes it difficult to compare analyses [99, 148]. However, this general approach is common in regularisation methods generally, and usually such parameters can be set through cross validation or otherwise. TGV adopts a generalisation of total variation (TV) regularisation that has shown its superiority in the literature, and applies it to both ASL difference images and control images [149]. In their original publication, Spann *et al* examined denoising in both standard resolution and high resolution ASL data, showing the superiority of their method to several others from both classical computer vision and the ASL literature.

Of particular note is the denoising approach adopted in the popular BASIL software for ASL processing [74]. Adopting a fully Bayesian model, this method avoids the need to set tuning hyperparameters, and handles parameter inference under several variations of the different ASL models. Adaptive spatial regularisation can be combined with a physiological prior through extension to a Gaussian process model, although in practice the computational challenges associated with this lead to a spatial-only Laplace prior being used by default, initialised based on inference with a physiological prior. However, this approach has shown significant over-smoothing in single-TI ASL in some settings [90]; marginalisation and evidence optimisation significantly benefit from rich kinetic curve information provided by

multi-TI ASL, and may struggle in single-TI ASL.

4.3 Material and methods

4.3.1 Arterial spin labelling

In ASL, images are acquired with and without magnetic labelling of water in the feeding vessels. The difference between these images is related to perfusion by the model of Buxton *et al* [60]:

$$\Delta M = \frac{2\alpha M_0 f (1 - \exp(-\tau/T_{1b}))}{6000\lambda e^{PLD/T_{1b}}} \quad (4.1)$$

where ΔM is the difference image; f is the perfusion in ml/100g/min; PLD (in seconds) is the post-label delay at which the signal is measured; T_{1b} (in seconds) is the decay constant for longitudinal magnetisation of water in blood; τ (in seconds) is the bolus temporal length; α is the labelling efficiency; and λ (in ml/g) is the blood-tissue partition coefficient. All constants, where not otherwise stated, use the recommended values given by Alsop *et al* [23]. The model above is for ASL using pseudo-continuous labelling, but a similar model applies for pulsed labelling [60].

4.3.2 Theory and implementation

We use a hierarchical model in which spatial correlation is modelled by regions containing voxels with similar values, to the extent supported by the data [3]. Related approaches using manually defined regions of interest and different statistical models have been used in IVIM diffusion imaging [148]. Intra-region similarity is a justifiable assumption due to the later use of a noninformative prior: if intra-region similarity is not supported by the data, the fit will fall back to a broad and very weak spatial correlation. This would not typically be the case, however, as intra-region similarity is supported by the literature and perfusion statistics are often reported on a per-region basis [150, 151]. Parameter inference makes use of this intra-region correlation, resulting in large-scale, data-driven spatial smoothness. To define the regions in this work, we use lobar parcellations automatically derived from T_1 images [140], thresholded at per-voxel grey matter volume fraction greater than or

equal to 0.8[90]. In principle, however, our method could use any parcellation, and is not restricted to neuroimaging data.

We begin from the data likelihood for a voxel, index i , with ASL measurements $y_{i,:}$: where \mathcal{N} is a normal distribution, \mathcal{MVN} is a multivariate normal, $g(f)$ is the forward model of Equation 4.1, $\mathbb{1}$ is a vector of ones of the same length as the number of measurements, and \mathbb{I} is the identity matrix of dimension equal to the number of measurements:

$$p(y_{i,:}|f_i, \sigma_n) = \prod_j \mathcal{N}(y_{i,j}; g(f_i), \sigma_n) = \mathcal{MVN}(y_{i,:} - g(f_i)\mathbb{1}, \sigma_n^2\mathbb{I}) \quad (4.2)$$

As the noise standard deviation, σ_n , is unknown at this stage, we marginalise over it: $p(y_{i,:}|f_i) = \int_0^\infty p(y_{i,:}|f_i, \sigma_n)p(\sigma_n)d\sigma_n$. We use a conjugate inverse gamma prior, $p(\sigma_n) = \mathcal{IG}(\sigma_n^2; \alpha, \beta)$ with shape parameter α and scale parameter β , later intentionally setting $\alpha, \beta \rightarrow 0$ to make the prior noninformative. Reparameterising and combining these, where \mathcal{NIG} is normal-inverse-gamma, we get the following:

$$p(y_{i,:}|f_i) = \int_0^\infty \mathcal{NIG}(y_{i,:} - g(f_i)\mathbb{1}, \sigma_n^2\mathbb{I}, \alpha, \beta)d\sigma_n^2 \quad (4.3)$$

The integral of a \mathcal{NIG} distribution with respect to noise variance has a known solution in the form of a multivariate t-distribution, here denoted t_ν with ν degrees of freedom:

$$\Rightarrow p(y_{i,:}|f_i) = t_{2\alpha}(y_{i,:} - g(f_i)\mathbb{1}, \frac{\beta}{\alpha}) \quad (4.4)$$

Next we introduce the hierarchical prior structure: we assume that each region (throughout this work a lobe of the cortex) contains voxels with normally distributed perfusion values. μ and σ are hyperparameters controlling this distribution, and are unknown, so we again use a noninformative hyperprior to make them wholly data-driven. We use a noninformative Jeffreys' prior: $p(\mu, \sigma) = \frac{1}{\sigma^3}$. Applying Bayes' theorem, the joint posterior distribution for a region containing N voxels is:

$$p(f_{1:N}, \mu, \sigma | y_{1:N};) \propto \prod_{i=1}^N \{p(y_{i,:} | f_i) p(f_i | \mu, \sigma)\} p(\mu, \sigma) \quad (4.5)$$

We use a Monte Carlo Markov Chain (MCMC) approach to perform inference on the per-voxel perfusion, f_i , as well as the per-region distribution hyperparameters, μ and σ , using Gibbs sampling. This is initialised with least squares estimates, and over 100,000 iterations (1,000 discarded for burn-in), yields robust estimates on a timescale of minutes.

MCMC update equations for step $k + 1$ are derived as follows:

1. Update the mean hyperparameter by randomly sampling from $\mu^{k+1} \sim p(\mu | \sigma^k, y) = \mathcal{N}(\frac{\sum f_i^k}{N}, \frac{\sigma^k}{N})$, the marginal posterior for the unknown mean of a Gaussian distribution with N measurements [152].
2. Update the covariance hyperparameter by randomly sampling from $\sigma^{k+1} \sim \sigma | f_{1:N}^k, \mu^k, y = \mathcal{IG}(\sum_i (f_i^k - \mu^{k+1})^2, N - 1)$, where \mathcal{IG} is the inverse gamma distribution, the marginal posterior for the unknown variance of a Gaussian distribution [152].
3. For each voxel, generate a proposal voxelwise parameter, f_i^* , by sampling f_i in turn from $\mathcal{N}(f_i, w f_i^2)$, where w is a proposal variance.
4. Update f_i^{k+1} with this new value if $r \sim U(0, 1) < \alpha(f_i^*, f_i^k)$, otherwise $f_i^{k+1} = f_i^k$.

Proposal variance, w , is initialised at $w = 0.5$ for all regions. Periodically these are tuned to ensure approximately 25% of samples are being accepted [148]. The acceptance probability, α , is given by the ratio of likelihoods, as is standard in Metropolis-Hastings: $\alpha(f^*, f^0) = \frac{p(y | \mu, \sigma, f=f^*)}{p(y | \mu, \sigma, f=f^0)}$, where the proposal is f^* and the previous value is f^0 . That is, α is higher when the proposal f^* better matches the observed data than f^0 . The evaluation of α here uses our derivation of the joint distribution in Equation 4.5, evaluated at the limit $\alpha, \beta \rightarrow 0$.

In this way, ADRIMO provides full posterior distributions for the per-voxel perfusion, similarly to the inference approach of BASIL. To help analyse ADRIMO

results and compare with methods such as TGV, we set ADRIMO’s per-voxel perfusion estimates equal to their expected values from these posterior estimates. This is done by evaluating the average of the inferred f distribution, for each voxel.

4.3.3 Comparison methods

To assess ADRIMO, we compare with several other ASL spatial regularisation approaches: standard voxelwise fitting with no denoising, smoothing with a Gaussian kernel, regularisation via inference under an explicit prior [99], and spatio-temporal total generalised variation filtering [92]. These were chosen to provide a mix of the most commonly used approaches (voxelwise fitting, Gaussian smoothing), and the state of the art in both filtering methods (total generalised variation) and Bayesian methods (BASIL). All comparison methods are described in more detail below.

Voxelwise fitting Voxelwise fitting attempts no denoising; it simply involves estimation of perfusion values. To achieve this, we solve for f in the appropriate perfusion model (pulsed or continuous ASL, depending upon which type of labelling was used during acquisition) [60]. Throughout this work, fitting was performed using the open-source NiftyFit package [15]. Voxelwise fitting provides a simple baseline against which to compare all methods: it is still common not to perform spatial regularisation, and this is currently the recommended approach for standard ASL experiments [23].

Gaussian smoothing Gaussian smoothing is conceptually the simplest approach to spatial denoising, and is common in ASL processing pipelines. This consists of smoothing with a Gaussian filter with standard deviation σ mm. Smoothing is performed only within the masked region in which perfusion values are estimated, to prevent blurring noisy voxels without significant perfusion signal into the estimates. σ is a tuning parameter, and to some extent arbitrary. Here we first examine a range of σ values to show ADRIMO out-performs Gaussian smoothing for any value of σ . We subsequently use only $\sigma = 1$ mm to represent Gaussian smoothing, as this has lowest absolute error and is closest to common practice [23]. As with voxelwise fitting, Gaussian smoothed images were calculated using the NiftyFit package [15]. When voxels outside the grey matter mask were within the extent of the smoothing

kernel, they were excluded, and the kernel was subsequently re-weighted to correct for their exclusion. This prevents biasing of the smoothed values by partial volume effects [76].

BASIL inference The popular ASL processing software, BASIL, uses a data-driven method in which the extent of smoothing is estimated from the data, estimated jointly with the perfusion values. Conceived of by Groves *et al* [99], we use the software implementation in FSL’s *oxford_asl* [74]. This is a very effective technique in multiple-delay ASL, where there is a wealth of information to be derived from measurements of the full kinetic model [99, 74]. In single-delay ASL, however, there is less information available over the full extent of the model, and consequently the BASIL approach may be less suitable [90]. It is included here as the state of the art for Bayesian regularisation without tuning parameters, although it is plausible that methods with tuning parameters may out-perform it. We use default settings for parameters such as label efficiency, partition coefficient, etc. In practice, at least in this work, BASIL provides results comparable to or better than the other state of the art method (TGV), albeit with a larger bias.

Spatio-temporal total generalised variation (TGV) filtering TGV is a recently developed regularisation technique for single-delay data, which has demonstrated the best performance of any regularisation approach in an extensive comparison [92]. We include this technique as a gold standard in filter-based regularisation, and in single-delay ASL denoising generally. TGV regularisation for ASL requires the use of three tuning parameters: α_1/α_0 , s and λ .

The first parameter, α_1/α_0 , is inherent to TGV regularisation, and controls how regularisation behaviour varies between smooth and non-smooth parts of the decomposed image. Essentially, this weights the balance between TV and TV^2 regularisations, both different variants of TV. Notably, TGV is fairly insensitive to α_1/α_0 , and here we follow Spann *et al* in setting $\alpha_1/\alpha_0 = 1/\sqrt{2}$ throughout.

s and λ are specific to TGV for ASL, and must be tuned on representative data for optimal performance [92]. s controls the balance between regularisation of label images and regularisation of difference images. λ provides an overall trade-

off between data fidelity and regularisation, and in the ASL implementation is most sensitive to noise level [92]. In this work, both these parameters are optimised for experimental data and simulated data in a similar fashion to Spann *et al*, using the single subject dataset with a large number of repeats as gold standard (Section 4.3.4) and optimising on a smaller subset of this data to minimise mean absolute error (MAE). A subset with 30 random repeats was used, to be comparable to the number of repeats in the main dataset. This yielded $s = 0.525$ and $\lambda = 3.8$, which are similarly to parameters used in the literature [92]. These values are used throughout this work.

4.3.4 Validation for representative patterns of typical perfusion

4.3.4.1 Simulations

In order to validate our method for typical patterns and values of cerebral perfusion, we initially used simulations. Simulations were of the following types: *flat*, where white matter perfusion was 20 ml/100g/min and grey matter perfusion was 70 ml/100g/min; and *normal*, where white matter perfusion was sampled randomly from a normal distribution with mean 20 ml/100g/min and standard deviation 5 ml/100g/min and grey matter perfusion was sampled randomly from a normal distribution with mean 70 ml/100g/min and standard deviation 20 ml/100g/min. These were chosen to represent relative extremes of spatial correlation: in *flat* simulations there is maximal spatial correlation as the underlying perfusion values are the same throughout, whereas in *normal* simulations, spatial correlation exists only to the slight extent that all perfusion values are sampled from the same physiologically-plausible normal distribution. Both of these cases are based upon simulations from previous denoising and regularisation publications in arterial spin labelling and diffusion imaging [92, 153], and by covering the extremes of spatial correlation, the results form a complete picture of how different regularisation methods perform. Partial volume effects were included in all simulations, with each voxel contributing both white matter and grey matter perfusion, weighted by its corresponding volume fraction – this is why the *flat* simulation does not simply show one uniform value of perfusion throughout the brain.

Simulations were based on high-quality ASL images taken using the same readout as our main experimental dataset (Section 4.3.4.2), but with many more repeats (120 repeats, as opposed to the 30 used in the main dataset). The subject used to provide these simulation basis scans was a healthy male 22-year-old volunteer. The benefits of this approach are twofold. Firstly, we have representative and realistic control images to use as the basis for simulations with imposed perfusion values. Secondly, this gives a realistic estimate for the spatial distribution of noise magnitude: in ASL noise is well-approximated by white noise [99, 92, 23], but without these data we would not know the magnitude of noise to impose across different voxels.

Simulations used 40 difference images, with white noise added to each voxel according to the noise magnitude estimated from the initial high-quality dataset. On average the SNR of this noise (for a single measurement) was approximately 40%, relative to the ASL signal for $f = 70$ ml/100g/min in a purely grey matter voxel. This varied spatially from as high as 10:1 to approximately 1:1 near the edges of the brain. To compare methods, we denoised and estimated perfusion for each simulation using each regularisation method. We examine the resulting fits for bias and variance, which both contribute to the overall estimate error – that is, the distance of an estimated value from the underlying true value. Ideally, bias should be close to zero and the estimate error should come mostly from the fit variance, so we first assess bias from the simulation results to verify that this is the case. Subsequently, if a given method produces estimates with smaller error in simulation fits, this is taken as evidence of its superiority.

4.3.4.2 Experimental validation

In addition to simulations of typical perfusion maps, we perform experimental validation using a large dataset (N=130) of preterm-born adolescents and age-matched controls. First, we assess the different regularisation methods based on their test-retest repeatability. Each subject's ASL series is split in half, and then the test-retest coefficient is calculated for that subject as the correlation between per-voxel values in the first half and the second half of the data. This can serve as a proxy for the

error in the absence of a ground truth, given (as we show in simulation results) there is no increase in bias. If the test-retest repeatability is significantly increased, this may be taken to demonstrate improved performance. This is analysed with repeated measures ANOVA, extending a paired t-test to multiple methods and allowing for post hoc paired t-test comparison between methods [139].

We make further use of the experimental data by testing how different fitting methods affect the statistical power with which perfusion differences can be detected between groups. It has been shown previously that differences in cerebral perfusion exist between male and female subjects [154], so these differences can be used to assess detection capabilities for each method. A method that detects these known differences with higher statistical power is better – assuming it does not also increase the false positive rate. We compare, for each method, the p value with which significant differences between these groups are detected. We perform similar tests for differences between term-born and preterm-born subjects, where differences are less well-established but have previously been shown in this cohort [16]. As a control, we compare preterm subjects delivered by Caesarean section with preterm subjects who had normal vaginal delivery, where no perfusion difference is expected to be present. To show improved performance, a regularisation method should increase the statistical power (decrease the p value) where differences truly do exist between groups, without falsely detecting inter-group differences that do not exist.

In this set of experiments we use ASL images from 130 19-year-old subjects, 81 born extremely preterm (F:M=48:33, < 26 wks gestation) and 49 term-born peers (F:M=31:18). Informed consent was obtained from all subjects. Images were acquired on a 3T Phillips Achieva with 2D EPI pseudo-continuous ASL using 30 control-label pairs, PLD=1800ms+41ms/slice, $\tau = 1650$ ms, voxel size $3 \times 3 \times 5$ mm. We also acquired SI_{PD} images and 3D T_1 -weighted volumes at 1 mm isotropic resolution for segmentation and parcellation. Analysis was restricted to grey matter, masked by thresholding the segmentation at 0.8. We use a pre-existing tool to derive lobar parcellations corresponding to the Neuromorpho-

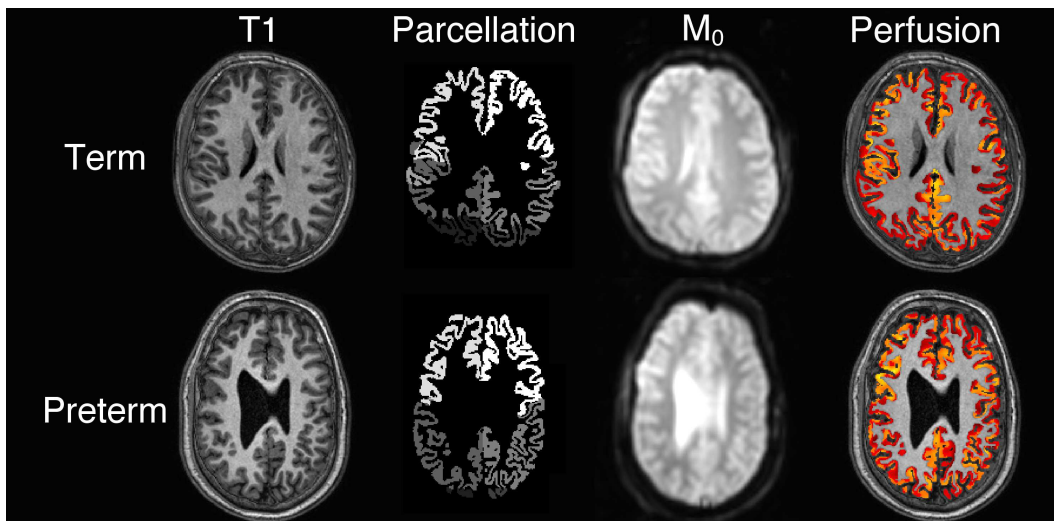


Figure 4.1: Examples of the experimental data, shown in an axial slice. From left to right: T1-weighted structural image, lobar parcellation, proton density image, perfusion-weighted image. The top row shows data from a term-born subject, while the bottom row shows data from a preterm-born subject.

metrics atlas [140]. We separate the grey matter into parietal, temporal, frontal and occipital lobes, dividing each of these lobes into left and right halves. Example data are shown in Figure 4.1.

4.3.5 Validation in the presence of focal perfusion changes

A common pitfall for spatial regularisation methods is to smooth excessively, resulting in the loss of fine detail in the data. This is particularly problematic when there is a small region of focal perfusion change, which may be lost from the image during regularisation. To assess the extent of this for different spatial regularisation techniques, we used simulations and experimental data with highly focal perfusion changes.

4.3.5.1 Simulations

We performed similar simulations to those discussed in Section 4.3.4, referring to them as *focal* simulations. In *focal* simulations, the *normal* perfusion map had focal regions of grey matter hypoperfusion (20 ml/100g/min) and hyperperfusion (120 ml/100g/min) imposed. To assess the effects of spatial regularisation across different sizes of focal perfusion change, the focal regions were chosen to be cuboids of length two voxels, four voxels and eight voxels. As with previous

	Voxelwise	Gaussian			TGV	BASIL	ADRIMO
		1mm	2mm	3mm			
Bias (%)	0.33	0.02	2.29	3.71	0.19	0.43	0.01
CoV (%)	20.9	18.7	10.4	8.51	7.53	3.01	0.590
MAE (ml/100g/min)	3.66	3.34	3.21	3.80	1.86	1.37	0.317

Table 4.1: *Flat* simulation: bias (as a percentage of the mean perfusion), coefficient of variation ($\text{CoV} = \sigma/\mu$) and mean absolute error (MAE) for each method.

simulations, we assess regularisation techniques based upon their error relative to the ground truth values, where smaller error is preferable. We further examine how well the focal regions were preserved, qualitatively, in the fitted images.

4.3.5.2 Experimental validation

Subsequently, we compared fitted perfusion values and regularisation performance in a 73-year-old female subject with a large arachnoid cyst, which leads to a large focal hypointensity in both the T1-weighted image and the perfusion values. The ASL data from this subject were used to estimate perfusion, where ADRIMO-generated perfusion maps are compared with maps from other approaches. Images were acquired on a Siemens Trio 3T with Q2TIPS pulsed ASL using seven control-label pairs, with $PLD = 1200$ ms and $\tau = 800$ ms, 3D-GRASE read-out, voxel size $1.875 \times 1.875 \times 4$ mm. A proton density image was acquired for calibration, with $TR = 5000$ ms, and the total acquisition time was five minutes.

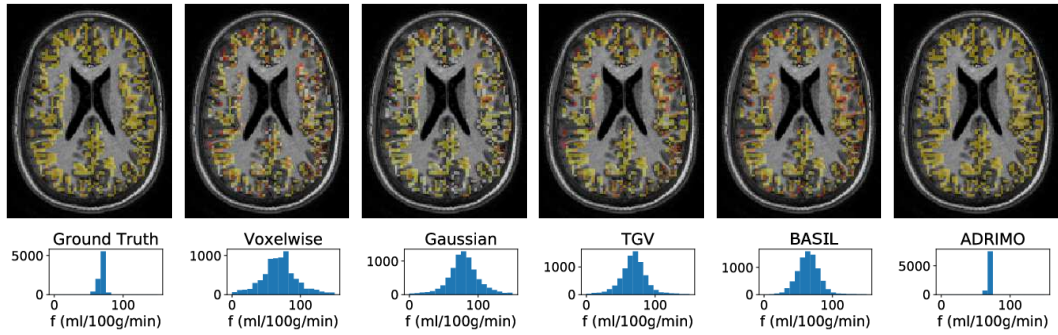
4.4 Results

4.4.1 Typical perfusion

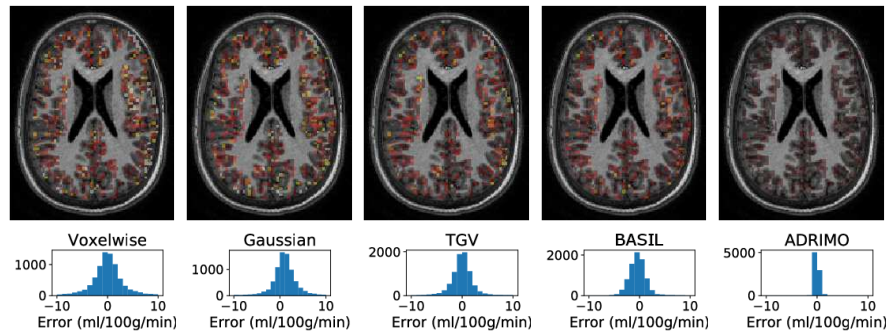
4.4.1.1 Simulation results

Simulation ground truths and fits are shown in the top row of Figure 4.2 for *flat* simulations and the top row of Figure 4.3 for *normal* simulations. Maps of absolute error between ground truth and fitted value are shown in the bottom row of each figure. Finally, Tables 4.1 and 4.2 show mean absolute error, error bias, and coefficient of variation for the errors.

Bias in the *flat* simulation is negligibly small in almost all methods, becoming



(a) Ground truth and estimated perfusion for several regularisation methods. Top row: example slice, bottom row: distribution of grey matter perfusion estimates.



(b) Error between estimated perfusion and ground truth. Top row: example slice showing absolute error, bottom row: distribution of error over grey matter.

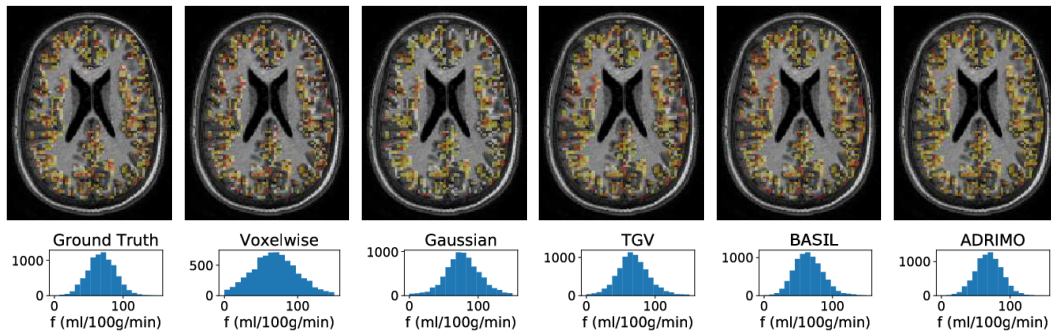
Figure 4.2: Ground truth, fitted values and errors for the *flat* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$.

appreciable only in Gaussian smoothing with unrealistically large kernel widths, where it can reach levels (3.7%, $p < 1 \times 10^{-10}$) that might have a meaningful effect on data analysis. Notably, Gaussian smoothing is very sensitive to the kernel width: when using $\sigma = 1\text{mm}$, the bias falls to non-detectable levels (0.02%, $p = 0.94$). Otherwise, the only method to increase bias beyond voxelwise fitting is BASIL, where bias is 0.4% with $p = < 1 \times 10^{-10}$. ADRIMO produces the smallest bias of any detectable method, at 0.01%, $p = 0.025$. Bias in the *normal* simulation shows a similar pattern, although here any amount of Gaussian smoothing increases bias. Again, Gaussian smoothing with extremely large kernel widths has an appreciable bias (3.86%, $p < 1 \times 10^{-10}$), but all other methods produce negligible bias and only BASIL significantly increases bias beyond voxelwise fitting. Again, ADRIMO is one of the methods presenting least bias of those with any detectable bias, and is far

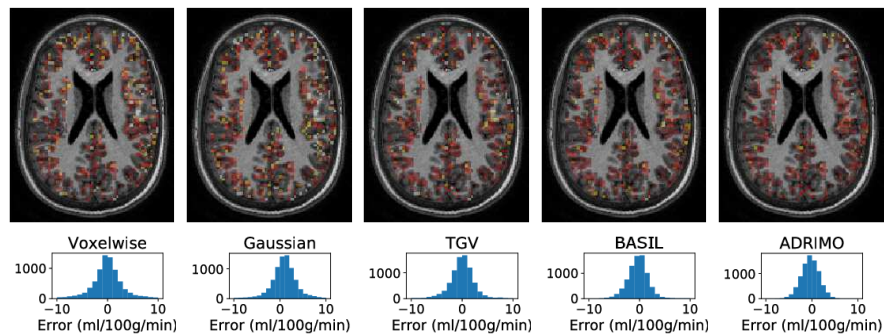
	Voxelwise	Gaussian			TGV	BASIL	ADRIMO
		1mm	2mm	3mm			
Bias (%)	0.02	0.286	2.54	3.86	0.03	0.326	0.03
CoV (%)	17.5	15.7	9.38	8.49	8.65	3.50	3.23
MAE (ml/100g/min)	3.59	3.27	3.36	3.92	2.05	1.63	1.30

Table 4.2: Normal simulation: bias (as a percentage of the mean perfusion), coefficient of variation ($\text{CoV} = \sigma/\mu$) and mean absolute error (MAE) for each method.

below the level of practical significance (0.03%, $p < 1 \times 10^{-10}$).



(a) Ground truth and estimated perfusion for several regularisation methods. Top row: example slice, bottom row: distribution of grey matter perfusion estimates.



(b) Error between estimated perfusion and ground truth. Top row: example slice showing absolute error, bottom row: distribution of error over grey matter.

Figure 4.3: Ground truth, fitted values and errors for the *normal* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$.

ADRIMO produces the lowest mean absolute error in both *flat* and *normal* simulations. This is most striking in the *flat* simulation, where ADRIMO's shrinkage property allows accurate detection of the near-constant perfusion map, leading to a mean absolute error of 0.32ml/100g/min compared with 1.37ml/100g/min for BASIL. ADRIMO also achieves the best performance in the *normal* simulation, al-

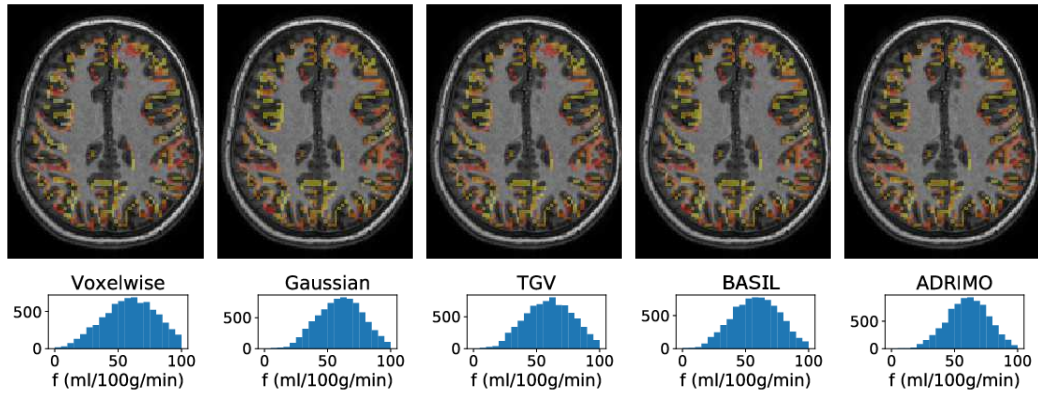
though the margin is smaller here: 1.30ml/100g/min versus 1.63ml/100g/min. The perfusion and error maps for the *flat* simulation, shown in Figure 4.2, illustrate the difference in methods' performance. Denoising methods lead to smoother, more accurate perfusion estimates in all cases. Absolute error is made smaller by all methods, and is reduced most visibly in the ADRIMO image. The more advanced denoising methods (TGV, BASIL, ADRIMO) do particularly well at ridding the error map of extreme errors, which corresponds to their lower coefficient of variation and narrower tails. These points are also true of the maps from the *normal* simulation, shown in Figure 4.3, with ADRIMO creating the perfusion map closest to ground truth and reducing errors.

4.4.1.2 Experimental results

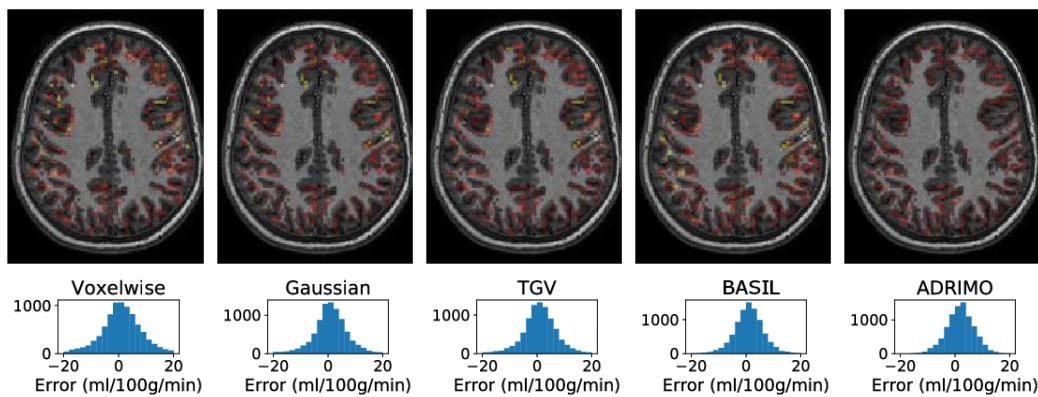
Figure 4.4 shows an axial slice from an example term-born subject, with perfusion estimates in the top row and absolute test-retest differences in the bottom row for each regularisation method. Consistently with the simulations, the test-retest differences are visibly smaller for the ADRIMO method than for others, while the estimated values show no sign of a bias in any method. Moreover, the test-retest differences for ADRIMO show less of a recognisable spatial structure than seen with the other approaches, suggesting that more of what is removed is noise rather than fine image detail.

Whole grey matter test-retest coefficients and per-subject mean perfusion values are shown in Figure 4.5, along with a table of statistical comparisons between methods using repeated measures ANOVA and post hoc testing with adjustment of p values to correct for multiple comparisons. These show the same pattern as the example slices: test-retest coefficients are significantly higher for ADRIMO than any other method ($p < 0.05$), with BASIL being closest and voxelwise fitting performing worst.

It should be noted that although Gaussian smoothing with a large kernel leads to a high test-retest coefficient (higher even than TGV or BASIL), this is simply due to the large bias introduced by smoothing away all spatial detail, and as shown in simulations such large kernels produce poor results. Unlike large-kernel Gaussian



(a) Perfusion estimates. Top row: example slice, bottom row: distribution of estimated perfusion in grey matter.

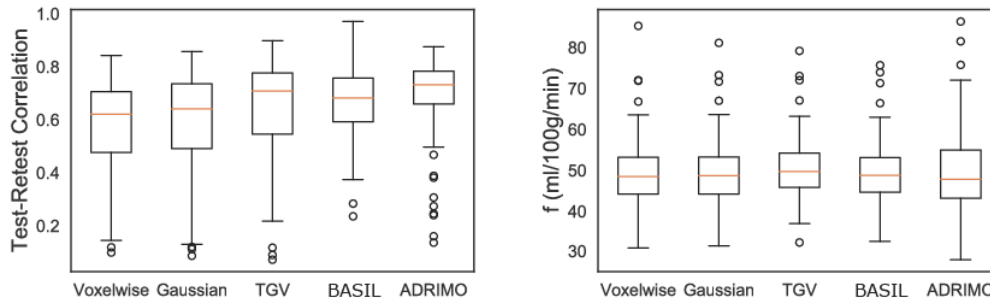


(b) Test-retest errors. Top row: absolute error shown in an example slice, bottom row: distribution of error in grey matter.

Figure 4.4: Example perfusion estimates and test-retest errors in experimental data.

smoothing, ADRIMO showed minimal bias in simulations, and ADRIMO's marked increase in test-retest correlation is compelling evidence of reduced estimate error, and hence of more precise perfusion estimates.

Figure 4.6 shows perfusion estimates for the following groups: term vs preterm, male vs female, Caesarean vs vaginal delivery. Each boxplot shows distributions of whole grey matter perfusion for each method, as well as the p value from post hoc tests for significant differences between the groups after a three-way ANOVA, controlling for covariates in each case. For example, when comparing term/preterm, we control for male/female and Caesarean/vaginal delivery. Table 4.3 summarises group statistics, including the mean perfusion and confidence interval for each group, as well as the p values for post hoc significant inter-group differ-



(a) Test-retest coefficients (left) and grey matter perfusion distributions (right) for all subjects.

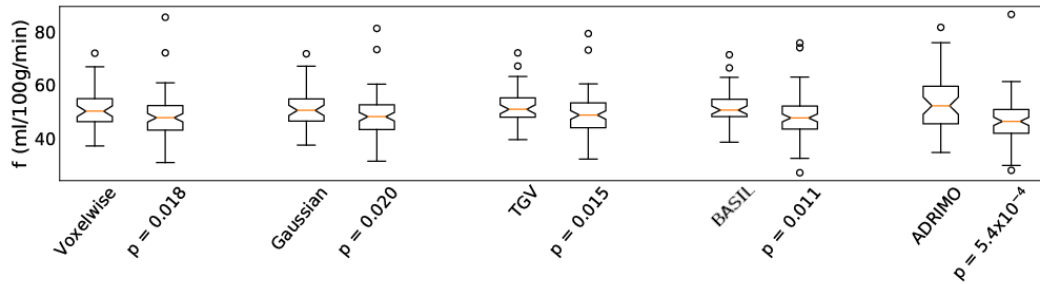
	Voxelwise	Gaussian	TGV	BASIL	ADRIMO
Test-Retest μ	0.57	0.59	0.64	0.67	0.70
Test-Retest σ	0.19	0.19	0.18	0.12	0.14
p	$< 1 \times 10^{-6}$	$< 1 \times 10^{-6}$	5.14×10^{-4}	0.0314	–

(b) Test-retest statistics for each method. μ and σ are the mean and standard deviation of the test-retest coefficients, while p is calculated by post hoc testing for significant differences compared to ADRIMO, adjusted for multiple comparisons using Holm-Bonferroni correction.

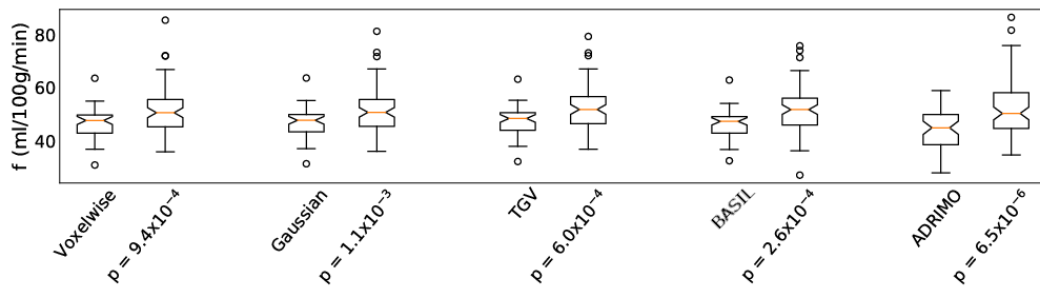
Figure 4.5: Results of test-retest experiments. Top row: distributions of test-retest correlation and estimated perfusion, bottom row: test-retest statistics and comparison against ADRIMO.

ences. We note that this is *not* a repeated measures ANOVA, unlike Figure 4.5, as we are no longer *testing* for differences between different methods. Instead, we are showing the ANOVA result an imaging study would get under each processing method.

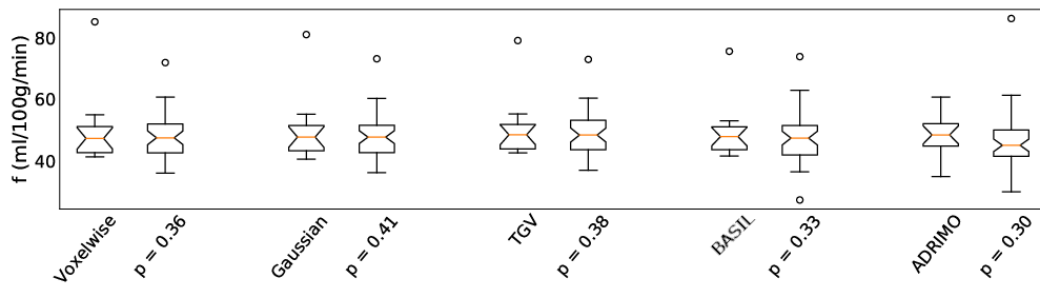
Broadly speaking, regularisation methods improve the statistical power to detect real differences – regularised results have smaller p values than the voxelwise case, apart from in Gaussian smoothing. With ADRIMO, however, the p value for real differences is an order of magnitude smaller when a difference exists. For example, we can compare male versus female subjects. Sex differences in perfusion as measured by ASL are well-established in the literature [155], and in this work all methods agree that female subjects have higher perfusion than males. However, the statistical power with which this is detected is greater when using ADRIMO; $p = 2.0 \times 10^{-5}$ for ADRIMO, whereas $p = 2.8 \times 10^{-3}$ without regularisation and $p = 7.8 \times 10^{-4}$ for BASIL.



(a) Term-born (left) vs preterm-born (right) for each method.



(b) Male (left) vs female (right) for each method.



(c) Caesarean (left) vs vaginal delivery (right) in preterm subjects for each method.

Figure 4.6: Grey matter perfusion estimates for different groups, and p value for post hoc tests for statistical difference between them, compared across methods. p values here are stated before adjustment for multiple comparisons – for adjusted versions, see Table 4.3.

Similarly, when considering term-born versus preterm-born subjects, all methods agree that term-born subjects have higher perfusion than preterm-born subjects before adjustment for multiple comparisons ($p = 1.1 \times 10^{-3}$ for ADRIMO, $p = 0.036$ for voxelwise fitting, $p = 0.022$ for BASIL and $p = 0.030$ for TGV). All methods agree that term-born subjects have higher perfusion than preterm-born subjects, but ADRIMO again shows this with greater power, similarly to the comparison of male versus female subjects. The effect of preterm birth on perfusion is much less established in the literature than sex differences in perfusion, but our finding does seem to agree with what previous work there is in this dataset and others [156, 16]. Assuming that this effect is present in the data, it is found with higher power by ADRIMO. Moreover, the increase in statistical power seems to be achieved without compromising the false positive rate: when comparing preterm-born subjects who were born via Caesarean section with those who were delivered vaginally, all methods agree there is no significant difference ($p \sim 0.3$).

4.4.2 Focal perfusion changes

4.4.2.1 Simulations

For *focal* simulations, the ground truth and fitted values are shown in the top row of Figure 4.7, and maps of absolute error by method are shown in the bottom row. In the *focal* simulations, as discussed in Section 4.3.5, we imposed several focal regions on the *normal* perfusion map: focal regions have $2 \times 2 \times 2$, $4 \times 4 \times 4$ and $8 \times 8 \times 8$ voxels, both hypoperfused (20 ml/100g/min) and hyperperfused (120 ml/100g/min). These regions are indicated with arrows in Figure 4.7. Regularisation methods again produce smoother images with little sign of overall bias, and reduced error variance relative to voxelwise fitting in each case, as in the original simulations. ADRIMO does not smooth away the focal regions, which remain visible in Figure 4.7, but it does exhibit a tendency to drag these regions towards the average perfusion value, particularly in the case of the smaller regions.

This tendency to drag focal regions of hypo- or hyperperfusion towards the average perfusion value is quantified in Table 4.4. BASIL, which also uses a prior (and hence involves some degree of statistical shrinkage), is nonetheless less prone

	Voxelwise				BASIL			
	μ	σ	p	Δf	μ	σ	p	Δf
Preterm	48.4	7.80	0.0360	2.90 ± 3.21	48.2	7.73	0.0222	3.25 ± 3.07
Term	51.3	7.44			51.5	7.01		
Male	46.5	5.69	2.82×10^{-3}	4.74 ± 2.70	46.3	5.28	7.80×10^{-4}	5.06 ± 2.59
Female	51.2	8.40			51.4	8.25		
Caesarean	49.5	10.9	0.36	1.38 ± 6.34	49.2	8.21	0.33	1.50 ± 5.03
Vaginal	48.1	7.49			47.7	8.04		
	TGV				ADRIMO			
	μ	σ	p	Δf	μ	σ	p	Δf
Preterm	49.2	7.24	0.0301	2.95 ± 3.02	46.7	8.95	1.08×10^{-3}	6.58 ± 4.15
Term	52.1	7.07			53.3	10.3		
Male	47.5	5.47	1.80×10^{-3}	4.60 ± 2.54	44.2	7.31	1.95×10^{-5}	7.95 ± 3.35
Female	52.1	7.78			52.2	10.2		
Caesarean	50.2	9.05	0.38	1.21 ± 5.37	47.7	7.15	0.30	1.38 ± 4.75
Vaginal	49.0	7.27			46.4	9.42		

Table 4.3: Perfusion statistics by group, compared under different methods: preterm-born versus term-born, male versus female, Caesarean versus vaginal delivery. μ and σ are perfusion mean and standard deviation (both in ml/100g/min), p is for significant differences between groups and Δf is the perfusion difference with 95% confidence intervals (also in ml/100g/min). p values and confidence intervals are stated after Holm-Bonferroni adjustment for multiple comparisons.

to drag focal regions towards the mean perfusion condition. The difference between methods is most dramatic for extremely small focal regions, in both hypo-perfused and hyperperfused cases. For the small hypo-perfused region, for example, ADRIMO overestimates perfusion as 38.2 ml/100g/min, whereas the true value is significantly lower at 18.6 ml/100g/min. TGV and BASIL also overestimate perfusion, but to a smaller degree (25.6 ml/100g/min and 26.2 ml/100g/min respectively). A similar pattern of overestimation exists for hyperperfusion. In “large” regions, ADRIMO performs more similarly to BASIL and other methods, for example estimating 107.5 ml/100g/min on the large hyperperfusion, compared to 115.5 ml/100g/min ground truth and 113.3 ml/100g/min for BASIL.

Focal changes would be expected to be most challenging situation for ADRIMO: when there are abrupt voxel-scale changes in the data, these may be less well-modelled with ADRIMO than with TGV (which explicitly models step

	Ground Truth		Voxelwise		Gaussian		TGV		BASIL		ADRIMO	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Small Hypo (ml/100g/min)	9.6	0.77	18.6	14.9	37.9	11.9	26.2	11.5	25.6	11.7	38.2	8.0
Medium Hypo (ml/100g/min)	9.8	0.50	6.5	15.5	16.7	18.3	18.8	11.6	25.7	10.9	34.9	12.2
Large Hypo (ml/100g/min)	9.7	0.66	8.0	26.9	17.4	20.4	11.1	12.5	16.7	10.4	28.6	12.8
Small Hyper (ml/100g/min)	113.8	7.2	120.4	23.8	130.2	18.3	115.5	14.2	100.6	10.6	87.4	2.0
Medium Hyper (ml/100g/min)	116.5	6.3	142.2	26.5	166.8	25.9	131.1	22.6	131.6	18.1	98.1	9.0
Large Hyper (ml/100g/min)	115.5	7.9	116.0	29.0	135.5	27.0	120.6	23.6	113.3	19.6	107.5	11.9

Table 4.4: *Focal* simulation: mean perfusion within ROI (and standard deviation) for each method.

changes and ramp changes in intensity) or BASIL (which, like ADRIMO, is data-driven – but also has an explicit model that weights parameter smoothness at the voxel level rather than the regional level). These results show a need for caution when very small focal changes in perfusion might be present, although we note that changes on this scale would in general be difficult to reliably detect using ASL.

4.4.2.2 Experimental results

Figure 4.8 shows example axial slices from the subject with focal hypoperfusion. Although there is no way to ascertain the ground truth, and the data are too scarce for the test-retest difference to be meaningful, it is clear that ADRIMO succeeds in detecting the focal perfusion difference, similarly to other regularisation methods. The estimated perfusion maps are broadly similar, with a mean perfusion of approximately 25 ml/100g/min and all methods preserving the hypoperfused region.

4.5 Discussion

We have presented a new technique for spatial regularisation of ASL images, and shown its superior performance in typical perfusion maps using both simulations and test-retest experiments. Compared with state of the art regularisation techniques such as TGV, ADRIMO produces more accurate results in simulations and has higher test-retest repeatability. Moreover, we have shown ADRIMO significantly improves the statistical power to detect perfusion differences between groups: using data from a cohort of 130 preterm-born adolescents and age-matched controls, ADRIMO is more able to detect perfusion differences between groups where they

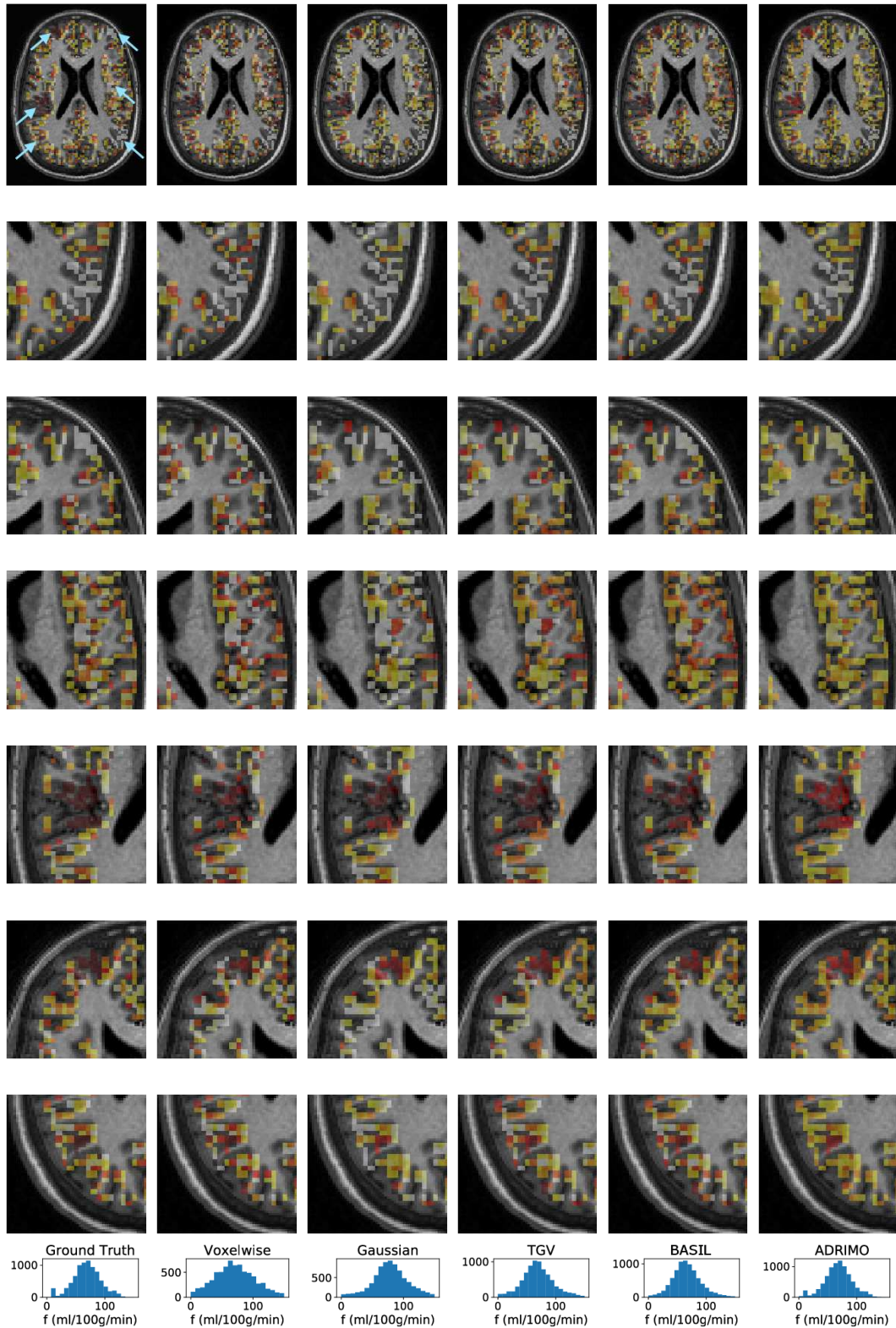


Figure 4.7: Ground truth and estimated values for the *focal* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$. Focal regions are indicated with arrows in the ground truth image and then zoomed in on the lower rows.

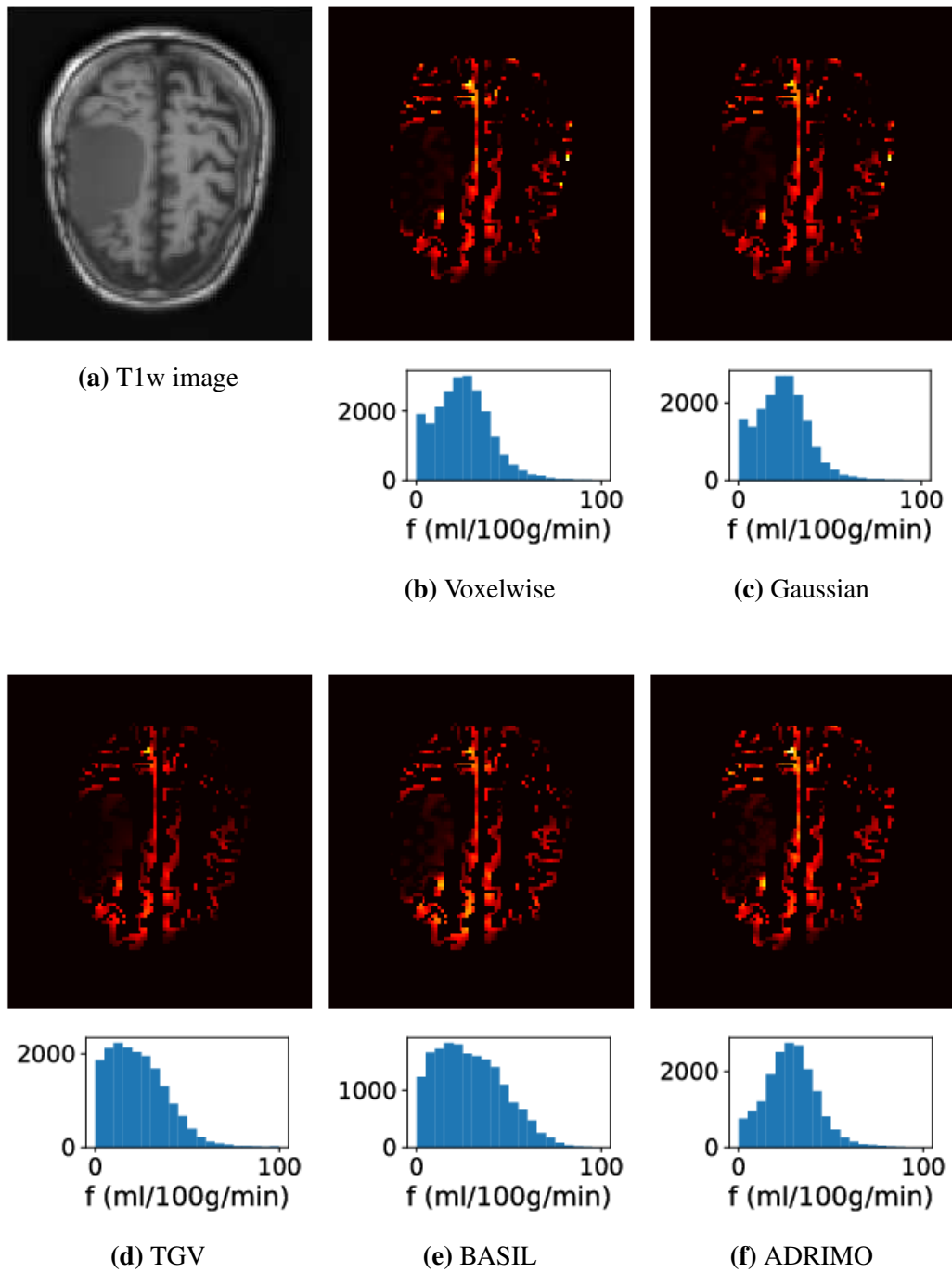


Figure 4.8: T1-weighted image and example fits from a subject with focal hypoperfusion, for all methods. The top row shows an example axial slice, while the bottom row shows the distribution of perfusion values within the mask.

are present, while giving comparable (negative) results where differences are not present. Finally, we have shown that ADRIMO copes with highly focal changes in perfusion – both in simulations with challenging small focal hypo- or hyperperfusion, and in experimental data with a highly focal arachnoid cyst.

4.5.1 Increased accuracy of perfusion estimates in representative simulations

ADRIMO offers the best fit accuracy in simulations with both *flat* and *normal* perfusion values. In the *flat* simulation (see Table 4.1), previously used in the literature for comparison of regularisation techniques with TGV [92], ADRIMO outperforms BASIL and TGV, with a mean absolute error of 0.317 ml/100g/min (ADRIMO), compared with 1.86 ml/100g/min (TGV) and 1.37 ml/100g/min (BASIL). This clear improvement in performance stems from the underlying statistical model: ADRIMO is more able to detect these non-varying perfusion values because they are well modelled by Gaussians in a region. In the less spatially correlated *normal* simulation (see Table 4.2), ADRIMO outperforms BASIL and TGV, with a mean absolute error of 1.30 ml/100g/min (ADRIMO) compared with 1.63 ml/100g/min (BASIL) and 2.02 ml/100g/min (TGV). This simulation again shows significant performance improvements for ADRIMO. This is likely because the reference methods explicitly weight perfusion estimates towards voxel-scale (rather than region-scale) similarity, which is not present in this simulation. Together, these simulations show ADRIMO is better able to handle spatial similarity and also better able to model its comparative absence, which is reflected in the improved estimate accuracy in each case.

4.5.2 Significant experimental evidence of improved accuracy in typical perfusion maps

Test-retest correlations support the idea that ADRIMO offers improved fit accuracy; assuming no significant increase in bias (and none is predicted by simulations) then more accurate fits should lead to increased test-retest correlation. ADRIMO fitting results in greater and more consistent test-retest correlation than other regularisation

techniques, improving on voxelwise fitting by 22.8%, TGV by 9.4% and BASIL by 4.5%, with $p < 0.05$ for all. This corresponds to smaller test-retest differences: in Figure 4.4, example slices show how ADRIMO gives visibly smaller test-retest errors over the whole image. ADRIMO also leaves less spatial information in the error map, as seen by the more constant intensity of the test-retest difference across the slice: more of what is smoothed is noise, rather than fine-scale information.

The improvement in fit accuracy has a striking effect when the values from different methods are used to test for perfusion differences between groups. ADRIMO-processed results are more statistically powerful for the cases where perfusion differences are believed to exist: in male versus female subjects and term-born versus preterm-born subjects, the p value is reduced by orders of magnitude (see Figure 4.6 and Table 4.3). This directly corresponds to statistical power, and hence needing a smaller number of subjects to detect a difference. We argue that this is achieved without increasing the false positive rate, but rather by improving the fit accuracy: as can be seen in the case of subjects born via Caesarean section versus vaginal delivery, ADRIMO-processed results agree with other methods that there is no difference between subjects. Sample size calculations illustrate how significant an improvement in statistical power can be: based on the statistics in Table 4.3, a male/female perfusion difference can be detected ($p < 0.05$) with 19 subjects using ADRIMO and 24 subjects using the next-best method (BASIL). Similarly, for the smaller preterm-born/term-born difference, ADRIMO requires 40 subjects and BASIL requires 58 subjects on average.

4.5.3 Sustained performance in the presence of focal perfusion changes

Simulations show that ADRIMO copes tolerably with preserving the focal hypo- and hyperintensities: the focal regions – although smoothed to a greater degree than by other methods – remain clear in the image, as shown in Figure 4.7. Very small focal changes on the scale of a 1–2 voxel radius are smoothed towards the region average, but just about remain distinguishable from their surroundings. This represents an extremely challenging example: focal changes with a larger spatial extent

are more realistically detectable with ASL (small focal changes would be difficult to differentiate from artifact in ASL), and these spatially larger changes are preserved with less smoothing and are clearly distinguishable from the baseline perfusion for both hyperperfusion and hypoperfusion. Focal changes of interest in ASL studies, for example in the study of dementia, would typically be comparable to these large regions if not even larger [150, 151], and hence ADRIMO would tolerate them well based on these simulations. Perhaps more importantly, experimental evidence for focal hypo- and hyperperfusion shows ADRIMO performs similarly to TGV and BASIL regularisation, preserving the focal difference while smoothing the image. This confirms that, even in the presence of challenging and highly focal perfusion changes, ADRIMO is able to preserve spatial detail, although this is an area of weakness compared to other methods.

4.5.4 Limitations and future work

One key limitation of our work, so far, is that we do not model partial volume effects within the signal model, instead relying on a grey matter threshold to minimise them. In practice this is a common approach, especially for single-delay ASL, and should not significantly limit our method's applicability [23]. Moreover, we do model partial volume effects within our simulations, which still show an improvement in performance from our method, without significant deterioration from this unmodelled factor. Nevertheless, in future work we hope to implement partial volume modelling within our method, which might work by incorporating volume fractions into the forward model, either directly [91] or by exploiting spatial correlation to simultaneously improve the resulting partial volume and perfusion estimates [75]. Indeed, it should be possible to directly implement ADRIMO within the spatial regularisation framework that enables simultaneous partial volume estimation in BASIL [99]: in the Gaussian process's covariance matrix, the parameter variance should be included as a per-region parameter to be inferred, rather than modelled by a noninformative hyperprior. This would involve substantial work on tractable implementation, however, and we argue it is out of scope for the work as presented here.

In the future, we also intend to explore how statistical models with adaptive complexity might further improve the performance of ADRIMO in regions with heterogeneous perfusion patterns, as in our *focal* simulations in this work. For example, using a mixture of Gaussians and automatic relevance detection [74], one could explicitly model regions with focal hypoperfusion, hence obtaining higher estimate accuracy. Another promising avenue of research would be to incorporate a Gaussian process or other more fine-scale smoothing prior into the model, potentially combining the benefits of ADRIMO with existing regularisation techniques. (As discussed above, this be achieved through integrating ADRIMO assumptions within the prior structure used in BASIL). Finally, it might be advisable to model spatially-varying noise, allowing some degree of data-driven noisemap estimation in a similar fashion to popular diffusion imaging denoising methods [153], or even incorporating noisemap estimates based on coil sensitivity measurements [74].

4.5.5 Conclusion

ADRIMO is a novel Bayesian approach to spatial regularisation of ASL images, exploiting anatomical information from structural images to improve the quality of perfusion estimates in single inflow time data. Entirely data-driven, it eliminates the need for tuning parameters as are used in most existing methods. It outperforms other methods in simulations (as measured by absolute error) and experiments (as measured by test-retest coefficients), and can be seen in simulations and experiments to preserve regions with focal perfusion changes. This work shows that ADRIMO significantly improves the quality of perfusion estimation using ASL, and increase the statistical power of ASL-based experiments. This could enable more reliable detection of perfusion differences using fewer subjects, and would be invaluable for ASL in research and the clinic.

Chapter 5

Deep convolutional filtering for spatio-temporal denoising and artifact removal

This chapter presents work on the problem of artifacts in ASL images, showing how doing artifact filtering and denoising jointly may yield superior results to doing these as separate processing steps. This chapter is based on the original publication in which this work first appeared [1].

5.1 Introduction

As discussed at length in Chapter 2, ASL suffers from the twin problems of having low signal-to-noise ratio (SNR) and being prone to artifacts from patient motion, RF coil instability and several other sources. Typically, to address these problems, denoising and artifact filtering are used. Denoising uses statistical properties of the ASL signal to improve the effective SNR, for example by modelling the signal using total variation priors, a wavelet basis, or anatomy-derived spatial correlation [92, 3]. Denoising methods tend to assume Gaussian noise, and are not usually robust to non-Gaussian artifacts, for example due to patient motion or hardware instability. Artifact filtering methods, conversely, remove or down-weight parts of the ASL signal that have severe artifacts, allowing subsequent processing to assume Gaussian noise [103, 100, 107, 106].

Denoising and artifact removal are usually considered in isolation from one another, but are overlapping problems: noise is often neither strictly Gaussian nor spatially homogeneous, and artifact filtering often results in the rejection of entire image volumes when only a fraction of the image is thoroughly corrupted. In this work, we develop a deep convolutional neural network (CNN) [110] for simultaneous denoising and artifact filtering, making full use of the available data. Inspired by cutting-edge developments in computer vision, we create a novel deep learning architecture that can relate noisy, artifact-corrupted ASL images to the true underlying perfusion. This architecture uses joint convolutional filtering [157] in order to efficiently extract spatio-temporal information from the ASL signal, allowing our method to distinguish artifact from noise. We present results from our method in ASL data from 15 healthy volunteers, showing that our method improves on the state of the art for both artifact filtering and denoising, increasing the peak SNR by up to 50%. These promising initial results show that deep convolutional joint filtering holds great promise for ASL processing, and suggest our approach might be useful both for improving individual subjects' images and for increasing the statistical power of neuroimaging studies.

5.2 Methods

5.2.1 Arterial spin labelling

We use the Buxton kinetic model (described in Chapter 2) to relate the underlying perfusion parameters to the measured ASL signal. In this work, we use an ASL dataset from 35 healthy 19-year-old volunteers (F/M=17/18). Images were acquired on a 3T Phillips Achieva with 2D EPI pseudo-continuous ASL using 30 control-label pairs, PLD=1800ms+41ms/slice, $\tau = 1650\text{ms}$, $3 \times 3 \times 5\text{mm}$. We also acquired M_0 images and 3D T_1 -weighted volumes at 1mm isotropic resolution. All ASL data were motion corrected via rigid registration before being used – note, however, that motion correction often does not fully compensate for subject motion, and this is one of the artifacts that should ideally be filtered when estimating the true perfusion.

5.2.2 Denoising and artifact removal in natural images and medical images

Denoising and artifact removal are long-standing problems in computer vision, dating back to the early days of image processing [143]. For decades, the main approaches to these problems used explicit image prior models to regularise the inverse problem of estimating an original image after corruption. In recent years these problems have, like many other problems in computer vision, seen radical progress by using discriminative approaches based on convolutional neural networks (CNNs) [110]. Here we discuss these recent advances, focusing on CNN architecture and results in natural images.

Much work on image denoising focuses on the case of additive white Gaussian noise (AWGN), due to its being a commonplace property of noise in many practical settings [143]. Early work on AWGN denoising focused on formulating explicit image priors, subsequently using these to regularise the inverse problem of estimating the original image. These were based on sparsity in some image basis (e.g. wavelets, total variation), non-local self-similarity, or some model of the image content variation (e.g. Markov Random Fields, Gaussian processes) [143]. These have all been used extensively in medical image denoising as well as natural image denoising, and typically they can easily be adapted for the problem of *blind* image denoising, where noise magnitude is unknown ahead of time. This is done by marginalising over (or estimating) the appropriate noise parameter, which in the case of AWGN is simply the noise variance.

Attempts to learn discriminative models of image content offered the promise of richer, data-driven image models. As machine learning in general moved towards neural networks [110], it was natural to use these to learn such a discriminative model. The work of Zhang *et al* was a significant advance in this direction, using a CNN to learn blind AWGN denoising and demonstrating significant improvements relative to the state of the art at the time [143]. This work used an architecture stacking several identically sized 3×3 convolutional layers, batch normalisation after each convolution, and residual linear unit (ReLU) nonlinearity layers after

each batch normalisation. This architectural design is fairly “standard”, to the extent standards exist in this new and rapidly developing field, and was heavily inspired by the work of He *et al* [158]. Another design feature was the use of a residual connection; this had already shown notable success in training deep networks, and was a natural fit to the problem of AWGN denoising, where the noise itself may be seen as an additive residual on the uncorrupted image.

Subsequent work has achieved its largest improvements primarily through careful processing or augmentation of the training data (hence improving generalisation), with the benefits from subsequent architectural changes being relatively small [159]. One notable architectural development was that of using shallow, wider networks to achieve similar or even superior results – seemingly in disagreement with CNN design principles for other problems such as image classification [160]. Another novel subsequent direction of research was to use related methods in different applications – for instance burst denoising and video denoising, where there are multiple noisy images to be used [161]. In theory this could allow a higher quality reconstruction, but carries challenges related to any non-stationarity between the images.

Related to this question of stationarity between images in burst denoising and video denoising, artifact removal (as opposed to denoising) has commonly been associated with anomaly identification [106, 103]. Rather than modelling the noise process as stationary, and the noisy image as a sample from the noise process applied to the ground truth image, artifacts are occasional dramatic changes in apparent image content. Artifacts often have no clearly defined model, and hence are particularly amenable to machine learning approaches [110]. Excitingly, CNN denoising approaches like those discussed above have shown a promising capability to learn robustness to artifacts, without the need for a separate step for anomaly identification [143]. In general, the ability to learn invariance to nuisance features that are hard to explicitly define is a recognised advantage of rich, automatically learned discriminative models such as CNNs [110]. In our approach, outlined below, we take advantage of this property. We use a CNN primarily designed for denoising,

similar to [143] and [160], ensuring that it learns robustness to relevant artifacts in our problem.

As set out above, CNNs are a well-established means of processing images for a variety of tasks including denoising [110, 143]. Application of deep learning based denoising methods to MR imaging (particularly ASL) remains in its infancy, but has already achieved several promising results. Here, we focus on pixel-to-pixel denoising networks: those which take input images and produce a higher quality output image. We also focus our review on methods that existed prior to the work in this chapter was conducted. Subsequent developments are discussed in Section 5.4.

Outside ASL, much of the initial work on applying deep learning to medical imaging focused on classic computer vision tasks such as segmentation or registration [162]. Aside from this, several authors used deep learning approaches for denoising and artifact removal in low-dose CT reconstruction – with architectures ranging from an encoder-decoder approach [163] to generative adversarial networks (GANs) [164]. There is as yet no clear consensus on which approach is best, with comparisons being made (if at all) on different data under different conditions. Arguably more relevant to work in ASL, Benou *et al* used an ensemble of autoencoders for denoising of DCE-MRI, with evidence of improvement over traditional methods [165]. Their approach was quite different to the rest of the literature, which has mostly focused on CNNs. Benou *et al* focused on fully-connected (rather than convolutional) autoencoders, and trained a classifier to route a given kinetic curve to the most appropriate autoencoder in an ensemble trained across their parameter space. Finally, another exciting area of application was the application of CNNs to image quality transfer in diffusion MRI by Tanno *et al* [166], where variational dropout was used in a shallow convolutional architecture for simultaneous super-resolution and quality enhancement.

Within ASL, Gong *et al* did early work on a deep learning method to fuse information across ASL and other contrasts for patch-based denoising, showing proof of concept results with dramatic SNR improvements [167], but their technique was never developed further or published in full. Kim *et al* applied deep

learning based denoising to Hadamard-encoded ASL, showing significant improvements in image quality although not comparing with any other method [96]. Kim *et al* used a relatively shallow two-pathway network in an attempt to encourage the extraction of multi-scale information – one pathway focused on higher-level global information, one pathway focused on low-level information. It is perhaps surprising that this was necessary, given the success of single-pathway architectures for denoising in the computer vision literature [143]. Part of the reason may be the great benefit to ASL processing of *spatio-temporal* models: it is easier to distinguish artifact from perfusion when one is aware of whether a particular part of the signal was transient or permanent [92]. Processing a sequence of several 3D images rapidly becomes computationally expensive, however, and although there is work on recurrent-convolutional networks in the literature, they are mostly for 2D images and are technically challenging even then [161].

5.2.3 Deep convolutional joint filtering for ASL

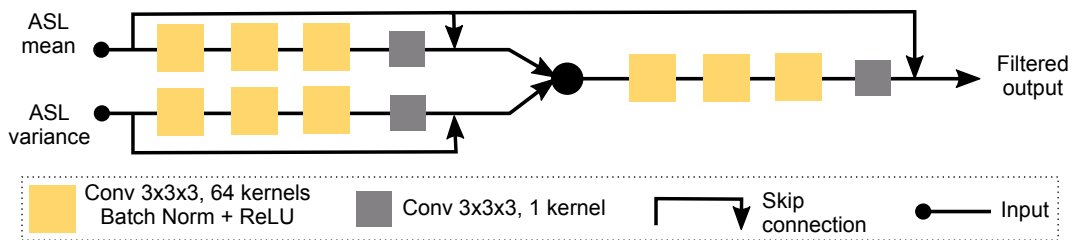


Figure 5.1: Architecture diagram for our deep convolutional joint filter network. Skip connections at the end of first and second stages improve convergence. In the case of the mean-only filter, the ASL variance branch is not used.

In order to provide some temporal information in our network, we decided on a simpler and computationally cheaper approach: we explicitly feed in temporal variance information so the network has two inputs: the mean ASL signal, and the ASL signal variance over time. The naive approach to using spatio-temporal information in this way would be to feed the signal mean and variance maps into the same CNN, similarly to colour channels in a natural image. However, initial testing showed poor performance, as the network was unable to translate voxel-level features into meaningful cross-channel information, and produced denoised images very similar to those from using only the mean ASL signal image.

Our approach to solve this problem is a *joint convolutional filter* architecture, inspired by image processing approaches to integrating RGB cameras and depth sensors [157]. In such sensor fusion problems, simply providing inputs as different channels of an image is also ineffective, likely due to the dramatically different domain inhabited by RGB images and depthmaps [157]. We reasoned that there may be a similar discrepancy between temporal variation and ASL signal mean images, which might benefit from a similar approach. In our architecture, information is extracted and processed in parallel from the mean ASL image and the ASL temporal variance. These images are combined at a later stage in the network, with several more layers used to extract meaningful features from their combination – again, inspired by the literature on multi-modality image fusion, in which similar architectures are established and successful [157].

We settled on a relatively shallow number of layers (the network is essentially six layers deep) as performance seemed to plateau, which is actually consistent with state of the art results in natural image denoising networks [160]. This is also beneficial for computational cost, particularly during training; indeed, in other image fusion techniques, shallow networks are also used [157]. Skip connections in the parallel stages improve network convergence and robustness [158], as well as transferring global information in the learning process [157]. We also use a skip connection from the mean ASL image to the output for the same purpose. We avoid using a skip connection from the temporal variance, as this data is much further from the desired output than the ASL mean is, and initial experiments showed negative effects on convergence and performance.

To train our models, we first identified artifact-corrupted volumes using the filtering method of Tan *et al* [103]. We generated gold standard high-quality perfusion maps by removing these outlying volumes and using all of the remaining data to fit perfusion according to literature recommendations [23]. These gold standard images were used as ground truth. The inputs to our network were derived by taking 10 random volumes from the ASL series, including 1–3 artifact-corrupted volumes per subject to train the network to correct for artifacts in addition to denoising. This

selection of volumes was repeated 10 times for each subject in the training set. The loss function used was mean squared error within the brain mask.

We implemented our CNNs in Keras, using the Adam optimiser with learning rate 0.01 with 20 subjects for training and 15 subjects for validation. As previously discussed, each training subject had 10 different random subsamples of its original data used, yielding a total training pool of 200 images derived from the original 20. The learning rate was set empirically by choosing a rate low enough to converge but high enough to converge in reasonable time; learning rate optimisation or scheduling are generally less beneficial for the Adam optimiser compared to Stochastic Gradient Descent (SGD), and tend not to yield drastic performance improvements [110]. To avoid overfitting and improve generalisation, we augmented with random translations sampled uniformly up to 5mm in each dimension. We also augmented input images with Gaussian noise, magnitude approximately 1% of the ASL noise as estimated from gold standard data. Augmentations were applied with random fresh values at every iteration of training, meaning each epoch used slightly different training images [110]. We note that, although it is typical to use datasets with thousands of images in deep learning, here we used 3D images, which means there was significantly more information per image. We also, as discussed, used a relatively shallow network. We trained to convergence (within 1000 epochs), which took 12 hours using an NVIDIA K80 graphics card. We subsequently used the model with lowest validation set loss.

5.2.4 Comparison to pre-existing methods and validation

For both denoising and artifact filtering, we compare our method with a state of the art spatial regularisation technique using total generalised variation (TGV), which has been shown to produce reliable and accurate denoising with built-in artifact rejection via robust statistics [92]. For reference, we also compare against voxelwise fitting with no spatial regularisation – this remains a very common way to process ASL data, and acts as a representative baseline.

We evaluate our method, using the full spatio-temporal information as discussed in Section 5.2.3, and we also evaluate a simpler CNN architecture using

only spatial information (see Figure 5.1), to show the benefit of the joint filter. We evaluate the performance by examining filtered images for residual artifacts, and by producing maps of absolute error relative to the gold standard perfusion map. These are shown in Figure 5.3. We show slices from subject 7, where there is extreme artifact; and subject 4, with less severe artifact. Subsequently we perform quantitative validation by calculating the peak SNR (PSNR) for each denoising method, again calculated relative to the gold standard and expressed in decibels, dB^1 . We subsequently statistically compare the performance of different methods on a per subject basis by bootstrapping volumes for subjects in the validation set, i.e. running inference with the trained network on different subsets of the original full per-subject data, in a similar fashion to how the training data was subsampled in Section 5.2.3. For each subject, we run paired nonparametric tests (Wilcoxon signed-rank) comparing the different methods.

Often, outlier filtering is performed as a separate step prior to denoising; so we also compare against TGV and voxelwise fitting with explicit outlier rejection via z-score filtering [103]. Our validation dataset was chosen such that each subject

¹ $PSNR = 20\log_{10}(S_{max}/RMSE)$, where S_{max} is the maximum ASL signal over all voxels and $RMSE$ is the root mean square error

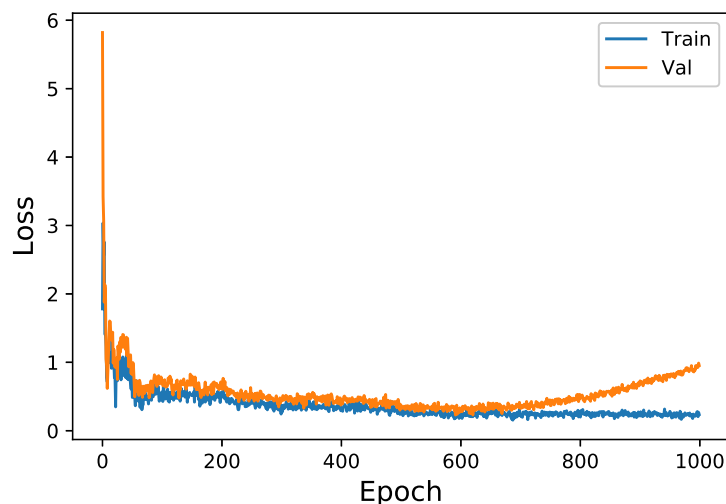


Figure 5.2: Train and validation loss versus epoch for the deep convolutional filter. Overfitting begins around epoch 6000, and the lowest val loss is obtained at epoch 6084.

contains one or more artifact volumes, as identified by z-score filtering on the full dataset. We remove artifact volumes for the reference methods, showing how they would perform in conjunction with z-score filtering. For our joint filter, however, we do not remove the volumes in this comparison, as the purpose of the joint filter was to use the non-artifact information within partially-corrupted images. As before, we evaluate example images from subjects 4 and 7, and then we present quantitative validation over all subjects using PSNR calculations.

5.3 Results

5.3.1 Example images

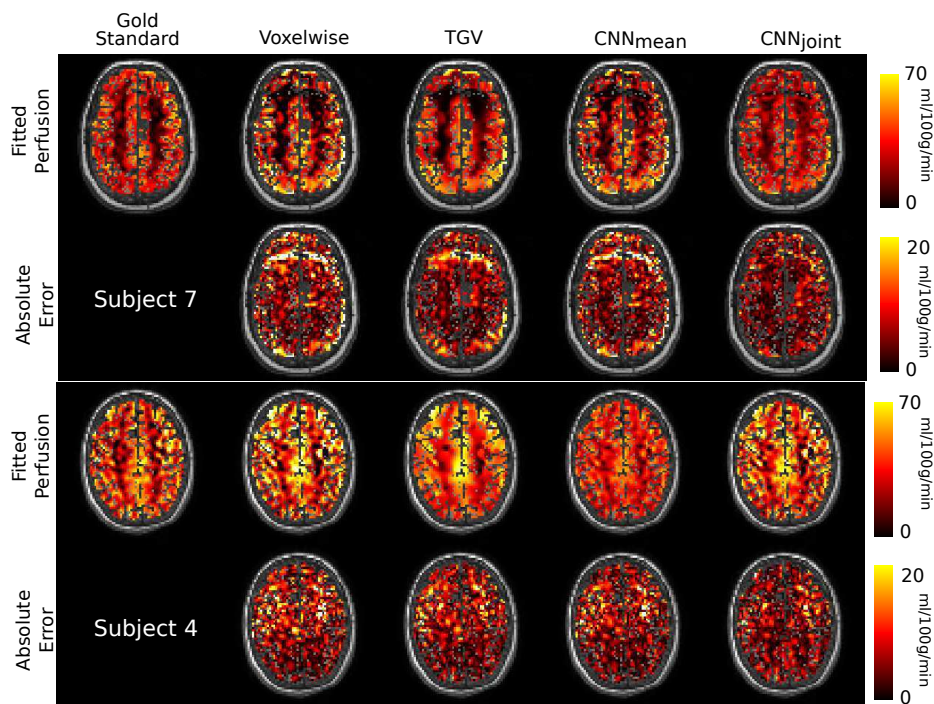


Figure 5.3: Example perfusion estimates and absolute errors for several denoising methods, with no separate artifact filtering step, shown in subjects 4 and 7. Our two CNN filter approaches are shown on the right (CNN_{mean} and CNN_{joint}).

Figure 5.3 shows example axial slices from subjects 4 and 7, as well as maps of absolute error. For subject 7, there is a strong hyperintense ring artifact near the front of the brain. Similarly, for subject 4, several artifacts present as extreme intensity changes, mostly near the edges of the brain. Voxelwise fitting shows these most plainly in both subjects, as the fitting has no implicit artifact removal. TGV results

in heavily smoothed images, removing some of the artifact seen in the voxelwise images, but also losing detail in the image. CNN mean-only smooths away even more spatial detail than TGV, and shows a similar pattern of artifact to voxelwise fitting. However, the joint filter produces a significantly less artifact-prone image, as well as improved denoising.

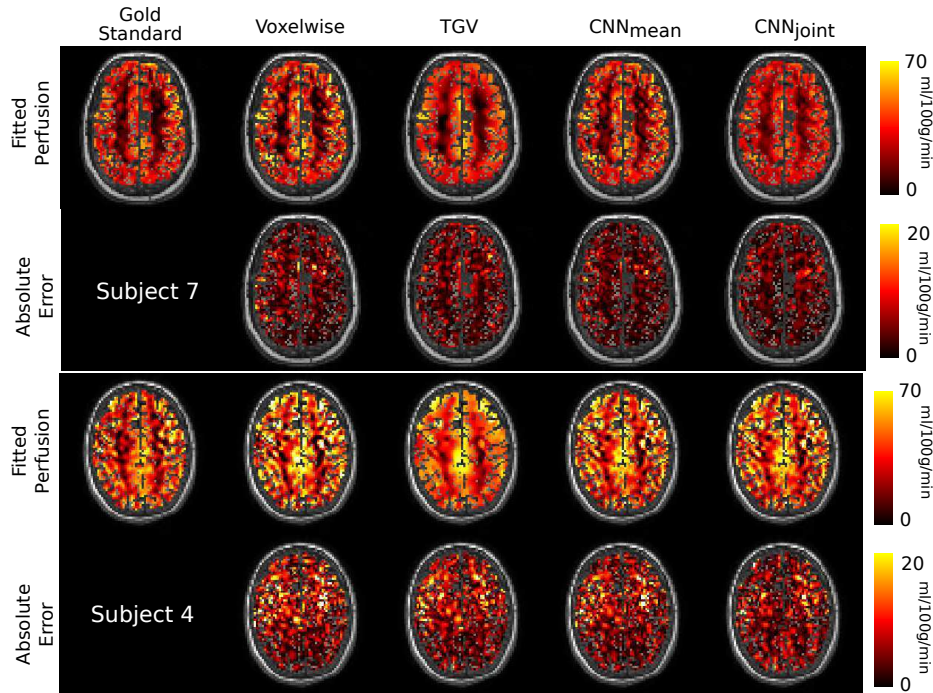


Figure 5.4: Example perfusion estimates and absolute errors for several denoising methods, with separate artifact filtering performed before denoising, shown in subjects 4 and 7. Our two CNN filter approaches are shown on the right (CNN_{mean} and CNN_{joint}).

Figure 5.4 shows example axial slices for subjects 4 and 7 again, this time preprocessed with artifact removal as a separate step. This is a more realistic comparison – certainly for voxelwise smoothing, which has no built-in artifact rejection. Here, the mean-only CNN performs closer to the joint CNN, although the joint CNN continues to produce visibly better denoising. Moreover, the remaining artifacts have been better removed by the joint CNN, despite the joint CNN being the only method to have no explicit artifact rejection before fitting.

5.3.2 Quantitative evaluation via PSNR

Figure 5.5 shows the PSNR for each subject and method, when there is no explicit artifact filtering. Because there are relatively few ASL images, and there is large inter-subject variability in the artifacts and global perfusion, PSNR varies greatly across subjects. To assist comparison between methods, Figure 5.5 shows change in PSNR relative to voxelwise fitting for each subject. The joint CNN produces the best result in all subjects except subject 15, where TGV performs marginally better. The average per-subject improvement of the joint CNN over TGV is 1.25dB ($p < 0.01$), although for some subjects with extreme artifact the improvement can be as high as 6dB. Crucially, while the mean-only CNN is often worse than TGV, the joint filter outperforms it significantly ($p < 0.05$ for each) in 11/15 subjects, marginally in 3/15 subjects, and is marginally worse (0.72dB, $p = 0.07$) only in subject 15. Although joint filtering does not always produce significantly better results than the mean-only CNN, it is significantly better than the mean-only CNN ($p < 0.05$) in nine subjects. Figure 5.7 supports these results by showing confidence intervals: differences are significant where there is small overlap of confidence intervals.

Figure 5.6 shows the PSNR for each subject and each method, when there is an explicit artifact filtering step as described in Section 5.2.4. Joint filtering again performs the best, always better than or comparable to the runner-up. Joint filtering is significantly ($p < 0.05$) better than TGV in 13/15 subjects, and marginally superior in subjects 9 and 13. The average improvement over TGV per subject is 1.64dB ($p < 0.001$). Moreover, the second-best method is typically the mean-only CNN – with this explicit filtering step, even the mean-only CNN consistently outperforms TGV filtering ($p < 0.05$ for 12/15 subjects). This is reasonable: when there is less artifact influence, temporal information is less important and the problem becomes one of spatial regularisation, where CNNs excel. Additionally, TGV often performs worse than voxelwise fitting – over-regularising the fits based on the scarce data remaining after filtering. Figure 5.7 again shows confidence intervals for these PSNR results, demonstrating the significant improvement of the CNN methods in most subjects.

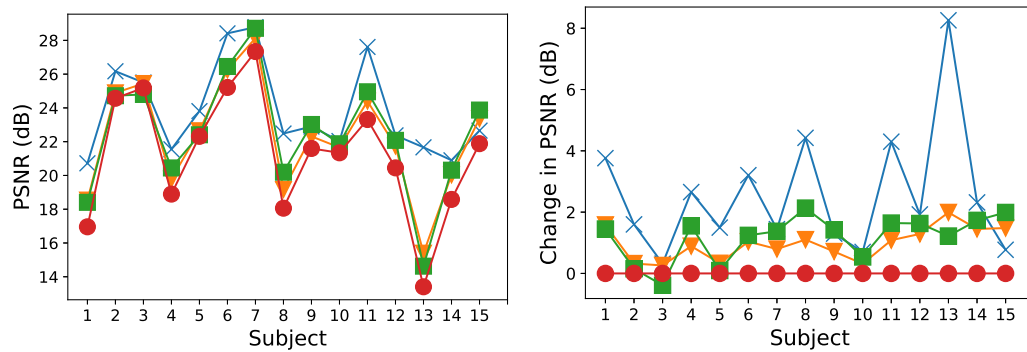


Figure 5.5: Left: PSNR for each subject and filtering method, with no separate artifact filtering step. Right: change in PSNR, expressed relative to simple voxelwise fitting, shown to make comparison of methods easier. Key: \times joint CNN, ∇ mean-only CNN, \blacksquare TGV, \bullet voxelwise.

5.4 Discussion and conclusions

As demonstrated by the visible improvements in image quality (Figure 5.3) and the significant increase in PSNR (Figure 5.5), our joint filtering approach performs better than state of the art for denoising in the presence of artifact. Of particular note is the filter's strong performance in artifact removal – in subject 7, for example, a prominent edge artifact is removed completely from the output image without any appreciable drop in denoising. The superior performance of the joint filter, compared with a mean-only CNN, shows the value in providing temporal variance information when processing artifact-prone data.

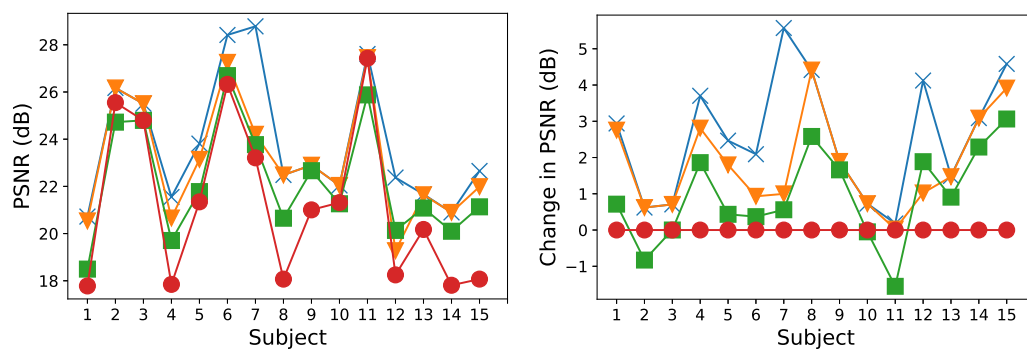


Figure 5.6: Left: PSNR for each subject and filtering method, with a separate artifact filtering step before denoising as described in Section 5.2.4. Right: change in PSNR, expressed relative to simple voxelwise fitting, shown to make comparison of methods easier. Key: \times joint CNN, ∇ mean-only CNN, \blacksquare TGV, \bullet voxelwise.

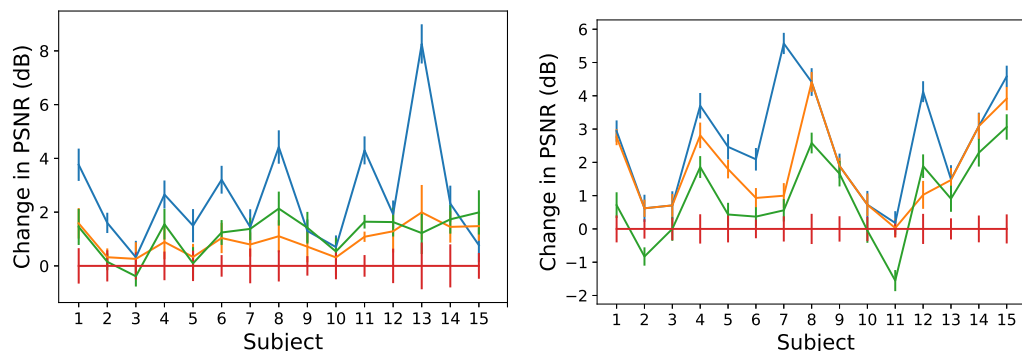


Figure 5.7: Change in PSNR for each subject and each processing method, expressed relative to voxelwise fitting in each case. These are the same plots as in Figure 5.5 (right) and Figure 5.6 (right) but now include errorbars generated by bootstrapping volumes. This helps to see where differences between methods are significant. Left: no separate artifact filtering step, right: separate preprocessing for all methods except the joint filter, as described in Section 5.2.4. Key: blue – joint CNN, orange – mean-only CNN, green – TGV, red – voxelwise.

Compared with pipelines involving separate filtering and denoising, our method again outperforms state of the art (Figures 5.4 and 5.6). By retaining parts of a corrupted volume, more information can be used in denoising, meaning the joint filter performs better than mean-only CNN filtering in most subjects. This is evidence the joint filter is able to perform better, on average, than combining a simpler CNN approach with explicit artifact filtering. Moreover, even the mean-only CNN is itself an advance on state of the art: this approach outperforms TGV in 12/15 subjects when filtering is applied separately.

Future work will involve validation across different ASL acquisitions and subject populations, leading the way for use in neuroimaging studies and the clinic. Such work would also benefit from a more thorough comparison across more pre-existing methods: denoising methods have been compared relatively thoroughly in ASL [95], but artifact filtering methods have not seen head-to-head comparison for the most part. In this work we have focused on comparison against a small number of methods in the literature, shown to perform close to state of the art (in the case of denoising) or to be in common use (in the case of artifact filtering). We argue that it is outside of the scope of this work to produce a fuller comparison, which would be a publication in itself [95], but it would be valuable work for the future of the field.

Future work might also explore alternative ways to exploit temporal information, for example through a recurrent-convolutional architecture. More importantly, any method should handle variations in ASL data such as readout and label type. Currently this requires retraining on each new dataset, but it may be possible for a single network to handle these different cases. Finally, a limitation of this work is the necessity for higher-quality data (e.g. more ASL volumes) in a subset of subjects for training; so we wish to explore how cross-validation derived loss functions might ameliorate this. To this end, transfer learning may be helpful to reduce the computational cost of retraining in several cross-validation folds.

Since the original publication of this work, deep learning for medical imaging has flourished, with thousands of new publications covering many diverse applications [162, 168]. Our work was one of the first applications of deep learning to ASL processing, with preceding work discussed in Section 5.2.2. Roughly contemporaneous with our work was that of Ulas *et al*, who adopted a similar approach of CNN-based residual learning for denoising, albeit using only a single network branch [97]. They also included a loss term that encouraged perfusion estimates to fall close to normal reference values – conceptually similar to a physiological prior [74], albeit without the analytical advantages of generative models. This approach showed a PSNR improvement of 1.5dB relative to simple averaging in real data, i.e. the “voxelwise” reference method in this work. This seems to be smaller than the benefit we derived from our method in this work, although the datasets are not directly comparable, so it is uncertain.

The innovative joint approach to denoising and artifact filtering presented here has the potential to substantially increase the quality of ASL images, even salvaging datasets that were previously considered unusable. By fusing temporal variance information with spatial information in a novel network architecture, our deep convolutional joint filter method outperforms state of the art in both denoising and filtering. Our method is applicable to any ASL data, subject to training requirements, and could even be used in other imaging modalities. Consequently, deep convolutional joint filtering presents an exciting future direction for medical image

processing in noisy and artifact-prone modalities, and may eventually be used to improve the statistical power of neuroimaging studies.

Chapter 6

Conclusions

In this final chapter we summarise our contributions, examine subsequent developments, and discuss future research directions. This thesis has presented three strands of original work within ASL imaging: optimising acquisitions with experimental design (Chapter 3), regularising parameter estimation in a forward model using anatomical information (Chapter 4), and spatio-temporal regularisation for simultaneous filtering of noise and artifact (Chapter 5). These may all be seen within the context of an “intelligent imaging” paradigm: optimised acquisitions tailor the acquisition to the patient, making the measurement process “intelligent”; ADRIMO uses the patient’s individual anatomical information to generate a shrinkage prior which improves parameter estimates; and spatio-temporal regularisation using a convolutional neural network is an example of using state-of-the-art pattern recognition methods to provide more informed filtering.

6.1 Experimental design

In Chapter 3, we presented an original approach to Bayesian experimental design for ASL acquisitions. The objective of this work is to make maximally efficient use of scanner time, and experimental results show it can significantly improve the accuracy of perfusion parameter estimates in a given duration. The primary novelty of our approach lies in its use of global design optimisation, and its use of a direct constraint on scanner time rather than number of measurements. Ours was also among the first such work on simultaneous optimisation of inversion times and

label durations, and was among the first such work on optimal design in pseudo-continuous ASL.

Among subsequent work on this topic, the most directly relevant is that of Woods *et al* [131]. Their work contained the first publication of optimal design in pseudo-continuous ASL and, citing our work, also followed a truly global design optimisation. They explored the usage of flat priors, and presented an ASL-specific approximation that dramatically speeds up design optimisation with negligible effect on optimisation quality. They also confirmed improvements in perfusion parameter estimation from their method, both in simulations and ASL experiments, which is consistent with our results. Their work provides further compelling evidence for the value of experimental design techniques, and its computational improvements make it more practical to use than our approach, albeit at the cost of generality.

Bladt *et al* build on ASL optimal design work in a less directly comparable way. They examine the extended problem of optimal design for ASL with tissue T_1 as a parameter to be estimated, rather than using a fixed value [169]. They use Bayesian design optimisation inspired by Woods *et al* and our work, optimising inversion times alone and jointly optimising inversion times and label durations. Their emphasis is not on the improvements from design optimisation, however; they focus on the trade-off between estimating T_1 and perfusion parameters. They show that the estimation of T_1 substantially impairs estimation of other parameters, and make a convincing argument that it is impractical in shorter acquisitions. They also explore how population-derived T_1 values can reduce bias when using a fixed T_1 value, and compare this with both T_1 estimation and other fixed T_1 values.

There are other subsequent developments that are less directly similar, due to not being within ASL, but are relevant nonetheless. Gómez *et al* use optimal design to create a bespoke acquisition for rapid estimation of several parameters simultaneous with angiography [170]. We attempted something related in our own subsequent work, in which we showed initial results from using our optimal design framework in joint T_1 and T_2 relaxometry [4]. However, the work of Gómez *et*

al is considerably more advanced in its acquisition methods and validation. Their approach is inspired by magnetic resonance fingerprinting (MRF), in which acquisition parameters vary constantly across the sequence and lead to a transient signal that can be used for parameter estimation. However, rather than a pseudorandom variation of acquisition parameters, they use a Bayesian optimal design approach inspired by our own to improve time efficiency.

Finally, we note the work of Lahiri *et al*, which adapts experimental design for MRF in ASL, comparing it with experimental design for a more conventional multi-TI ASL sequence [171]. MRF remains a cutting-edge technique, and this is even more the case for its usage in ASL. However, in a similar fashion to Gómez *et al*, Lahiri *et al* use a transient acquisition as in MRF, but with Bayesian optimal design instead of a pseudorandom acquisition sequence. Their preliminary results indicate this may lead to significant improvements in parameter accuracy compared to MRF without optimisation or “conventional” optimised ASL.

Optimal design for ASL acquisitions has several natural areas for future work. The most crucial problem is that of reliably gathering an informative prior for use in the optimisation process: if the prior is unrepresentative, optimisation can harm the acquisition; if the prior is too vague, optimisation offers limited benefits. As discussed in Chapter 3, there are two natural ways of doing this. One way would be to acquire a rapid low resolution pre-scan to calibrate the parameter priors [136]. The other would be to move towards a closed loop of scan/estimate/optimize/... although this requires a more specialist hardware setup [125].

Another important area for future work is to investigate the performance of optimised designs in various pathologies, demographics, and acquisitions. This may best be done in several stages rather than in a small number of large studies, as best practices in ASL acquisitions are still under active improvement, especially for multi-TI ASL [23]. Moreover, advanced changes in acquisition can interact with the details of optimal design; for example, the use of Hadamard-encoded ASL [62], velocity-sensitive ASL [172], or ASL MRF [82] could all substantially interact with the design process.

Eventually, we might hope to see optimal design become an automated part of scanning procedure in a variety of imaging contrasts. This is a long term prospect, as it would require the aforementioned validation across a large variety of settings. Indeed, we cannot yet be certain that the benefits from optimal design will be large enough to justify widespread use. A significant challenge when varying acquisition parameters between subjects is that their scans are no longer directly comparable. Despite its status as a quantitative imaging methodology, ASL has many ways in which it can produce biased estimates of the underlying perfusion parameters [23]. It is an open question whether this problem is severe enough to prevent widespread use of per-subject optimal design. Even if this problem is severe, it is also possible that post-processing could help to reduce its severity [107, 103, 106]. In the meantime, perhaps optimal design may best be used for designing protocols on a per-study basis, somewhat similar to how it is often used in the statistics literature [113]. Nonetheless, optimal design work both by ourselves and other authors has shown that there is room to improve acquisition parameters in ASL acquisitions. This can certainly be used to inform future acquisition design, and we hope that future research will allow it to be used more widely.

6.2 Anatomy-driven modelling for spatial regularisation

In Chapter 4, we present our original approach for improving spatial regularisation in ASL through a shrinkage prior. Generative Bayesian models have seen marked and sustained success in ASL processing, even compared to their widespread use in medical image processing generally [74]. This is partly due to a well-established culture of use and open-source tools, but also due to their perfect suitability for ASL: methods which would be less tractable with larger numbers of data points remain tractable at the resolutions used in ASL, and meaningful priors provide valuable regularisation in a challenging imaging modality. Our work is primarily an attempt to strengthen the value of the prior in Bayesian methods for ASL analysis, making spatial regularisation more effective while remaining entirely data-driven. We

derive the method, then explore its performance in synthetic and real world experiments. Experiments suggest our method may have a benefit not only in individual image quality, but also in group studies of perfusion.

Much of the subsequent work in this area has focused on using various advanced methods for ASL denoising. This is compelling work, and we discuss it below, but it has a different focus to our approach, which was intended to avoid the need for tuning parameters and was inspired more by previous Bayesian inference methods in ASL and other imaging modalities [74, 99]. This may be partially due to the limitations of our method, which as initially presented did not incorporate partial volume correction or local spatial similarity. We discuss this in more detail below, when talking about future work.

The most direct application of our work in the literature, to date, is our group's usage of it in placental diffusion imaging [173]. This is somewhat similar to the work of Orton *et al* in the liver [148], but uses our implementation and a more similar prior formulation to that of Chapter 4. The use of our spatial regularisation led to dramatic improvements in image quality in the placental imaging work, and highlights the advantage of not needing to set tuning parameters for analysis.

Our work has also been cited by Ulas *et al*, in their work on denoising ASL using deep learning [97]. They argue that their incorporation of the ASL kinetic model into the loss function may provide more robust results. This is loosely connected to our method, and the other Bayesian approaches that inspired it, inasmuch as they are also based on the ASL kinetic model. However, the method of Ulas *et al* is otherwise very different from our method and its predecessors: we attempted to remove the need for tuning parameters, whereas Ulas *et al* introduce a neural network, which contains thousands of tuning parameters, learned from example data. We discuss this in more detail, along with several related developments, in Section 6.3.

Another subsequent development in ASL denoising, although further from our work, has been the further usage of the total generalised variation (TGV) regularisation developed by Spann *et al* [174]. We used TGV as comparison denoising

method in our work, but in their original publication, Spann *et al* also applied it to the problem of denoising higher-than-usual spatial resolution ASL images to reduce partial volume effects [92]. In their more recent work, they used similar denoising in conjunction with a novel undersampled acquisition, showing its capabilities to improve perfusion estimates substantially.

Building on our work in Chapter 4, a clear direction for future research is to reconcile our high-level priors with the full wealth of models and techniques used in ASL processing. In particular, it should be possible to simultaneously have anatomy-driven large-scale similarities as modelled by ADRIMO side-by-side with data-driven local smoothing and inference of auxiliary parameters such as partial volume and multiple compartments, all supported by a Bayesian generative model. Incorporating shrinkage into the Gaussian process prior approach explored by Groves *et al* [99] would solve this, as well as addressing other simplifications in our work concerning partial volume effects and errormap estimation. We discuss this in more detail in Chapter 4, including how it might be implemented through a change to the covariance matrix used in the Gaussian process approach. A full reconciliation of these methods remains elusive, but we believe it is achievable. Handling partial volume correction would be particularly important for practical use, and ASL post-processing methods have often been focused on this step [75, 91, 94, 89, 90].

In ASL denoising and spatial regularisation literature, there is a fairly large divide between Bayesian methods such as BASIL [74] and our work, which use easily interpretable generative models and do not require manual tuning of parameters; and methods inspired more by the literature on natural image denoising. The latter has increasingly turned to neural networks, and we discuss these more in Section 6.3. However, despite excitement in research about such methods, it remains unclear whether they will offer a *practical* benefit in ASL processing, due to the need for training data, challenges in robustness across different datasets, and lack of interpretability [168]. This lack of clarity is worsened by the lack of a standardised, side-by-side comparison of similar methods. It would be valuable for future

work to provide this, possibly through establishing an independent leaderboard on held out data, similar to those used in many computer vision tasks [175]. This could also allow the comparison of methods under different conditions, for example when similar training data is provided versus when there is no training data for a particular scanner and sequence.

6.3 Deep learning for joint denoising and artifact filtering

In Chapter 5, we set out a joint approach to artifact filtering and denoising, with promising results in real data. Inspired by state of the art results in natural image denoising, we use a convolutional neural network (CNN) that uses the mean ASL difference image and its temporal variance. Ours was among the first works on using CNNs in ASL, and this was a source of novelty in its own right. Our approach primarily differs from its predecessors and contemporaries in its use of temporal information, and an explicit consideration of artifacts. Several authors have subsequently investigated similar approaches, demonstrating the interest in this area.

As discussed in Chapter 5, work performed by Ulas *et al* took a different approach to ours, attempting to directly incorporate the ASL kinetic model in the network's loss function [97]. Their network also does not use temporal information as an input, instead using only the average difference image. This achieved significant improvements in parameter estimates and image quality, relative to least squares regression, although was not directly compared with other denoising methods. We speculate that incorporation of kinetics might be even more valuable in a multi-TI ASL acquisition, similar to its value in pre-existing Bayesian methods for denoising [99]. However, this was not tested. The work of Ulas *et al* provides further confirmation of the promise in translating CNN methods from natural image denoising, although as in our work, issues remain concerning the need for complex training and potentially poor transfer across pathologies, sequences and scanners.

Subsequent to our work and that of Ulas *et al*, Xie *et al* used a two-stream network architecture for ASL denoising [176]. Similarly to Ulas *et al*, they used mean

CBF as the input rather than using temporal information like in our method. In their method, the separate streams were to incorporate local and large-scale inter-voxel relationships, rather than to model spatial and temporal correlations as in our work. They demonstrated a substantial improvement compared to several other network architectures, although did not compare to a traditional non-CNN approach. Xie *et al* used a larger number of subjects than ourselves or Ulas *et al*, with a train/val/test split of 200/20/60 subjects respectively. This provides more confidence in the results, and may also be a reason why the more complex architecture performed better – because it had more data on which to train. This provides yet more evidence for the potential benefits of these approaches, although, similar to our work and that of Ulas *et al* and others, the method suffers a significant disadvantage: it needs a complex training procedure before use, may not transfer well across different pathologies, sequences and scanners.

Most recently, Gong *et al* have presented a CNN approach that attempts to overcome the problems of complex training procedures and poor transfer across different datasets [177]. Inspired by recent developments in the computer vision literature, they formulate denoising as a so-called deep prior problem, in which the network can be initialised with random weights and used as a prior during parameter estimation. This effectively allows “training” on a single subject’s data at the time of parameter inference. This approach also allows for the use of a structural image to further improve the prior, although they do not study the effect of this in detail. Strikingly, Gong *et al* achieve significant improvements over comparison methods in three subjects, without the need for a separate training step. They also show that their method can help with reconstructing undersampled data, although they note that artifact removal remains a challenge, perhaps because it is difficult for the network to filter these without conventional training on artifact-free examples.

Finally, we note the use of a CNN-based method in a slightly different ASL application: since our work, Li *et al* have subsequently attempted to use a deep learning based method for simultaneous denoising and super-resolution, achieving promising early results [178]. They use a two-stage architecture with multiple

losses, preserving high-frequency detail during the super-resolution stage. This is compared against several other CNN-based approaches for denoising and/or super-resolution, with their method generally showing comparable or superior performance in 30 subjects. It is possible that performance may be over-estimated in this work, due to cross-validation across *slices* rather than subjects. According to the authors' description, different slices from a given subject may be seen in both training and validation for a given fold, which might lead to overfitting. This method also suffers from the usual problem of requiring a complex training procedure and potentially struggling to transfer across datasets. Nevertheless, it demonstrates the potential of CNN-based methods in an innovative application, and shows the interest in this area in the ASL literature.

Future work, for all CNN-based methods, will need substantially more validation, as well as examining the problem of transferability across scanners, patients, and acquisitions. As discussed in Section 6.2, there has not been a systematic comparison of different denoising methods that includes many of the recently developed methods. Moreover, despite the existence of several artifact filtering methods, there has been no thorough comparison of them on the same dataset, nor an examination of which methods work best on which types of ASL data. This would be a natural next step in the problem of artifact filtering, and could prove to be an invaluable contribution. Again, we emphasise the success of independent leaderboards with held-out data in the computer vision literature [175]. Such a leaderboard could allow for the fairest comparison of different methods, prevent overfitting, and make it easier for groups without large datasets to conduct research in this field. However, it would be a substantial task, and while there is progress in such leaderboards in medical imaging more broadly (for example the challenges at MICCAI and elsewhere [175]), we do not expect to see this in the immediate future for ASL research.

CNNs have become widespread in recent years for a wide variety of computer vision problems, and justifiably so. However, they suffer from several shortcomings relative to more traditional methods, and these are particularly harmful for medical imaging [168]. CNNs require a complex training procedure and training data, and

may not generalise to different domains. Many techniques exist to improve robustness across different domains and reduce – or even remove – the need for training data. This has even been explored in ASL data specifically, in the aforementioned work of Gong *et al* [177]; however, the problem remains open, both for research at the level of fundamentals and applications.

Another weakness of CNNs, compared to popular Bayesian approaches for ASL post-processing [74], is that CNNs do not robustly model uncertainty, and hence can fail without adequate warning or visible sign of a problem [110]. Several researchers have begun to explore this topic, but it largely remains unsolved. Relatedly, another deficiency of CNN-based approaches is that they are not easily interpretable [168, 110]. A generative model has a clear and understandable logic for how MR images are used to infer parameter distributions. This allows us to reason about unusual cases, understand poor performance, and know the limits of our model. CNNs learn relationships from data in a way that is mathematically simple (stacked convolutions and non-linearities) but can be conceptually opaque. This is another area of active research in the machine learning literature, and it will be valuable future work to address these problems – to the extent possible – and use those solutions for medical imaging applications.

6.4 Conclusion

ASL is becoming increasingly widespread, with great excitement about its usage for research, diagnostics, prognostics and drug evaluation – primarily in the brain, but increasingly in other organs. Through various improvements in acquisition, processing and analysis, one can significantly improve the quality of ASL images and their practical usefulness. This thesis has given an overview of these methods within the context of so-called intelligent imaging, and has also presented original work (including theory, simulations and experiments) in detail. We hope that some of these techniques may be of use in ASL's exciting future, and we conclude on a note of optimistic anticipation: blood flow imaging is only beginning to show its vast capabilities, and seems poised to do more and more in the future.

Appendix A

Analytic solutions for the optimal design problem of Xie *et al*

In this appendix we show how the problem of finding a weighted ensemble of local designs, as posed by Xie *et al* and described in Chapter 2, can be optimised mostly analytically, without requiring the use of iterative exchange.

Xie *et al* use a bounded line search inspired by coordinate exchange, which is commonly used in classic experimental design problems. In a coordinate exchange algorithm, at every step, the list of design point coordinates is updated one by one, replacing each point in turn with another candidate point. These candidate points are typically taken from a pre-evaluated set of optimal design points which form the support of the optimal design. This approach is more useful for local optimal design, in which the support is often relatively sparse. In robust optimal design through a nonlinear function, as is the case for ASL, the support often has many more points. Consequently, instead of attempting to find the support in advance, the replacement point is found via a line search between adjacent points on either side.

The time consuming step in this process is that it must be repeated many times, for different settings of the parameters, f and θ . As discussed in Chapter 2, this assumes that the ensemble of local optimal, weighted by the prior, approximates the global optimum. However, we note that it is possible to exploit the sparsity of the design support to cast this as a linear mapping: the optimal TIs' distribution is a mixture of two normal distributions: the Δt prior, and a translated version of the Δt

prior centred at $\Delta t + \tau$. We show this below.

Assuming we seek the distribution of local optima as an approximation to the global optimum, we note that an analytic solution for local optimisation would suffice for analytic solution of the entire global optimisation. The local utility function is $u(t) = \frac{1}{N} \det(\Sigma_i^N H(t_i))$, with $H_{j,k} = \frac{d\Delta M}{d\theta_j} \frac{d\Delta M}{d\theta_k}$ and $\theta = (f, \Delta t)$.

$$\text{Hence } u(t) \propto \Sigma_i \left(\frac{d\Delta M}{df} \right)^2 \Sigma_i \left(\frac{d\Delta M}{d\Delta t} \right)^2 - \left\{ \Sigma_i \frac{d\Delta M}{df} \frac{d\Delta M}{d\Delta t} \right\}^2.$$

If we define $A : A_i = \frac{d\Delta M}{df}$ and $B : B_i = \frac{d\Delta M}{d\Delta t}$, we can express this a metric over vectors: $u(t) \propto (A \cdot A)(B \cdot B) - (A \cdot B)^2$. This captures the trade-off in optimal design: one seeks to maximise sensitivity to each parameter ($A \cdot A$ and $B \cdot B$), but measurements should also disentangle the effect of different parameters on the signal ($A \cdot B$).

By inspection of the sensitivity functions, $\frac{d\Delta M}{d\theta}$, the locally optimal design must have a support of only two points: $t_{opt} \in \{\Delta t, \Delta t + \tau\}$. If a design point were placed anywhere other than one of these points, then shifting it to $\Delta t + \tau$ would necessarily improve $(A \cdot A)(B \cdot B)$ or shifting it to Δt would necessarily improve $(A \cdot B)^2$.

Xie *et al* observed that the locally optimal designs used only these two points, but neglected to exploit this sparsity. When the priors are taken to be Gaussian, as in this case, this yields a bimodal distribution of design points – essentially a mixture-of-Gaussians, responsibilities weighted by the relative sensitivity function inner-product. This is easiest to evaluate numerically. Consequently, the approach of Xie *et al* can be evaluated without the use of an iterative exchange algorithm, and almost entirely analytically. Unfortunately, as previously discussed, the weighted ensemble of local optima is a poor approximation to the global optimum, which motivated our subsequent work in Chapter 3.

Appendix B

Software Packages

There are several pieces of software underpinning the work presented in this thesis. Here, I briefly recap these, as well as explaining my own contributions. Open source releases can be found at <https://github.com/karnival> and <https://cmiclab.cs.ucl.ac.uk/CMIC>.

NiftyFit A package for nonlinear, multi-modality model fitting. My contributions lie in the ASL modelling and statistical modelling of shrinkage priors for other modalities. NiftyFit is described in detail in its associated publication [15].

NiftyOpt Bayesian optimal experiment design for a constrained duration, using quadrature and global optimisation. This was the basis of the optimal design work in Chapter 3.

ASL-TGV An open-source implementation of the total generalised variation filtering described by Spann *et al.* This was used as a baseline for comparison in Chapters 4 and 5, prior to Spann *et al* sharing their implementation. At the time of writing, ASL-TGV is the only publicly available implementation.

ADRIMO This implementation of ADRIMO underpinned the work in Chapter 4, and was also the basis of shrinkage priors and fitting for other modalities in several publications [3, 14, 12].

ASL-DL Based on the Keras framework, ASL-DL was used to implement the convolutional neural network architecture of Chapter 5, in addition to its training and validation.

Appendix C

Acronyms and abbreviations

- ABV – Arterial Blood Volume
- ADC – Apparent Diffusion Coefficient
- ADRIMO – Anatomy-Driven Modelling
- ASL – Arterial Spin Labelling
- ATT – Arterial Transit Time
- BASIL – Bayesian Inference for Arterial Spin Labelling MRI
- BOLD – Blood-Oxygenation Level Dependent
- CASL – Continuous ASL
- CBF – Cerebral Blood Flow
- CNN – Convolutional Neural Network
- CT – Computer Tomography
- DCE – Dynamic Contrast Enhanced
- DECIDE – Diffusion-Relaxation Combined Imaging for Detailed Placental Evaluation
- DL – Deep Learning
- DSC – Dynamic Susceptibility Contrast
- EMF – Electromagnetic Field, Electromotive Force
- EPI – Echo-Planar Imaging
- FDG – Fluorodeoxyglucose
- FLIRT – FMRIB's Linear Image Registration Tool
- FSL – FMRIB Software Library

- GF – Growth Factor
- GM – Grey Matter
- GPP – Gaussian Process Prior
- GRASE – Gradient Spin Echo
- GRE – Gradient Echo
- ICA – Independent Component Analysis
- IVIM – Intravoxel Incoherent Motion
- KL – Kullback-Leibler
- KW – Kruskal-Wallis
- MAE – Mean Absolute Error
- MPRAGE – Magnetisation-Prepared Rapid Acquisition GRE
- MR – Magnetic Resonance
- MRI – Magnetic Resonance Imaging
- MVN – Multivariate Normal
- NIG – Normal Inverse Gamma
- NIRS – Near-Infrared Spectroscopy
- PASL – Pulsed ASL
- PCASL – Pseudo-Continuous ASL
- PD – Proton Density
- PDF – Probability Density Function
- PET – Positron Emission Tomography
- PLD – Post-Label Delay
- PSNR – Peak SNR
- PV – Partial Volume
- Q2TIPS – QUIPSS-II with Thin-Slice TI_1 Periodic Saturation
- QUASAR – Quantitative STAR Labelling of Arterial Regions
- QUIPSS – Quantitative Imaging of Perfusion Using a Single Subtraction
- RF – Radio Frequency
- RGB – Red/Green/Blue
- RMSE – Root Mean Square Error

- ROI – Region of Interest
- SAR – Specific Absorption Rate
- SD – Standard Deviation
- SNR – Signal to Noise Ratio
- SPECT – Single Photon Emission CT
- SPM – Statistical Parametric Mapping
- TE – Echo Time
- TGV – Total Generalised Variation
- TI – Inflow Time or Inversion Time
- TR – Repetition Time
- VB – Variational Bayes

List of Figures

- 2.1 A high-level view of ASL measuring perfusion: a perfusion-weighted image (right) is the difference between labelled and control images (left and centre respectively). Figure adapted from [33] . 20
- 2.2 Overview of ASL timing, and how it differs between PASL and PCASL. First a label is applied before the imaging region, typically at the neck for neuroimaging. Subsequently, after waiting to allow labelled blood to reach the organ, imaging occurs. In PASL (bottom) the label duration is not necessarily known, although it can be fixed with a quantitative imaging of perfusion with single subtraction (QUIPSS-II) pulse. PASL is usually specified in terms of TI, whereas PCASL is usually specified in terms of PLD. Figure adapted from [63] 29
- 2.3 The black curve shows the ASL kinetic model (PASL), with f and Δt set to typical values of 40 ml/100g/min and 0.8 s respectively. The shaded region shows how the curve varies as f and Δt vary over common values (20–60 ml/100g/min for f and 0.5–1.0 s for Δt). Blood arrives at $t = \Delta t$, the first turning point of the curve; then collects, increasing the measured signal, until the second turning point, $t = \Delta t + \tau$, when there is no more inflowing blood and the signal begins to decay. Red crosses show example noisy measurements from this forward model: the task in model fitting is to infer f and Δt from these noisy measurements. 49

- 2.4 MR signal decay due to T_1 relaxation (red curve) was used as an argument for weighting the number of averages (blue bars) at different TIs differently, with quadratically more averages at the lower intensity regions. The figure is reproduced from Kramme *et al* [128]. 52
- 3.1 Scanner estimates of SAR, as a percentage of the allowable limit, versus label duration for PCASL. Energy deposition is linear with τ , matching published theoretical models. 65
- 3.2 The black curve shows the PASL kinetic model, without noise, simulated at $f = 40$ ml/100g/min and $\Delta t = 0.8$ s. Labelled blood first arrives at $t = \Delta t$, and the signal magnitude peaks at $t = \Delta t + \tau$. The shaded region shows the variation in the curve within one standard deviation of the parameter means ($20 < f < 40$ ml/100g/min, $0.5 < \Delta t < 1.1$ s). \times shows example noisy measurements, which can be used to invert the model to estimate f and Δt 67
- 3.3 PASL optimised design (top) and reference design (bottom), defined by their TIs. Shifting TIs to more informative values near Δt and $\Delta t + \tau$ makes the acquisition more efficient, and avoiding longer TIs allows more measurements to be made in the same scan time. . . 69
- 3.4 PCASL reference design (top), TI-optimised design (middle) and TI+LD-optimised design (bottom). \circ marks TI and \times marks LD. Note that we are specifying TIs rather than PLDs, for consistency with the PASL results. 70
- 3.5 TI against LD for the the PCASL TI+LD optimised design. The line shows identity (TI=LD), highlighting how the gap between LD and TI increases at longer TIs. Long LDs are generally favoured (near as possible to the identity line) at lower TI values, but at higher values, this is deprioritised. Somewhat similarly to the PASL optimised design, measurements cluster roughly around Δt and $\Delta t + \tau$ 71

- 3.6 Boxplots of whole-GM test-retest coefficients for f and Δt in the PASL experiments. Boxes show lower and upper quartiles of the data, and the line within each box shows the sample median. Crosses mark individual subjects' coefficients. When there is no increase in bias, a higher coefficient is preferable, as it reflects a lower variance between test and retest. 76
- 3.7 Example grey matter maps of f (left) and Δt (right) for Subject 4 from the PASL experiments. Absolute test-retest difference maps are also shown for each parameter. The same axial slice is shown throughout for comparison. Estimated maps are similar for both parameters, indicating no great change in bias, whereas test-retest difference maps are smaller for f and similar for Δt , indicating lower variability and hence improved performance. 77
- 3.8 Boxplots of whole-GM test-retest coefficients for f and Δt from the PCASL experiments. Boxes show lower and upper quartiles of the data, and the line within each box shows the sample median. Crosses mark individual subjects' coefficients. When there is no increase in bias, a higher coefficient is preferable, as it reflects a lower variance between test and retest. 77
- 3.9 Example grey matter maps of f (left) and Δt (right) from the PCASL experiments. Absolute test-retest error maps are also shown. The same axial slice is shown for each parameter, for comparison. Estimated parameter maps are broadly similar, indicating no change in bias, whereas the maps of absolute test-retest error are clearly smaller in the optimised acquisitions, and lowest of all in the joint optimisation. 78

- 3.10 Simulation results for abnormal perfusion values. The top row shows results for hypoperfusion and the bottom row shows results for hyperperfusion. Bars show the 95% confidence intervals for f and Δt residuals at each f value, calculated over 10,000 noisy simulations, with reference design on the left (red) and optimised design on the right (blue) in each case. In each case, \times marks the mean residual. 80
- 3.11 Simulation results for abnormal transit time. The top row shows results for prolonged Δt and the bottom row shows results for shortened Δt . Bars show the 95% confidence intervals for f and Δt residuals at each Δt value, calculated over 10,000 noisy simulations, with reference design on the left (red) and optimised design on the right (blue) in each case. In each case, \times marks the mean residual. 81
- 3.12 Optimised designs, Xie method versus truly global method, flat prior $\Delta t \sim U(0.5, 1.8)$ versus normal prior $\Delta t \sim \mathcal{N}(\mu = 1.15, \sigma = 0.325)$ such that two standard deviations of the normal prior reflect the entire range of the flat prior. 82
- 4.1 Examples of the experimental data, shown in an axial slice. From left to right: T1-weighted structural image, lobar parcellation, proton density image, perfusion-weighted image. The top row shows data from a term-born subject, while the bottom row shows data from a preterm-born subject. 100
- 4.2 Ground truth, fitted values and errors for the *flat* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$ 102
- 4.3 Ground truth, fitted values and errors for the *normal* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$ 103
- 4.4 Example perfusion estimates and test-retest errors in experimental data. 105

- 4.5 Results of test-retest experiments. Top row: distributions of test-retest correlation and estimated perfusion, bottom row: test-retest statistics and comparison against ADRIMO. 106
- 4.6 Grey matter perfusion estimates for different groups, and p value for post hoc tests for statistical difference between them, compared across methods. p values here are stated before adjustment for multiple comparisons – for adjusted versions, see Table 4.3. 107
- 4.7 Ground truth and estimated values for the *focal* simulation. Gaussian smoothing is shown for $\sigma = 2\text{mm}$. Focal regions are indicated with arrows in the ground truth image and then zoomed in on the lower rows. 111
- 4.8 T1-weighted image and example fits from a subject with focal hypoperfusion, for all methods. The top row shows an example axial slice, while the bottom row shows the distribution of perfusion values within the mask. 112
- 5.1 Architecture diagram for our deep convolutional joint filter network. Skip connections at the end of first and second stages improve convergence. In the case of the mean-only filter, the ASL variance branch is not used. 122
- 5.2 Train and validation loss versus epoch for the deep convolutional filter. Overfitting begins around epoch 6000, and the lowest val loss is obtained at epoch 6084. 125
- 5.3 Example perfusion estimates and absolute errors for several denoising methods, with no separate artifact filtering step, shown in subjects 4 and 7. Our two CNN filter approaches are shown on the right (CNN_{mean} and CNN_{joint}). 126
- 5.4 Example perfusion estimates and absolute errors for several denoising methods, with separate artifact filtering performed before denoising, shown in subjects 4 and 7. Our two CNN filter approaches are shown on the right (CNN_{mean} and CNN_{joint}). 127

5.5 Left: PSNR for each subject and filtering method, with no separate artifact filtering step. Right: change in PSNR, expressed relative to simple voxelwise fitting, shown to make comparison of methods easier. Key: \times joint CNN, \blacktriangledown mean-only CNN, \blacksquare TGV, \bullet voxelwise. 129

5.6 Left: PSNR for each subject and filtering method, with a separate artifact filtering step before denoising as described in Section 5.2.4. Right: change in PSNR, expressed relative to simple voxelwise fitting, shown to make comparison of methods easier. Key: \times joint CNN, \blacktriangledown mean-only CNN, \blacksquare TGV, \bullet voxelwise. 129

5.7 Change in PSNR for each subject and each processing method, expressed relative to voxelwise fitting in each case. These are the same plots as in Figure 5.5 (right) and Figure 5.6 (right) but now include errorbars generated by bootstrapping volumes. This helps to see where differences between methods are significant. Left: no separate artifact filtering step, right: separate preprocessing for all methods except the joint filter, as described in Section 5.2.4. Key: blue – joint CNN, orange – mean-only CNN, green – TGV, red – voxelwise. 130

List of Tables

3.1	Synthetic results for PASL, evaluated within one and two standard deviations of the prior, and over the whole space. Root mean square error (RMSE) is measured in the same units as the parameter, i.e. ml/100g/min for f and seconds for Δt . Δ RMSE and Δ CoV are the percentage changes between reference and optimised acquisitions, where positive change indicate better performance in the optimised design.	73
3.2	Synthetic results for PCASL, evaluated within one and two standard deviations of the prior, and over the whole space. Root mean square error (RMSE) is measured in the same units as the parameter, i.e. ml/100g/min for f and seconds for Δt . Δ RMSE and Δ CoV are the percentage changes between reference and optimised acquisitions, where positive change indicate better performance in the optimised design. $p < 0.05$ for all.	74
3.3	Simulation results for optimised vs reference designs under various combinations of abnormal parameter values. Positive change in RMSE and CoV indicate better performance in the optimised design.	79
3.4	PASL inversion times (ms) for reference and optimised acquisitions. Label duration is 0.8s for all measurements.	87
3.5	PASL relaxation times (ms) for reference and optimised acquisitions.	87
3.6	PCASL inversion times (ms) for reference, TIs-optimised and joint-optimised acquisitions.	88

3.7	PCASL label durations (ms) for reference, TIs-optimised and joint-optimised acquisitions.	88
3.8	PCASL relaxation times (ms) for reference, TIs-optimised and joint-optimised acquisitions.	88
4.1	<i>Flat</i> simulation: bias (as a percentage of the mean perfusion), coefficient of variation ($\text{CoV} = \sigma/\mu$) and mean absolute error (MAE) for each method.	101
4.2	<i>Normal</i> simulation: bias (as a percentage of the mean perfusion), coefficient of variation ($\text{CoV} = \sigma/\mu$) and mean absolute error (MAE) for each method.	103
4.3	Perfusion statistics by group, compared under different methods: preterm-born versus term-born, male versus female, Caesarean versus vaginal delivery. μ and σ are perfusion mean and standard deviation (both in ml/100g/min), p is for significant differences between groups and Δf is the perfusion difference with 95% confidence intervals (also in ml/100g/min). p values and confidence intervals are stated after Holm-Bonferroni adjustment for multiple comparisons. .	109
4.4	<i>Focal</i> simulation: mean perfusion within ROI (and standard deviation) for each method.	110

Bibliography

- [1] D. Owen, A. Melbourne, Z. Eaton-Rosen, D. L. Thomas, N. Marlow, J. Rohrer, and S. Ourselin. Deep convolutional filtering for spatio-temporal denoising and artifact removal in arterial spin labelling MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 21–29. Springer, 2018.
- [2] D. Owen, A. Melbourne, Z. Eaton-Rosen, D. L. Thomas, N. Marlow, J. Rohrer, and S. Ourselin. Anatomy-driven modelling of spatial correlation for regularisation of arterial spin labelling images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 190–197. Springer, 2017.
- [3] D. Owen, A. Melbourne, D. Thomas, J. Beckmann, J. Rohrer, N. Marlow, and S. Ourselin. ADRIMO: Anatomy-DRiven MOdelling of spatial correlation to improve analysis of arterial spin labelling data. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*. ISMRM, 2017.
- [4] D. Owen, A. Melbourne, M. Sokolska, D. Thomas, J. Rohrer, and S. Ourselin. Bayesian experimental design for multi-parametric T1/T2 relaxometry and diffusion. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*. ISMRM, 2017.
- [5] D. Owen, A. Melbourne, D. Thomas, E. De Vita, J. Rohrer, and S. Ourselin. Optimisation of arterial spin labelling using Bayesian experimental design.

In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*. Springer, 2015.

- [6] A. Melbourne, R. Aughwane, M. Sokolska, D. Owen, G. Kendall, D. Flouri, A. Bainbridge, D Atkinson, J. Deprest, T. Vercauteren, et al. Separating fetal and maternal placenta circulations using multiparametric MRI. *Magnetic resonance in medicine*, 81(1):350–361, 2019.
- [7] L. Smith, A. Melbourne, D. Owen, M. J. Cardoso, C. Sudre, T. Tillin, M. Sokolska, D. Atkinson, N. Chaturvedi, S. Ourselin, A. Hughes, F. Barkhof, and R. Jäger. Cortical cerebral blood flow in aging: Effects of haematocrit, sex and ethnicity. *European Radiology*, 2019.
- [8] R. Pratt, A. Melbourne, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, J. Deprest, G. Kendall, T. Vercauteren, S. Ourselin, et al. Spatial vascular heterogeneity in the normal placenta assessed with multicompart ment placental mri. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France*. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018.
- [9] L. Smith, A. Melbourne, D. Owen, MJ. Cardoso, C. Sudre, T. Tillin, M. Sokolska, D. Atkinson, N. Chaturvedi, S. Ourselin, et al. Cortical cerebral blood flow in aging: Effects of haematocrit, sex and ethnicity. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France*. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018.
- [10] A. Melbourne, R. Pratt, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, J. Deprest, G. Kendall, T. Vercauteren, A. David, et al. Placental insufficiency investigated with multi-compartment placental MRI. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Paris, France*. Joint Annual Meeting ISMRM-ESMRMB 2018, 2018.

- [11] R. Pratt, A. Melbourne, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, S. Ourselin, et al. Novel placental evaluation using multimodal MRI. *Ultrasound in Obstetrics & Gynecology*, 50(S1):62–62, 2017.
- [12] A. Melbourne, R. Pratt, M. Sokolska, D. Owen, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. Separation of fetal and maternal circulations using multi-modal MRI. *Placenta*, 57:291, 2017.
- [13] A. Melbourne, R. Pratt, D. Owen, M. Sokolska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. DECIDE: Diffusion-rElaxation combined imaging for detailed placental evaluation. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*. ISMRM, 2017.
- [14] A. Melbourne, R. Pratt, D. Owen, M. Sokloska, A. Bainbridge, D. Atkinson, G. Kendall, J. Deprest, T. Vercauteren, A. David, et al. Placental image analysis using coupled diffusion-weighted and multi-echo T2 MRI and a multi-compartment model. In *Medical Image Computing and Computer-Assisted Intervention: Perinatal, Preterm and Paediatric Image Analysis workshop—MICCAI 2016*. MICCAI, 2016.
- [15] A. Melbourne, N. Toussaint, D. Owen, I. Simpson, T. Anthopoulos, E. De Vita, D. Atkinson, and S. Ourselin. NiftyFit: a software package for multi-parametric model-fitting of 4D magnetic resonance imaging data. *Neuroinformatics*, 14(3):319–337, 2016.
- [16] A. Melbourne, Z. Eaton-Rosen, D. Owen, M.J. Cardoso, J. Beckmann, D. Atkinson, N. Marlow, and S. Ourselin. Measuring cortical neurite-dispersion and perfusion in preterm-born adolescents using multi-modal MRI. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, pages 378–386. Springer, 2015.
- [17] K. Miller, F. Alfaro-Almagro, N. Bangerter, D.L. Thomas, E. Yacoub, J. Xu,

- A. Bartsch, S. Jbabdi, S. Sotiropoulos, and J. Andersson. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 2016.
- [18] R. Klabunde. Cardiovascular Physiology Concepts. <http://cvphysiology.com>.
- [19] A. Pappano and W. Gil Wier. 1 - overview of the circulation and blood. In *Cardiovascular Physiology (Tenth Edition)*, pages 1 – 9. Elsevier, Philadelphia, tenth edition edition, 2013.
- [20] E. Gabryś, M. Rybaczuk, and A. Kędzia. Fractal models of circulatory system. symmetrical and asymmetrical approach comparison. *Chaos, Solitons & Fractals*, 24(3):707–715, 2005.
- [21] M. Wintermark, M. Sesay, E. Barbier, K. Borbély, W. Dillon, J. Eastwood, T. Glenn, C. Grandin, S. Pedraza, J. F. Soustiel, et al. Comparative overview of brain perfusion imaging techniques. *Stroke*, 36(9):e83–e99, 2005.
- [22] J. Detre, J. Leigh, D. Williams, and A. Koretsky. Perfusion imaging. *Magnetic Resonance in Medicine*, 23(1):37–45, 1992.
- [23] D. C. Alsop, J. A. Detre, X. Golay, M. Günther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications. *Magnetic Resonance in Medicine*, 73(1):102–116, 2015.
- [24] J. Detre, J. Wang, Z. Wang, and H. Rao. Arterial spin-labeled perfusion MRI in basic and clinical neuroscience. *Current Opinion in Neurology*, 22(4):348–355, 2009.
- [25] J. A. Detre, H. Rao, D. Wang, Y. F. Chen, and Z. Wang. Applications of arterial spin labeled MRI in the brain. *Journal of Magnetic Resonance Imaging*, 35(5):1026–1037, 2012.

- [26] J. Hendrikse, E. T. Petersen, and X. Golay. Vascular disorders: insights from arterial spin labeling. *Neuroimaging Clinics*, 22(2):259–269, 2012.
- [27] F. Barkhof, N. Fox, A. Bastos-Leite, and P. Scheltens. Neuroimaging in dementia. *Neuroimaging in Dementia*, 8(1):1–278, January 2011.
- [28] D. Wolk and J. Detre. Arterial spin labeling MRI: an emerging biomarker for alzheimer’s disease and other neurodegenerative conditions. *Current Opinion in Neurology*, 25(4):421, 2012.
- [29] S. Payne. *Cerebral autoregulation: control of blood flow in the brain*. Springer, 2016.
- [30] K. Leenders, D. Perani, A. Lammertsma, J. Heather, P. Buckingham, T. Jones, M. Healy, J. M. Gibbs, R. Wise, J. Hatazawa, et al. Cerebral blood flow, blood volume and oxygen utilization: normal values and effect of age. *Brain*, 113(1):27–47, 1990.
- [31] P. W. Hales, J. Kawadler, S. Aylett, F. Kirkham, and C. Clark. Arterial spin labeling characterization of cerebral perfusion during normal maturation from late childhood into adulthood: normal reference range values and their use in clinical studies. *Journal of Cerebral Blood Flow & Metabolism*, 34(5):776–784, 2014.
- [32] R. Tootell, J. Reppas, K. Kwong, R. Malach, R. Born, T. Brady, B. Rosen, and J. Belliveau. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, 15(4):3215–3230, 1995.
- [33] C. Rorden. Chris Rorden’s Neuropsychology Lab. <http://www.mccauslandcenter.sc.edu/CRNL>.
- [34] S. Ogawa, T. M. Lee, A. Kay, and D. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

- [35] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, 1986.
- [36] D. Le Bihan and R. Turner. The capillary network: a link between IVIM and classical perfusion. *Magnetic resonance in medicine*, 27(1):171–178, 1992.
- [37] P. Pandharipande, G. Krinsky, H. Rusinek, and V. Lee. Perfusion imaging of the liver: current challenges and future goals. *Radiology*, 234(3):661–673, 2005.
- [38] S. Ichikawa, U. Motosugi, T. Ichikawa, K. Sano, H. Morisaka, and T. Araki. Intravoxel incoherent motion imaging of the kidney: alterations in diffusion and perfusion in patients with renal dysfunction. *Magnetic resonance imaging*, 31(3):414–417, 2013.
- [39] B. Rosen, J. Belliveau, J. Vevea, and T. Brady. Perfusion imaging with NMR contrast agents. *Magnetic resonance in medicine*, 14(2):249–265, 1990.
- [40] G. H. Jahng, K. L. Li, L. Ostergaard, and F. Calamante. Perfusion magnetic resonance imaging: a comprehensive update on principles and techniques. *Korean journal of radiology*, 15(5):554–577, 2014.
- [41] A. Khawaja, D. Cassidy, J. Al Shakarchi, D. McGrogan, N. Inston, and R. Jones. Revisiting the risks of MRI with Gadolinium based contrast agents: review of literature and guidelines. *Insights into imaging*, 6(5):553–558, 2015.
- [42] M. Rogosnitzky and S. Branch. Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *Biometals*, 29(3):365–376, 2016.
- [43] F. Mettler and M. Guiberteau. *Essentials of nuclear medicine imaging*. Elsevier Health Sciences, 2012.

- [44] D. Smith and A. Goel. Radiopaedia: SPECT vs PET. <https://radiopaedia.org/articles/spect-vs-pet>.
- [45] W. Moses. Fundamental limits of spatial resolution in pet. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648:S236–S240, 2011.
- [46] M. Lukies and F. Gaillard. Radiopaedia: CT perfusion in ischaemic stroke. <https://radiopaedia.org/articles/ct-perfusion-in-ischaemic-stroke>.
- [47] D. Cosgrove and N. Lassau. Imaging of perfusion using ultrasound. *European journal of nuclear medicine and molecular imaging*, 37(1):65–85, 2010.
- [48] W. Copen, P. Schaefer, and O. Wu. MR perfusion imaging in acute ischemic stroke. *Neuroimaging Clinics*, 21(2):259–283, 2011.
- [49] A. Jacobs, L. Kracht, A. Gossmann, M. Rüger, A. Thomas, A. Thiel, and K. Herholz. Imaging in neurooncology. *NeuroRx*, 2(2):333–347, 2005.
- [50] J. Folkman. Role of angiogenesis in tumor growth and metastasis. In *Seminars in oncology*, volume 29, pages 15–18. Elsevier, 2002.
- [51] A. Leeuwis, L. Smith, A. Melbourne, A. Hughes, M. Richards, N. Prins, M. Sokolska, D. Atkinson, T. Tillin, H. Jäger, et al. Cerebral blood flow and cognitive functioning in a community-based, multi-ethnic cohort: the sabre study. *Frontiers in aging neuroscience*, 10, 2018.
- [52] C. Luckhaus, M. Flüß, H. J. Wittsack, B. Grass-Kapanke, M. Jänner, R. Khalili-Amiri, W. Friedrich, T. Supprian, W. Gaebel, U. Mödder, et al. Detection of changed regional cerebral blood flow in mild cognitive impairment and early alzheimer’s dementia by perfusion-weighted magnetic resonance imaging. *Neuroimage*, 40(2):495–503, 2008.

- [53] A. Du, G. Jahng, S. Hayasaka, J. Kramer, H. Rosen, M. Gorno-Tempini, K. Rankin, B. Miller, M. Weiner, and N. Schuff. Hypoperfusion in frontotemporal dementia and alzheimer disease by arterial spin labeling mri. *Neurology*, 67(7):1215–1220, 2006.
- [54] J. Udelson, J. Beshansky, D. Ballin, J. Feldman, G. Griffith, J. and Heller, R. Hendel, J. Pope, R. Ruthazer, E. Spiegler, et al. Myocardial perfusion imaging for evaluation and triage of patients with suspected acute cardiac ischemia: a randomized controlled trial. *Jama*, 288(21):2693–2700, 2002.
- [55] P. Martirosian, U. Klose, I. Mader, and F. Schick. FAIR true-FISP perfusion imaging of the kidneys. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 51(2):353–361, 2004.
- [56] S. Francis, K. Duncan, R. Moore, P. Baker, I. Johnson, and P. Gowland. Non-invasive mapping of placental perfusion. *The Lancet*, 351(9113):1397–1399, 1998.
- [57] S. Huettel, A. Song, G. McCarthy, et al. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA, 2004.
- [58] D. McRobbie, E. Moore, M. Graves, and M. Prince. *MRI from Picture to Proton*. Cambridge university press, 2006.
- [59] D. Twieg. The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical physics*, 10(5):610–621, 1983.
- [60] R. Buxton, L. Frank, E. Wong, B. Siewert, S. Warach, and R. Edelman. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic Resonance in Medicine*, 40(3):383–396, 1998.
- [61] E. Wong, R. Buxton, and L. Frank. A theoretical and experimental compari-

- son of continuous and pulsed arterial spin labeling techniques for quantitative perfusion imaging. *Magnetic Resonance in Medicine*, 40(3):348–355, 1998.
- [62] J. Wells, M. Lythgoe, D. Gadian, R. Ordidge, and D.L. Thomas. In vivo Hadamard encoded continuous arterial spin labeling (H-CASL). *Magnetic Resonance in Medicine*, 63(4):1111–1118, 2010.
- [63] M. Chappell. FSL course notes: Quantification issues in arterial spin labeling perfusion MRI, 2014.
- [64] E. Wong, L. Frank, and R. Buxton. QUIPSS II: a method for improving quantitation of perfusion using pulsed arterial spin labeling. In *Proceedings of the ISMRM 5th Annual Meeting, Vancouver*, page 1761, 1997.
- [65] E. Wong, R. Buxton, and L. Frank. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). *Magnetic Resonance in Medicine*, 39(5):702–708, 1998.
- [66] W. Luh, E. Wong, P. Bandettini, and J. Hyde. QUIPSS II with thin-slice TI 1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. *Magnetic Resonance in Medicine*, 41(6):1246–1254, 1999.
- [67] B. J. MacIntosh, W. Swardfager, A. Robertson, E. Tchistiakova, M. Saleem, P. Oh, N. Herrmann, B. Stefanovic, and K. Lanctôt. Regional cerebral arterial transit time hemodynamics correlate with vascular risk factors and cognitive function in men with coronary artery disease. *American Journal of Neurology*, 36(2):295–301, 2015.
- [68] S. Gevers, M. van Osch, R. Bokkers, D. Kies, W. Teeuwisse, C. Majoie, J. Hendrikse, and A. Nederveen. Intra-and multicenter reproducibility of pulsed, continuous and pseudo-continuous arterial spin labeling methods for measuring cerebral perfusion. *Journal of Cerebral Blood Flow and Metabolism*, 31(8):1706–1715, 2011.

- [69] M. van Osch, J. Hendrikse, and J. van der Grond. Sensitivity comparison of multiple vs. single inversion time pulsed arterial spin labeling fMRI. *Journal of Magnetic Resonance Imaging*, 25(1):215–221, 2007.
- [70] J. Wang, D. Alsop, H.K. Song, J. Maldjian, K. Tang, A. Salvucci, and J. Detre. Arterial transit time imaging with flow encoding arterial spin tagging (FEAST). *Magnetic Resonance in Medicine*, 50(3):599–607, 2003.
- [71] J. Xie, D. Gallichan, R. Gunn, and P. Jezzard. Optimal design of pulsed arterial spin labeling MRI experiments. *Magnetic Resonance in Medicine*, 59(4):826–834, 2008.
- [72] E. Wong. Quantifying CBF with pulsed ASL: technical and pulse sequence factors. *Journal of Magnetic Resonance Imaging*, 22(6):727–731, 2005.
- [73] H. Lu, C. Clingman, X. Golay, and P. van Zijl. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. *Magnetic Resonance in Medicine*, 52(3):679–682, 2004.
- [74] M. Chappell, A. Groves, B. Whitcher, and M. Woolrich. Variational Bayesian inference for a nonlinear forward model. *Signal Processing, IEEE Transactions on*, 57(1):223–236, 2009.
- [75] M. Chappell, A. Groves, B. MacIntosh, M. Donahue, P. Jezzard, and M. Woolrich. Partial volume correction of multiple inversion time arterial spin labeling MRI data. *Magnetic Resonance in Medicine*, 65(4):1173–1183, 2011.
- [76] M. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. Smith. Bayesian analysis of neuroimaging data in FSL. *Neuroimage*, 45(1):S173–S186, 2009.
- [77] S. Haller, G. Zaharchuk, D. L. Thomas, K. O. Lovblad, F. Barkhof, and X. Golay. Arterial spin labeling perfusion of the brain: emerging clinical applications. *Radiology*, 281(2):337–356, 2016.

- [78] M. Vidorreta, Z. Wang, I. Rodríguez, M. Pastor, J. A. Detre, and M. Fernández-Seara. Comparison of 2d and 3d single-shot asl perfusion fmri sequences. *Neuroimage*, 66:662–671, 2013.
- [79] E. T. Petersen, T. Lim, and X. Golay. Model-free arterial spin labeling quantification approach for perfusion MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 55(2):219–232, 2006.
- [80] E. Wong. Vessel-encoded arterial spin-labeling using pseudocontinuous tagging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1086–1091, 2007.
- [81] M Gunther. Highly efficient accelerated acquisition of perfusion inflow series by cycled arterial spin labeling. In *Proc Intl Soc Mag Reson Med*, volume 15, page 380, 2007.
- [82] Pan Su, Deng Mao, Peiying Liu, Yang Li, Marco C Pinho, Babu G Welch, and Hanzhang Lu. Multiparametric estimation of brain hemodynamics with MR fingerprinting ASL. *Magnetic resonance in medicine*, 78(5):1812–1823, 2017.
- [83] O. Leontiev and R. B. Buxton. Reproducibility of BOLD, perfusion, and CMRO₂ measurements with calibrated-BOLD fMRI. *Neuroimage*, 35(1):175–184, 2007.
- [84] J. Wells, B. Siow, M. Lythgoe, and D. L. Thomas. Measuring biexponential transverse relaxation of the ASL signal at 9.4 t to estimate arterial oxygen saturation and the time of exchange of labeled blood water into cortical brain tissue. *Journal of Cerebral Blood Flow & Metabolism*, 33(2):215–224, 2013.
- [85] J. Gregori, N. Schuff, R. Kern, and M. Günther. T₂-based arterial spin labeling measurements of blood to tissue water transfer in human brain. *Journal of Magnetic Resonance Imaging*, 37(2):332–342, 2013.

- [86] P. W. Hales and C. Clark. Combined arterial spin labeling and diffusion-weighted imaging for noninvasive estimation of capillary volume fraction and permeability-surface product in the human brain. *Journal of Cerebral Blood Flow & Metabolism*, 33(1):67–75, 2013.
- [87] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.
- [88] C. Windischberger, H. Langenberger, T. Sycha, E. Tschernko, G. Fuchsjäger-Mayerl, L. Schmetterer, and E. Moser. On the origin of respiratory artifacts in bold-epi of the human brain. *Magnetic resonance imaging*, 20(8):575–582, 2002.
- [89] X. Liang, A. Connelly, and F. Calamante. Improved partial volume correction for single inversion time arterial spin labeling data. *Magnetic Resonance in Medicine*, 69(2):531–537, 2013.
- [90] R. A. Oliver. *Improved quantification of arterial spin labelling images using partial volume correction techniques*. PhD thesis, UCL (University College London), 2015.
- [91] I. Asllani, A. Borogovac, and T. Brown. Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. *Magnetic Resonance in Medicine*, 60(6):1362–1371, 2008.
- [92] S. Spann, K. Kazimierski, C. Aigner, M. Kraiger, K. Bredies, and R. Stollberger. Spatio-temporal TGV denoising for ASL perfusion imaging. *Neuroimage*, 2017.
- [93] J. Petr, G. Schramm, F. Hofheinz, J. Langner, and J. van den Hoff. Partial volume correction in arterial spin labeling using a Look-Locker sequence. *Magnetic resonance in medicine*, 70(6):1535–1543, 2013.
- [94] D. Bruening, S. Dharssi, R. Lazar, R. Marshall, and I. Asllani. Improved partial volume correction method for detecting brain activation in disease

- using Arterial Spin Labeling (ASL) fMRI. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 5441–5444. IEEE, 2015.
- [95] J. Wells, D. L. Thomas, M. King, A. Connelly, M. Lythgoe, and F. Calamante. Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image denoising. *Magnetic resonance in medicine*, 64(3):715–724, 2010.
- [96] K. Kim, S. Choi, and S. H. Park. Improving arterial spin labeling by using deep learning. *Radiology*, page 171154, 2017.
- [97] C. Ulas, G. Tetteh, S. Kaczmarz, C. Preibisch, and B. Menze. DeepASL: Kinetic model incorporated loss for denoising arterial spin labeled MRI via deep residual learning. *arXiv preprint arXiv:1804.02755*, 2018.
- [98] J. Petr, J. C. Ferré, J. Y. Gauvrit, and C. Barillot. Improving arterial spin labeling data by temporal filtering. In *Medical Imaging 2010: Image Processing*, volume 7623, page 76233B. International Society for Optics and Photonics, 2010.
- [99] A. Groves, M. Chappell, and M. Woolrich. Combined spatial and non-spatial prior for inference on MRI time-series. *Neuroimage*, 45(3):795–809, 2009.
- [100] Z. Wang. Improving cerebral blood flow quantification for arterial spin labeled perfusion MRI by removing residual motion artifacts and global signal fluctuations. *Magnetic Resonance Imaging*, 30(10):1409–1415, 2012.
- [101] Z. Zun, A. Shankaranarayanan, and G. Zaharchuk. Pseudocontinuous arterial spin labeling with prospective motion correction (PCASL-PROMO). *Magnetic Resonance in Medicine*, 72(4):1049–1056, 2014.
- [102] M. Hoßbach, J. Gregori, S. Wesarg, and M. Günther. Head motion compensation for arterial spin labeling using optical motion tracking. In K. Drechsler, M. Erdt, M. Linguraru, L. Oyarzun, K. Sharma, R. Shekhar, and S. Wesarg,

- editors, *Clinical Image-Based Procedures. From Planning to Intervention*, volume 7761 of *Lecture Notes in Computer Science*, pages 1–8. Springer Berlin Heidelberg, 2013.
- [103] H. Tan, J. Maldjian, J. Pollock, et al. A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI. *JMRI*, 29(5):1134–1139, 2009.
- [104] C. Maumet, P. Maurel, J. C. Ferré, and C. Barillot. Robust estimation of the cerebral blood flow in arterial spin labelling. *Magnetic resonance imaging*, 32(5):497–504, 2014.
- [105] S. Dolui, Z. Wang, R. Shinohara, D. Wolk, J. A. Detre, and Alzheimer’s Disease Neuroimaging Initiative. Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series. *Journal of Magnetic Resonance Imaging*, 45(6):1786–1797, 2017.
- [106] A. Tanenbaum, A. Snyder, M. Brier, et al. A method for reducing the effects of motion contamination in arterial spin labeling MRI. *Journal of Cerebral Blood Flow & Metabolism*, 35(10):1697–1702, 2015.
- [107] Z. Shirzadi, D. Crane, A. Robertson, et al. Automated removal of spurious intermediate cerebral blood flow volumes improves image quality among older patients: a clinical arterial spin labeling investigation. *MRM*, 42(5):1377–1385, 2015.
- [108] D. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.
- [109] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer’s disease and its prodromal stages. *NeuroImage*, 155:530–548, 2017.
- [110] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT Press, 2016.

- [111] K. Li, X. Zhu, N. Hylton, G. Jahng, M. Weiner, and N. Schuff. Four-phase single-capillary stepwise model for kinetics in arterial spin labeling MRI. *Magnetic Resonance in Medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 53(3):511–8, March 2005.
- [112] J. Zhou, D. Wilson, J. Ulatowski, R. Traystman, and P. van Zijl. Two-compartment exchange model for perfusion quantification using arterial spin tagging. *Journal of Cerebral Blood Flow and Metabolism*, 21(4):440–55, April 2001.
- [113] K. Chaloner. Bayesian experimental design: a review. *Statistical Science*, 1995.
- [114] D. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.
- [115] X. Huan and Y. Marzouk. Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, January 2013.
- [116] A. Overstall and D. Woods. Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, 59(4):458–470, 2017.
- [117] J. DiStefano. Optimized blood sampling protocols and sequential design of kinetic experiments. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 240(5):R259–R265, 1981.
- [118] J. Jones. Optimal sampling strategies for the measurement of relaxation times in proteins. *Journal of Magnetic Resonance*, 126(2):283–286, 1997.
- [119] X. Li, D. Feng, and K. Wong. A general algorithm for optimal sampling schedule design in nuclear medicine imaging. *Computer Methods and Programs in Biomedicine*, 65(1):45–59, 2001.

- [120] M. Cercignani and D. Alexander. Optimal acquisition schemes for in vivo quantitative magnetization transfer MRI. *Magnetic Resonance in Medicine*, 56(4):803–810, 2006.
- [121] R. Samson, M. Symms, M. Cercignani, D. Tozer, and P. Tofts. Optimisation of quantitative magnetization transfer (QMT) sequence acquisition parameters. In *Proceedings of the 14th Annual Meeting of ISMRM*, page 2493, 2006.
- [122] A. Pineda, S. Reeder, Z. Wen, and N. Pelc. Cramér–rao bounds for three-point decomposition of water and fat. *Magnetic Resonance in Medicine*, 54(3):625–635, 2005.
- [123] O. Brihuega-Moreno, F. Heese, and L. Hall. Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage. *Magnetic Resonance in Medicine*, 50(5):1069–1076, 2003.
- [124] D. Alexander and G. Barker. Optimal imaging parameters for fiber-orientation estimation in diffusion MRI. *Neuroimage*, 27(2):357–367, 2005.
- [125] J. Xie, S. Clare, D. Gallichan, R. Gunn, and P. Jezzard. Real-time adaptive sequential design for optimal acquisition of arterial spin labeling MRI data. *Magnetic Resonance in Medicine*, 64(1):203–210, 2010.
- [126] N. Santos, J. Sanches, I. Sousa, and P. Figueiredo. Optimal sampling and estimation in PASL perfusion imaging. *IEEE Transactions on Biomedical Engineering*, 58(11):3165–3174, 2011.
- [127] N. Maleki, W. Dai, and D. C. Alsop. Optimization of background suppression for arterial spin labeling perfusion imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 25(2):127–133, 2012.
- [128] J. Kramme, J. Gregori, V. Diehl, V. Madai, F. von Samson-Himmelstjerna, M. Lentschig, J. Sobesky, and M. Günther. Improving perfusion quantification in arterial spin labeling for delayed arrival times by using optimized

- acquisition schemes. *Zeitschrift fuer Medizinische Physik*, 25(3):221–229, 2015.
- [129] M. Johnston, K. Lu, J. Maldjian, and Y. Jung. Multi-TI arterial spin labeling MRI with variable TR and bolus duration for cerebral blood flow and arterial transit time mapping. *IEEE transactions on medical imaging*, 34(6):1392–1402, 2015.
- [130] X. Golay, M. Guenther, et al. COST Action BM1103: ASL Initiative in Dementia (AID). <http://www.aslindementia.org/>.
- [131] J. Woods, M. Chappell, and T. Okell. A general framework for optimizing arterial spin labeling MRI experiments. *Magnetic resonance in medicine*, 2018.
- [132] T. Hahn. Cuba – a library for multidimensional numerical integration. *Computer Physics Communications*, 168(2):78–95, 2005.
- [133] S. Johnson. The NLOpt nonlinear-optimization package, 2014.
- [134] P. Kaelo and M. Ali. Some variants of the controlled random search algorithm for global optimization. *Journal of Optimization Theory and Applications*, 130(2), 2006.
- [135] Roger J Ordidge, Marzena Wylezinska, James W Hugg, Edward Butterworth, and Florence Franconi. Frequency offset corrected inversion (foci) pulses for use in localized spectroscopy. *Magnetic resonance in medicine*, 36(4):562–566, 1996.
- [136] W. Dai, P. Robson, A. Shankaranarayanan, and D. Alsop. Reduced resolution transit delay prescan for quantitative continuous arterial spin labeling perfusion imaging. *Magnetic Resonance in Medicine*, 67(5):1252–65, May 2012.

- [137] Youngseob Seo and Zhiyue J Wang. Mri scanner-independent specific absorption rate measurements using diffusion coefficients. *Journal of applied clinical medical physics*, 18(4):224–229, 2017.
- [138] E. De Vita, M. Günther, X. Golay, and D.L. Thomas. Magnetisation transfer effects of Q2TIPS pulses in ASL. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 25(2):113–126, 2012.
- [139] MathWorks Inc. MATLAB documentation: ranova(). <https://uk.mathworks.com/help/stats/repeatedmeasuresmodel.ranova.html>.
- [140] M.J. Cardoso, M. Modat, R. Wolz, A. Melbourne, D. Cash, D. Rueckert, and S. Ourselin. Geodesic information flows: spatially-variant graphs and their application to segmentation and fusion. *IEEE transactions on Medical Imaging*, 34(9):1976–1988, 2015.
- [141] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001.
- [142] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.
- [143] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.
- [144] Adnan Bibic, Linda Knutsson, Freddy Ståhlberg, and Ronnie Wirestam. Denoising of arterial spin labeling data: wavelet-domain filtering compared with gaussian smoothing. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(3):125–137, 2010.
- [145] Kenya Murase, Youichi Yamazaki, Masaaki Shinohara, Kazunori Kawakami, Keiichi Kikuchi, Hitoshi Miki, Teruhito Mochizuki, and Junpei Ikezoe. An

- anisotropic diffusion method for denoising dynamic susceptibility contrast-enhanced magnetic resonance images. *Physics in Medicine & Biology*, 46(10):2713, 2001.
- [146] Ruogu Fang, Junzhou Huang, and Wen-Ming Luh. A spatio-temporal low-rank total variation approach for denoising arterial spin labeling mri data. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 498–502. IEEE, 2015.
- [147] Jan Petr, Jean-Christophe Ferré, Jean-Yves Gauthier, and Christian Barillot. Denoising arterial spin labeling mri using tissue partial volume. In *Medical Imaging 2010: Image Processing*, volume 7623, page 76230L. International Society for Optics and Photonics, 2010.
- [148] M. Orton, D. Collins, D. Koh, and M. Leach. Improved intravoxel incoherent motion analysis of diffusion weighted imaging by data driven Bayesian modeling. *Magnetic Resonance in Medicine*, 71(1):411–420, 2014.
- [149] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.
- [150] L. Parkes, W. Rashid, D. Chard, and P. Tofts. Normal cerebral perfusion measurements using arterial spin labeling: reproducibility, stability, and age and gender effects. *Magnetic Resonance in Medicine*, 51(4):736–743, 2004.
- [151] T. Yoshiura, A. Hiwatashi, T. Noguchi, K. Yamashita, Y. Ohyagi, A. Monji, E. Nagao, H. Kamano, O. Togao, and H. Honda. Arterial spin labelling at 3t MR imaging for detection of individuals with Alzheimers disease. *European radiology*, 19(12):2819, 2009.
- [152] Kevin P Murphy. Conjugate Bayesian analysis of the gaussian distribution. https://www.cse.iitk.ac.in/users/piyush/courses/tpmi_winter19/readings/bayesGauss.pdf, 2007.

- [153] J. Veraart, E. Fieremans, and D. Novikov. Diffusion MRI noise mapping using random matrix theory. *Magnetic resonance in medicine*, 76(5):1582–1593, 2016.
- [154] Y. Liu, X. Zhu, D. Feinberg, M. Guenther, J. Gregori, M. Weiner, and N. Schuff. Arterial spin labeling MRI study of age and gender effects on brain perfusion hemodynamics. *Magnetic Resonance in Medicine*, 68(3):912–922, 2012.
- [155] Raquel E Gur and Ruben C Gur. Gender differences in regional cerebral blood flow. *Schizophrenia Bulletin*, 16(2):247–254, 1990.
- [156] Maddie J Pascoe, Tracy R Melzer, L John Horwood, Lianne J Woodward, and Brian A Darlow. Altered grey matter volume, perfusion and white matter integrity in very low birthweight adults. *NeuroImage: Clinical*, 22:101780, 2019.
- [157] Y. Li, J. B. Huang, Ahuja N., and M. H. Yang. Deep joint image filtering. In *European Conference on Computer Vision*, 2016.
- [158] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
- [159] Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.
- [160] Peng Liu and Ruogu Fang. Wide inference network for image denoising via learning pixel-distribution prior. *arXiv preprint arXiv:1707.05414*, 2017.
- [161] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 538–554, 2018.

- [162] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. Deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.
- [163] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017.
- [164] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.
- [165] Ariel Benou, Ronel Veksler, Alon Friedman, and T Riklin Raviv. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced mri sequences. *Medical image analysis*, 42:145–159, 2017.
- [166] Ryutaro Tanno, Daniel E Worrall, Aurobrata Ghosh, Enrico Kaden, Stamatios N Sotiropoulos, Antonio Criminisi, and Daniel C Alexander. Bayesian image quality transfer with cnns: exploring uncertainty in dmri super-resolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 611–619. Springer, 2017.
- [167] E Gong, J Pauly, and G Zaharchuk. Boosting snr and/or resolution of arterial spin label (asl) imaging using multi-contrast approaches with multi-lateral guided filter and deep networks. In *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Honolulu, Hawaii*, 2017.
- [168] G Zaharchuk, E Gong, M Wintermark, D Rubin, and CP Langlotz. Deep learning in neuroradiology. *American Journal of Neuroradiology*, 39(10):1776–1784, 2018.

- [169] Piet Bladt, Arnold J den Dekker, Patricia Clement, Eric Achten, and Jan Sibbers. The costs and benefits of estimating T1 of tissue alongside cerebral blood flow and arterial transit time in pseudo-continuous arterial spin labeling. *NMR in Biomedicine*, page e4182, 2019.
- [170] Pedro A Gómez, Miguel Molina-Romero, Guido Buonincontri, Marion I Menzel, and Bjoern H Menze. Designing contrasts for rapid, simultaneous parameter quantification and flow visualization with quantitative transient-state imaging. *Scientific reports*, 9(1):1–12, 2019.
- [171] Anish Lahiri, Jeffrey A Fessler, and Luis Hernandez-Garcia. Optimizing MRF-ASL scan design for precise quantification of brain hemodynamics using neural network regression. *Magnetic resonance in medicine*, 2019.
- [172] Eric C Wong, Matthew Cronin, Wen-Chau Wu, Ben Inglis, Lawrence R Frank, and Thomas T Liu. Velocity-selective arterial spin labeling. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 55(6):1334–1341, 2006.
- [173] Dimitra Flouri, David Owen, Rosalind Aughwane, Nada Mufti, Kasia Maksym, Magdalena Sokolska, Giles Kendall, Alan Bainbridge, David Atkinson, Tom Vercauteren, et al. Improved fetal blood oxygenation and placental estimated measurements of diffusion-weighted mri using data-driven bayesian modeling. *Magnetic resonance in medicine*, 2019.
- [174] Stefan M Spann, Xingfeng Shao, Danny JJ Wang, Christoph S Aigner, Matthias Schloegl, Kristian Bredies, and Rudolf Stollberger. Robust single-shot acquisition of high resolution whole brain asl images by combining time-dependent 2d capirinha sampling with spatio-temporal tgv reconstruction. *NeuroImage*, 206:116337, 2020.
- [175] Bram van Ginneken, Sjoerd Kerkstra, and James Meakin. Grand Challenges in Biomedical Image Analysis. https://grand-challenge.org/Why_Challenges/.

- [176] Danfeng Xie, Yiran Li, Hanlu Yang, Li Bai, Tianyao Wang, Fuqing Zhou, Lei Zhang, and Ze Wang. Denoising arterial spin labeling perfusion MRI with deep machine learning. *Magnetic Resonance Imaging*, 2020.
- [177] Kuang Gong, Paul Han, Georges El Fakhri, Chao Ma, and Quanzheng Li. Arterial spin labeling mr image denoising and reconstruction using unsupervised deep learning. *NMR in Biomedicine*, page e4224, 2019.
- [178] Zheng Li, Qingping Liu, Yiran Li, Qiu Ge, Yuanqi Shang, Donghui Song, Ze Wang, and Jun Shi. A two-stage multi-loss super-resolution network for arterial spin labeling magnetic resonance imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 12–20. Springer, 2019.