A frequency-domain machine learning method for dual-calibrated fMRI mapping of oxygen extraction fraction (OEF) and cerebral metabolic rate of oxygen consumption ($CMRO_2$)

Michael Germuska[1], Hannah Chandler[1], Thomas Okell[2], Fabrizio Fasano[3], Valentina Tomassini[1,4], Kevin Murphy[1], Richard Wise[1].

[1] CUBRIC, Department of Psychology, Cardiff University, Cardiff, United Kingdom

[2] Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom

[3] Siemens Healthcare Ltd, Frimley, Camberley, UK.

[4] Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, United Kingdom

*** Correspondence:**
Michael Germuska
germuskam@cardiff.ac.uk

Keywords: magnetic resonance imaging, metabolism, oxygen extraction fraction, $CMRO_2$, OEF, BOLD, artificial neural networks, machine learning, calibrated-fMRI.

Abstract

Magnetic resonance imaging (MRI) offers the ability to non-invasively map the brain's metabolic oxygen consumption ($CMRO_2$), which is essential for understanding and monitoring neural function in both health and disease. One MRI method of mapping $CMRO_2$ is based on the simultaneous acquisition of cerebral blood flow (CBF) and blood oxygen level dependent (BOLD) weighted images during respiratory modulation of both oxygen and carbon dioxide. Although this dual-calibrated methodology has shown promise in the research setting, current analysis methods are unstable in the presence of noise and/or are computationally demanding. In this paper, we present a machine learning implementation for the multi-parametric assessment of dual-calibrated fMRI data. An extremely randomized trees regressor and a multi-layer perceptron (MLP) are cascaded to provide quantitative estimates of the resting CBF, $CMRO_2$, and oxygen extraction fraction (OEF). The proposed implementation takes advantage of the inherent noise immunity of tree-based ensemble methods and MLPs to provide robust and computationally efficient estimates of CBF, $CMRO_2$, and OEF. Synthetic data with additive noise are used to train the regressors, and their performance is compared to conventional analysis methods both in simulation and with *in-vivo* data (n=30). The proposed method is demonstrated to significantly reduce computation time, error, and proportional bias in both $CMRO_2$ and OEF estimates.

1 Introduction

Under normal conditions the brain's energy needs are met via a continuous supply of oxygen and glucose for the local production of ATP via aerobic metabolism (Verweij et al., 2007). Any disruption of the supply of oxygen to the brain tissue can have significant consequences (Safar, 1988), and impaired cerebral oxygen metabolism is associated with a wide variety of neurological

conditions (Frackowiak et al., 1988; Ishii et al., 1996; Miles and Williams, 2008). Therefore, monitoring and mapping the brain's consumption of oxygen is vital for understanding the diseases and mechanisms by which the metabolic consumption of oxygen may be affected. The cerebral metabolic rate of oxygen consumption ($CMRO_2$) has traditionally been measured with positron emission tomography (Frackowiak et al., 1980). However, this method has some substantial limitations including the use of ionizing radiation and the need for local production of 15-oxygen labeled tracers. Due to these limitations there is great interest in developing alternative, non-invasive, methods of mapping $CMRO_2$. One promising technique of non-invasively mapping $CMRO_2$ is the so-called dual-calibrated fMRI (dc-fMRI) method (Bulte et al., 2012; Gauthier et al., 2012). This method is finding growing adoption in the research setting, and has already been applied in Alzheimer's disease (Lajoie et al., 2017), carotid artery occlusion (De Vis et al., 2015), and studies of pharmacological modulation (Merola et al., 2017). For a review of the method and details on the its practical application please see (Germuska and Wise, 2019). Despite the promise shown by this technique, the reported between-session repeatability is relatively low (Merola et al., 2018) and improvements in the data acquisition and/or analysis are required if individualized assessment is to be made possible.

One of the key difficulties in analyzing dual-calibrated fMRI data is noise propagation through the analysis pipeline, which leads to unstable parameter estimates. We have previously presented regularized non-linear least squares fitting approaches that utilize prior physiological knowledge to produce more robust parameter estimates (Germuska et al., 2019; Germuska et al., 2016). Even though such regularization reduces the mean square error it does so by trading off a reduction in variance with an increase in bias. An alternative approach to reduce the prediction error is the use of noise insensitive analysis methods. Multi-layer perceptrons (MLP) and tree-based ensemble methods are inherently robust to noise (Bernier et al., 1999; Breiman, 2001; Geurts et al., 2006) and can be trained to provide low biased parameter estimate in the presence of input noise. In this paper we present an analysis pipeline comprised of an extremely randomized trees regressor and a MLP, cascaded to infer resting CBF and $CMRO_2$ from dual-calibrated fMRI data. A frequency-domain representation of simulated MRI data with the additive noise is used to train each of the regressors. Simulated data has the advantage over *in-vivo* data in this application as it allows a balanced dataset to be generated that covers a broad range of physiological variation. Such a dataset is essential to avoid bias in parameter estimation and to provide generalizability across groups and diseases. A frequency-domain representation is chosen as it allows for convenient dimensionality reduction, with most of the information of interest encoded at low temporal frequencies, and takes advantage of the superior ability of artificial neural networks to learn discriminative features from frequency-domain representation of a signal compared to a time-domain representation (Hertel et al., 2016). The performance of the proposed machine learning (ML) implementation is compared to an existing regularized non-linear least squares (rNLS) method (Germuska et al., 2019) both in simulation and in data acquired from a cohort of 30 healthy volunteers. We hypothesized that the machine learning approach would be able to achieve comparable or reduced prediction error with significantly reduced bias and computational overhead.

2 MRI Data Acquisition

Thirty healthy volunteers (15 males, mean age 32.41 ± 6.13 years) were recruited to the study. The local ethics committee approved the study and written informed consent was obtained from each participant. Blood samples were drawn via a finger prick prior to scanning and were analyzed with the HemoCue Hb 301 System (HemoCue, Ängelholm, Sweden) to calculate the systemic [Hb] value for each participant. All data was acquired using a Siemens MAGNETOM Prisma (Siemens

Healthcare GmbH, Erlangen) 3T clinical scanner with a 32-channel receiver head coil (Siemens Healthcare GmbH, Erlangen). The acquisition protocol was as previously described (Germuska et al., 2019). Briefly, an 18-minute dual-excitation pseudo-continuous arterial spin labeling (pCASL) and BOLD-weighted acquisition was acquired during modulation of inspired oxygen and carbon dioxide. Gas modulation was performed according to a protocol previously proposed by our lab (Germuska et al., 2016), and end-tidal monitoring was performed throughout the acquisition from the volunteer's facemask using a rapidly responding gas analyzer (PowerLab®, ADInstruments, Sydney, Australia). The prototype pCASL sequence (Germuska et al., 2019) parameters were as follows: post-labeling delay and label duration 1.5 seconds, EPI readout with GRAPPA acceleration (factor = 3), TE_1 = 10ms, TE_2 = 30ms, TR = 4.4 seconds, 3.4 x 3.4mm in-plane resolution, and 15 (7mm) slices with 20% slice gap.

3 Synthetic MRI Data Generation

Synthetic data was simulated to match the 18-minute *in-vivo* acquisition protocol using standard physiological models for the change in BOLD signal (Bulte et al., 2012; Gauthier and Hoge, 2013; Wise et al., 2013), as summarized by equation 1.

$$\frac{\Delta BOLD}{BOLD_0} = \mathrm{M}\left\{1 - \left(\frac{CBF}{CBF_0}\right)^{\alpha}\left(\frac{1 - \frac{CaO_2 - CMRO_2/CBF}{\varphi[Hb]}}{1 - \frac{CaO_{2,0} - CMRO_{2,0}/CBF_0}{\varphi[Hb]}}\right)^{\beta}\right\} \qquad (1)$$

Where, $\Delta BOLD/BOLD_0$ is the fractional change in BOLD signal due to a change in arterial oxygen content (CaO_2) or CBF due to either a hyperoxic or hypercapnic respiratory stimulus. M is a lumped parameter that is equal to $K \cdot \left((1 - SvO_2) \cdot [Hb]\right)^{\beta}$. Where K is a scaling factor dependent on the field strength, resting venous blood volume, tissue structure, and water diffusion effects in the extravascular space. [Hb] is the blood hemoglobin concentration and SvO_2 is the venous oxygen saturation. φ is the oxygen binding capacity for Hb (1.34 ml/g), α is the Grubb exponent that couples blood volume and blood flow changes, and β is a field strength dependent constant that summarizes the non-linear effects associated with the tissue structure and water diffusion effects. The values of α and β were fixed to the optimized values (0.06 and 1) found by (Merola et al., 2016), which minimize the error in OEF estimates over a range of vascular physiology. The subscript 0 represents the baseline or resting state. The hyperoxic and hypercapnic stimuli are assumed to be iso-metabolic, so $CMRO_2 = CMRO_{2,0}$.

The arterial spin labeling signal was modeled according to the simplified pCASL kinetic model (Alsop et al., 2015), and physiological constraints on baseline parameters were applied according to a simple model of oxygen exchange (Gjedde, 2002; Hayashi et al., 2003), equation 2.

$$CMRO_{2,0} = \mathrm{D}\left[P_{50}\sqrt[h]{\frac{2}{OEF_0} - 1} - P_{minO_2}\right] \qquad (2)$$

Where D is the effective oxygen diffusivity of the capillary network and can be expressed as a product of the effective oxygen permeability and the capillary blood volume, $D = \kappa \cdot CBV_{cap}$. P_{50} is the blood oxygen tension at which hemoglobin is 50% saturated (26 mmHg), h is the Hill coefficient (2.8) and P_{minO_2} is the minimum oxygen tension at the mitochondria (which is thought to be

negligible in healthy tissue (Gjedde, 2002)). In the modeling we assume a fixed value for κ of 3 μmol/mmHg/ml/min, corresponding to a typical diffusivity of 3 (Mintun et al., 2001) to 4 μmol/100g/mmHg/min (Vafaee and Gjedde, 2000) for CBV_{cap} = 1 to 1.33 ml/100g. The physiological parameter space encompasses a wide range of plausible physiology including both healthy and dysfunctional brain tissue, and is summarized in Table 1. A summary of MRI abbreviations and all model parameters used in the simulations is given in table 2.

The partial pressure of arterial oxygen (PaO_2) and change in carbon dioxide ($\Delta PaCO_2$) were modeled to match the range of end-tidal recordings acquired from healthy volunteers. The baseline PaO_2 had a range of 90-120 mmHg, ΔPaO_2 was 200 to 300 mmHg, and $\Delta PaCO_2$ was set to 8-12 mmHg. Rectangular stimulus blocks were convolved with a gamma density function with shape parameter 0.5-2.5 to account for the variation in biological rise and fall times of the hyperoxic and hypercapnic stimuli. Drift in $\Delta PaCO_2$, which was observed in some subjects, was included by adding a bandpass filtered noise signal (4th order IIR filter, lowcut/highcut = 0.005/0.05 of the Nyquist frequency). Change in the arterial blood longitudinal relaxation rate due to dissolved oxygen was included in pCASL calculations as per (Germuska et al., 2019). Noise (BOLD tSNR = 90, pCASL tSNR = 3 for CBF = 60 ml/100g/min) was added to simulated BOLD and pCASL time series. The pCASL noise was bandpass filtered (4th order IIR filter, lowcut/highcut = 0.05/0.8 of the Nyquist frequency) and the BOLD noise was lowpass filtered (1st order IIR filter, highcut = 0.5 of the Nyquist frequency) to match the noise characteristics of the *in-vivo* data. In addition, the BOLD timeseries data was highpass filtered with a 320 second cut-off using the filter implementation in FSL (Jenkinson et al., 2012), which is routinely used for de-trending fMRI data. Figure 1 shows 50 randomly generated pCASL and BOLD timeseries overlaid with the temporal mean to demonstrate the typical output of the simulations. Please note that the pCASL timeseries are divided by the equilibrium magnetization of arterial blood ($M0_{blood}$), and the baseline signal has been set to zero for display purposes.

4 Methods

Simulated BOLD and ASL timeseries data were Fourier transformed along the time axis to produce magnitude and phase data, each data series was truncated after 15 data points. The truncation removes high-frequency content that is unrelated to either the hyperoxic or hypercapnic respiratory modulations. Feature vectors were constructed by concatenating the ASL and BOLD magnitude and phase data with the modeled [Hb], ΔPaO_2, $SaO_{2,0}$, $CaO_{2,0}$, and the post-label delay (used in the ASL analysis to include slice-wise post-labeling time), thus each feature vector had 65 entries. These feature vectors were used to train an extremely randomized trees regressor to infer the baseline cerebral blood flow (CBF_0). The resting CBF was then included in feature vectors (66 entries) used to train a MLP to infer $CMRO_{2,0}$ / $CaO_{2,0}$, from which $CMRO_{2,0}$ was calculated. The output from both of these regressors were used to calculate OEF_0 via the Fick principle, where $CMRO_2$ / CaO_2 = OEF x CBF.

The extremely randomized trees regressor and MLP were implemented in Scikit learn (Pedregosa et al., 2011). The extremely randomized trees regressor was trained with the following options, number of estimators = 50, bootstrap = True, and out-of-bag samples were used to estimate the R^2 on unseen data. A total of 50,000 simulations were used for training. The MLP network has two-hidden layers and 50 nodes in each layer. The activation function for each node was chosen to be a rectified linear unit (ReLU). The ADAM solver was used for training with $1x10^6$ simulated feature vectors and 10% of the data were used for early stopping. Data simulation and training was repeated 40 times to create an ensemble of MLP networks to further reduce the uncertainty in parameter estimates (Sollich and Krogh, 1996).

The validation score for the extremely randomized trees regressor for predicting resting cerebral blood flow was 0.997, slightly greater than the results obtained for a random forest implementation (0.961). The validation score for the MLP estimation of $CMRO_{2,0}$ / $CaO_{2,0}$ were 0.923 ± 0.002. Training of the MLP network was also undertaken while eliminating key elements of the simulation or feature vectors to see how this affected the performance of the MLP. When BOLD data was excluded from the feature vector the validation score dropped to 0.577. Excluding the CO_2 and O_2 stimuli (but including the BOLD data) reduced the validation scores to 0.63 and 0.71 respectively.

A further 5,000 simulated datasets (with OEF restricted to 0.15 to 0.65, all other parameters as in table 1) were constructed to compare the performance of the proposed machine learning implementation with a previously implemented regularized non-linear least squares fitting method (Germuska et al., 2019). Each method was compared to the simulated data using a robust regression method (bisquare) in terms of the rms error and proportional bias. A bisquare cost function was used for the regression to reduce the influence of outliers and allow a robust estimate of the proportional bias. The rNLS fitting was implemented with regularization applied to the resting OEF and the effective oxygen diffusivity (D), as previously described. The relative weighting between OEF and diffusivity regularization was maintained constant, as per the optimization in (Germuska et al., 2019). However, the total weighting was varied to assess the impact on OEF and $CMRO_2$ error and proportional bias (slope of the simulated parameter values plotted against the parameter estimates).

5 Results

5.1 Simulations

Analysis of the simulated data demonstrated a substantial reduction in the rms error of machine learning OEF estimates compared to rNLS estimates. The bisquare rms error was 0.047 when using the mean prediction from the 40 MLP networks, and 0.055 for a randomly chosen MLP network. The rNLS approach produced a minimum bisquare rms error of 0.094. The ML approach displayed negligible proportional bias in OEF estimates (slope of true vs. estimated values = 0.982), whereas rNLS estimates had variable levels of bias depending on the level of regularization, see figure 2a for a summary of the results. As expected from the OEF results, ML estimates of $CMRO_2$ also had significantly reduced error and bias compared to the rNLS implementation. The proportional bias for the ML implementation was 0.977 compared to a minimum bias of 0.913 for the rNLS method. The bisquare rms error in $CMRO_2$ estimates for the ML implementation was 20.3 μmol/100g/min (22.6 for an individual MLP network) whereas the error for rNLS estimates ranged from 29.6 to 52.4 μmol/100g/min depending on the level of bias (with greater bias coinciding with reduced error), see figure 2b.

Training of the MLP with reduced feature vectors (excluding the BOLD data) or limited respiratory stimuli (excluding either CO_2 or O_2 modulation) highlights the importance of each signal and stimulus in estimate the rate of oxygen consumption. As expected, removing the BOLD signal resulted in a significant reduction in the network's ability to estimate $CMRO_2$ (validation R^2 reduced from 0.923 for the full model to 0.58). In this instance there should be no information relating to OEF in the feature vector and so the inference is based solely on the correlation between baseline flow and $CMRO_2$ in the simulated data. Adding the BOLD data back in but with only an O_2 stimulus does little to improve the performance of the network (R^2 = 0.63). This is not unexpected as the hyperoxic BOLD signal is largely related to venous blood volume (Blockley et al., 2013) with little influence from OEF. Perhaps unexpectedly, including the CO_2 stimulus but not the O_2 stimulus significantly improves the ability of the network to infer resting $CMRO_2$ (R^2 = 0.71). While this is still

significantly worse than the full model, it suggests that some quantitative metabolic information may be extracted from hypercapnic calibration studies that are normally employed to estimate relative changes in $CMRO_2$ (Hoge, 2012). Additionally, such results suggest that the simulation framework could be utilized to optimize data acquisition by designing respiratory stimuli that maximize the performance of the ML implementation, and that such respiratory paradigms may be different compared to those for standard analysis methods (which are unable to infer resting $CMRO_2$ information from a hypercapnic calibration experiment).

5.2 In-vivo

Due to the limited availability and technical challenges associated with acquiring 15-oxygen PET data for $CMRO_2$ mapping (the gold standard approach) it is difficult to directly validate the *in-vivo* results obtained in this study. However, a number of fundamental relationships between resting physiological parameters have consistently been observed across groups of healthy individuals. Here we compare these observed relationships against the acquired data to infer the relative error and bias for each analysis method. One of the most frequently reported relationships in the healthy human brain is that resting blood flow is linearly correlated with resting oxygen metabolism (Coles et al., 2006; Lebrun-Grandie et al., 1983; Leenders et al., 1990; Powers et al., 2011; Scheinberg and Stead, 1949). Additionally, PET data suggests that the OEF should be approximately uniform across the cerebral grey matter e.g. (Hyder et al., 2016). Thus, we can use the coefficient of variation (COV) of grey matter OEF estimates as an indicator of parameter error, and examine the variation in the slope of the CBF-$CMRO_2$ relationship to infer the proportional bias or sensitivity to physiological variation of $CMRO_2$ estimates.

As in the simulation experiments we investigated the *in-vivo* analysis for varying levels of regularization in the rNLS analysis and compare this to the ML results. Figure 3b plots the COV in OEF estimates for increasing levels of regularization against the slope of the CBF-$CMRO_2$ regression (normalized by the slope of the ML estimate). As predicted by the simulations, the slopes of the ML estimates and the rNLS estimates are similar when little regularization is applied, with the slope of the rNLS estimates slightly reduced compared to the ML approach. As more regularization is applied the COV of OEF estimates is reduced and the slope between CBF and $CMRO_2$ decreases, clearly demonstrating the trade-off between variance and bias. Again, as predicted by the simulations, the COV in ML estimates is significantly less than COV in rNLS estimates for a similar CBF-$CMRO_2$ slope.

To investigate the bias in OEF estimates we take advantage of another physiological relationship reported in the literature; cerebral oxygen extraction is inversely related to [Hb] (Ibaraki et al., 2010) and the closely related parameter Hct (Morris et al., 2018). Taking the same approach as before we observe *in-vivo* results that closely match predictions from the simulation, see figure 3a. As in the simulations, the slope in the [Hb]-OEF relationship is similar between the ML method and rNLS approach for a moderate amount of regularization. However, the slope is substantially increased when using minimal regularization, and reduced when applying strong regularization.

Figure 4 shows scatter plots of the grey matter CBF-$CMRO_2$ and [Hb]-OEF relationships observed with the ML and rNLS methods across the 30 healthy volunteers studied. The rNLS results are shown for a single level of regularization, where the slope of the [Hb]-OEF relationship most closely matches that of the ML analysis (see figure 3). The coefficient of determination is greater for the ML approach for each relationship, with R^2 values of 0.56 and 0.35 for the CBF-$CMRO_2$ and [Hb]-OEF relationships, compared to 0.34 and 0.14 for the rNLS approach ($p<0.05$ for all correlations).

Table 3 reports the results of a bivariate analysis of [Hb] against OEF and CBF for both analysis methods. The slopes of the relationship between OEF and [Hb] are similar to that reported in healthy subjects by (Ibaraki et al., 2010), -1.75 Hb (g/dL). As per Ibaraki et al. the relationship between CBF and OEF did not reach significance ($p=0.44$) for the ML approach, however a significant negative correlation was observed in the rNLS analysis ($p=0.005$). A univariate analysis of $CMRO_{2,0}$ against CBF_0 is consistent with that observed in healthy controls by (Powers et al., 2011) ($\beta1 = 0.2$) for both analysis methods, $\beta1 = 0.32$ ($p<0.001$) and $\beta1 = 0.24$ ($p<0.001$) for the ML and rNLS approaches respectively.

Figure 5 shows a comparison between CBF_0, OEF_0 and $CMRO_{2,0}$ parameter maps calculated with the ML method (single MLP network and ensemble of 40 networks) and the rNLS method. The image shows 7 slices from a single subject, which have been interpolated for display using cubic b-spline interpolation (Ruijters and Thevenaz, 2012) using FSLeyes (10.5281/zenodo.1470761). As expected OEF_0 is not well estimated in the white matter, due to the T_1 decay of the arterial spin labeling signal and the longer arrival time of white matter blood. Across grey matter containing voxels maps of OEF_0 calculated with the ML methods are more uniform than those calculated with the rNLS approach, with the ensemble approach visibly outperforming the singe network MLP estimates. These observations are consistent with the results of the simulations and the grey matter COV observed for *in-vivo* OEF_0 estimates. However, it is also apparent from the images that each method demonstrates sensitivity to regional susceptibility effects. For example, in the pre-frontal cortex and inferior temporal lobes the images show greater variability in OEF_0 estimates, with regions of both over and under-estimation apparent. This instability is likely due to reduced BOLD SNR in these locations and alteration of the susceptibility of air in and around the nasal cavity and paranasal sinuses due to modulation of the inspired oxygen content during data acquisition. It is clear that the ML estimates, in particular those made from the ensemble of MLPs, are more robust to such regional susceptibility effects.

The *in-vivo* analysis also highlights the improvement in computational efficiency of the proposed method. The rNLS approach took approximately 20 minutes to analyze a complete dataset on a standard laptop (2.8 Ghz Intel Core i7, 16GB memory), while the ML approach was able to complete the same analysis in approximately 10 to 20 seconds (depending on the number of networks in the ensemble of MLP regressors).

6 Discussion and Conclusions

Instability in parameter estimates made using noisy *in-vivo* data may be reduced by incorporating prior knowledge of physiological parameters, e.g. (Chappell et al., 2010; Frau-Pascual et al., 2014; Germuska et al., 2016; Mesejo et al., 2015). Previous investigation of such methods (Germuska et al., 2016) suggests that they are an effective means to increase the robustness of $CMRO_2$ estimates made with dc-fMRI. However, these methods are computationally expensive and must necessarily make a trade off between parameter uncertainty and parameter sensitivity. Thus, they are not well suited to high throughput or rapid data analysis and care must be taken when using such methods not to unduly bias parameter estimates towards the priors. In the work presented here we take a different approach by training a machine learning implementation that is robust to input noise. Given an appropriately selected (or generated) training dataset, a well-implemented solution will be unbiased, robust, and have a low computational overhead.

Computer modeling suggests that the proposed method outperforms previous analysis methods both in terms of uncertainty and bias. *In-vivo* data supports the predicted improvement in uncertainty with

a significant reduction in the COV of grey matter OEF_0 estimates when compared to a regularized non-linear least squares fitting of the data. Additionally, agreement was found between the predicted behaviors of each method and their associated biases when compared to reported physiological relationships. Qualitatively, the *in-vivo* parameter maps suggest that the ML approach, especially when paired with an ensemble implementation, is more robust to physiological noise; producing physiologically plausible parameter estimates in challenging brain regions, e.g. near the frontal sinuses. Such physiological noise was not modeled in the training data so it is perhaps unexpected that the ML method is robust to these noise sources. However, it is plausible that the discriminative features identified from the frequency-domain representation of the data during training are less sensitive to these regional susceptibility changes than a traditional time-domain fit of the data. It is possible that this aspect of the ML approach could be enhanced by extending the training data to include such regional susceptibility changes, either on their own or in combination with a spatially informed approach to data fitting.

The use of an ensemble of MLP networks reduced parameter uncertainty in simulation and reduced the coefficient of variation in grey matter OEF_0 estimates *in-vivo*, demonstrating its utility in this application. However, it is anticipated that enforcing network diversity during training could make further improvements in performance. As it is has previously been demonstrated that, in the presence of noise, the performance of an ensemble of networks can always be improved by explicitly encouraging diversity during training (Reeve and Brown, 2018).

A limitation of the proposed method is the need to train new regressors for a given gas paradigm and set of acquisition parameters, e.g. arterial spin labeling tagging duration, repetition time and duration of the acquisition. In addition, there is a requirement that the *in-vivo* gas manipulation does not deviate significantly from the range of simulated designs. While it is a relatively straightforward process to retrain the regressors with a new set of parameters, to match the local acquisition protocol, the scope of the method could be increased if individualized gas traces could be incorporated into the training data; allowing a single pre-trained implementation to be applied across studies.

The simulations and *in-vivo* results suggest that the proposed analysis method could significantly increase the utility of dc-fMRI, reducing the number of participants needed to detect a group difference in oxygen metabolism or oxygen extraction fraction and offering more physiological interpretability of metabolic differences or alteration due to a stimulus. In addition, the significant reduction in processing time and the improved robustness of the individual parameter maps reduces two of the hurdles restricting clinical implementation of such techniques.

Tables

OEF	CBF (ml/100g/min)	[Hb] (g/dL)	Mean capillary transit time (CBV_{cap} / CBF, seconds)	$PminO_2$ (mmHg)	Cerebral Vascular Reactivity (% CBF / mmHg CO_2)	K
0.05 – 0.75	1 – 250	10-18	0.25 – 4.0	0 - 30	1 - 7	0.01 – 0.25

Table 1. Range of physiological parameters used in the dc-fMRI data simulations for training of the machine learning regressors.

Variable / abbreviation	Expression (units)
OEF	Oxygen Extraction Fraction (dimensionless)
$CMRO_2$	Cerebral Metabolic Rate of Oxygen consumption (μmol/100g/min)
CBF	Cerebral Blood Flow (ml/100g/min)
φ	Oxygen binding capacity of hemoglobin (1.34 ml/g)
[Hb]	Hemoglobin concentration (g/dL)
CaO_2	Arterial oxygen content (ml/ml)
PaO_2	Arterial oxygen tension (mmHg)
SaO_2	Arterial oxygen saturation (dimensionless)
SvO_2	Venous oxygen saturation (dimensionless)
α	Grubb exponent
β	Venous morphology / deoxy-hemoglobin - BOLD exponent
BOLD	Blood Oxygenation Level Dependent signal
ASL	Arterial Spin Labeling
$M0_{blood}$	Arterial blood MRI signal equilibrium magnetization (dimensionless)
PLD	ASL post-label delay time (1.0 – 3.0 seconds)
M	Maximum possible BOLD signal (BOLD calibration parameter)
K	BOLD scaling factor = $M / ([Hb] \times (1-SvO_2))^{\beta}$
D	Effective oxygen diffusivity of the capillary network (μmol/100g/mmHg/min)
CBV_{cap}	Capillary blood volume (ml/100g)
$PminO_2$	Minimum oxygen partial pressure at the mitochondria (mmHg)
h	Hill coefficient (2.8)
κ	Effective permeability of capillary endothelium and brain tissue (μmol/mmHg/ml/min)

Table 2. Summary of model parameters and abbreviation used in the dc-fMRI data simulations and their definitions.

Predictor	ML β1 (p value)	rNLS β1 (p value)
OEF	-1.42 (0.001)	-2.23 (0.001)
CBF	-0.07 (0.44)	-0.37 (0.005)
Intercept	61.95 (<0.001)	89.48 (<0.001)

Table 3. Results of a bivariate regression of [Hb] against CBF_0 and OEF_0 grey mater estimates for 30 healthy volunteers analyzed with the ML (ensemble of MLPs) and rNLS fitting methods.

7 Conflict of Interest

Author FF was employed by company Siemens Healthcare Ltd. All other authors declare no competing interests.

8 Author Contributions

MG. Wrote the manuscript, developed and implemented the methods, and analysed the in-vivo data.

HC. Acquired and processed data and edited the manuscript.

TO. Created and provided code used in the prototype pseudo-continuous arterial spin labeling pulse sequence.

FF. Assisted in the implementation of the prototype arterial spin labeling pulse sequence.

VT. Chief investigator overseeing study design and data collection for a subset of healthy controls.

KM. Principal investigator overseeing study design and data collection for a subset of healthy controls.

RW. Principal investigator overseeing study design and data collection for a subset of healthy controls

All authors reviewed and edited the manuscript prior to submission.

9 Funding

Wellcome Strategic Award, ‘Multi-scale and multi-modal assessment of coupling in the healthy and diseased brain’, grant reference 104943/Z/14/Z (RW, MG, HC). RW is also supported by the Higher Education Funding Council for Wales. KM is supported by Wellcome grant 200804/Z/16/Z. TO is supported by the Royal Academy of Engineering and Wellcome Centre for Integrative Neuroimaging is supported by core funding from Wellcome (203139/Z/16/Z).

10 Acknowledgments

We would like to thank Wellcome for supporting this work: Wellcome Strategic Award, 'Multi-scale and multi-modal assessment of coupling in the healthy and diseased brain', grant reference 104943/Z/14/Z.

Data Availability Statement

The python code for the machine learning implementation proposed in this manuscript can be found in the fml_pMRI repository https://zenodo.org/badge/latestdoi/189416118. We do not have ethical consent to make the *in-vivo* datasets acquired for this study publically available.

References

Alsop, D.C., Detre, J.A., Golay, X., Gunther, M., Hendrikse, J., Hernandez-Garcia, L., Lu, H., MacIntosh, B.J., Parkes, L.M., Smits, M., van Osch, M.J., Wang, D.J., Wong, E.C., Zaharchuk, G., 2015. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med 73, 102-116.

Bernier, J.L., Ortega, J., Ros, E., Rojas, I., Prieto, A., 1999. A new measurement of noise immunity and generalization ability for MLPs. Int J Neural Syst 9, 511-521.

Blockley, N.P., Griffeth, V.E., Germuska, M.A., Bulte, D.P., Buxton, R.B., 2013. An analysis of the use of hyperoxia for measuring venous cerebral blood volume: comparison of the existing method with a new analysis approach. Neuroimage 72, 33-40.

Breiman, L., 2001. Random forests. Machine Learning 45, 5-32.

Bulte, D.P., Kelly, M., Germuska, M., Xie, J., Chappell, M.A., Okell, T.W., Bright, M.G., Jezzard, P., 2012. Quantitative measurement of cerebral physiology using respiratory-calibrated MRI. Neuroimage 60, 582-591.

Chappell, M.A., MacIntosh, B.J., Donahue, M.J., Gunther, M., Jezzard, P., Woolrich, M.W., 2010. Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI. Magn Reson Med 63, 1357-1365.

Coles, J.P., Fryer, T.D., Bradley, P.G., Nortje, J., Smielewski, P., Rice, K., Clark, J.C., Pickard, J.D., Menon, D.K., 2006. Intersubject variability and reproducibility of 15O PET studies. J Cereb Blood Flow Metab 26, 48-57.

De Vis, J.B., Petersen, E.T., Bhogal, A., Hartkamp, N.S., Klijn, C.J., Kappelle, L.J., Hendrikse, J., 2015. Calibrated MRI to evaluate cerebral hemodynamics in patients with an internal carotid artery occlusion. J Cereb Blood Flow Metab 35, 1015-1023.

Frackowiak, R.S., Herold, S., Petty, R.K., Morgan-Hughes, J.A., 1988. The cerebral metabolism of glucose and oxygen measured with positron tomography in patients with mitochondrial diseases. Brain 111 (Pt 5), 1009-1024.

Frackowiak, R.S., Lenzi, G.L., Jones, T., Heather, J.D., 1980. Quantitative measurement of regional cerebral blood flow and oxygen metabolism in man using 15O and positron emission tomography: theory, procedure, and normal values. J Comput Assist Tomogr 4, 727-736.

Frau-Pascual, A., Vincent, T., Sloboda, J., Ciuciu, P., Forbes, F., 2014. Physiologically Informed Bayesian Analysis of ASL fMRI Data. Bayesian and Graphical Models for Biomedical Imaging. Lecture Notes in Computer Science 8677.

Gauthier, C.J., Desjardins-Crepeau, L., Madjar, C., Bherer, L., Hoge, R.D., 2012. Absolute quantification of resting oxygen metabolism and metabolic reactivity during functional activation using QUO2 MRI. Neuroimage 63, 1353-1363.

Gauthier, C.J., Hoge, R.D., 2013. A generalized procedure for calibrated MRI incorporating hyperoxia and hypercapnia. Hum Brain Mapp 34, 1053-1069.

Germuska, M., Chandler, H.L., Stickland, R.C., Foster, C., Fasano, F., Okell, T.W., Steventon, J., Tomassini, V., Murphy, K., Wise, R.G., 2019. Dual-calibrated fMRI measurement of absolute cerebral metabolic rate of oxygen consumption and effective oxygen diffusivity. Neuroimage 184, 717-728.

411 Germuska, M., Merola, A., Murphy, K., Babic, A., Richmond, L., Khot, S., Hall, J.E., Wise, R.G.,
412 2016. A forward modelling approach for the estimation of oxygen extraction fraction by calibrated
413 fMRI. Neuroimage 139, 313-323.

414 Germuska, M., Wise, R.G., 2019. Calibrated fMRI for mapping absolute CMRO2: Practicalities and
415 prospects. Neuroimage 187, 145-153.

416 Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63, 3-42.

417 Gjedde, A., 2002. Cerebral blood flow change in arterial hypoxemia is consistent with negligible
418 oxygen tension in brain mitochondria. Neuroimage 17, 1876-1881.

419 Hayashi, T., Watabe, H., Kudomi, N., Kim, K.M., Enmi, J., Hayashida, K., Iida, H., 2003. A
420 theoretical model of oxygen delivery and metabolism for physiologic interpretation of quantitative
421 cerebral blood flow and metabolic rate of oxygen. J Cereb Blood Flow Metab 23, 1314-1323.

422 Hertel, L., Phan, H., Mertins, A., 2016. Comparing Time and Frequency Domain for Audio Event
423 Recognition Using Deep Learning. 2016 International Joint Conference on Neural Networks (Ijcnn),
424 3407-3411.

425 Hoge, R.D., 2012. Calibrated FMRI. Neuroimage 62, 930-937.

426 Hyder, F., Herman, P., Bailey, C.J., Moller, A., Globinsky, R., Fulbright, R.K., Rothman, D.L.,
427 Gjedde, A., 2016. Uniform distributions of glucose oxidation and oxygen extraction in gray matter of
428 normal human brain: No evidence of regional differences of aerobic glycolysis. J Cereb Blood Flow
429 Metab 36, 903-916.

430 Ibaraki, M., Shinohara, Y., Nakamura, K., Miura, S., Kinoshita, F., Kinoshita, T., 2010.
431 Interindividual variations of cerebral blood flow, oxygen delivery, and metabolism in relation to
432 hemoglobin concentration measured by positron emission tomography in humans. J Cereb Blood
433 Flow Metab 30, 1296-1305.

434 Ishii, K., Kitagaki, H., Kono, M., Mori, E., 1996. Decreased medial temporal oxygen metabolism in
435 Alzheimer's disease shown by PET. J Nucl Med 37, 1159-1165.

436 Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl.
437 Neuroimage 62, 782-790.

438 Lajoie, I., Nugent, S., Debacker, C., Dyson, K., Tancredi, F.B., Badhwar, A., Belleville, S.,
439 Deschaintre, Y., Bellec, P., Doyon, J., Bocti, C., Gauthier, S., Arnold, D., Kergoat, M.J., Chertkow,
440 H., Monchi, O., Hoge, R.D., 2017. Application of calibrated fMRI in Alzheimer's disease.
441 Neuroimage Clin 15, 348-358.

442 Lebrun-Grandie, P., Baron, J.C., Soussaline, F., Loch'h, C., Sastre, J., Bousser, M.G., 1983. Coupling
443 between regional blood flow and oxygen utilization in the normal human brain. A study with positron
444 tomography and oxygen 15. Arch Neurol 40, 230-236.

445 Leenders, K.L., Perani, D., Lammertsma, A.A., Heather, J.D., Buckingham, P., Healy, M.J., Gibbs,
446 J.M., Wise, R.J., Hatazawa, J., Herold, S., et al., 1990. Cerebral blood flow, blood volume and
447 oxygen utilization. Normal values and effect of age. Brain 113 (Pt 1), 27-47.

448 Merola, A., Germuska, M.A., Murphy, K., Wise, R.G., 2018. Assessing the repeatability of absolute
449 CMRO2, OEF and haemodynamic measurements from calibrated fMRI. Neuroimage 173, 113-126.

450 Merola, A., Germuska, M.A., Warnert, E.A., Richmond, L., Helme, D., Khot, S., Murphy, K.,
451 Rogers, P.J., Hall, J.E., Wise, R.G., 2017. Mapping the pharmacological modulation of brain oxygen

metabolism: The effects of caffeine on absolute CMRO2 measured using dual calibrated fMRI. Neuroimage 155, 331-343.

Merola, A., Murphy, K., Stone, A.J., Germuska, M.A., Griffeth, V.E.M., Blockley, N.P., Buxton, R.B., Wise, R.G., 2016. Measurement of oxygen extraction fraction (OEF): An optimized BOLD signal model for use with hypercapnic and hyperoxic calibration. Neuroimage 129, 159-174.

Mesejo, P., Saillet, S., David, O., Bénar, C., Warnking, J.M., Forbes, F., 2015. Estimating Biophysical Parameters from BOLD Signals through Evolutionary-Based Optimization. Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science 9350.

Miles, K.A., Williams, R.E., 2008. Warburg revisited: imaging tumour blood flow and metabolism. Cancer Imaging 8, 81-86.

Mintun, M.A., Lundstrom, B.N., Snyder, A.Z., Vlassenko, A.G., Shulman, G.L., Raichle, M.E., 2001. Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. Proc Natl Acad Sci U S A 98, 6859-6864.

Morris, E.A., Juttukonda, M.R., Lee, C.A., Patel, N.J., Pruthi, S., Donahue, M.J., Jordan, L.C., 2018. Elevated brain oxygen extraction fraction in preterm newborns with anemia measured using noninvasive MRI. J Perinatol 38, 1636-1643.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Powers, W.J., Videen, T.O., Markham, J., Walter, V., Perlmutter, J.S., 2011. Metabolic control of resting hemispheric cerebral blood flow is oxidative, not glycolytic. J Cereb Blood Flow Metab 31, 1223-1228.

Reeve, H.W.J., Brown, G., 2018. Diversity and degrees of freedom in regression ensembles. Neurocomputing 298, 55-68.

Ruijters, D., Thevenaz, P., 2012. GPU Prefilter for Accurate Cubic B-spline Interpolation. Computer Journal 55, 15-20.

Safar, P., 1988. Resuscitation from clinical death: pathophysiologic limits and therapeutic potentials. Crit Care Med 16, 923-941.

Scheinberg, P., Stead, E.A., 1949. The Cerebral Blood Flow in Male Subjects as Measured by the Nitrous Oxide Technique. Normal Values for Blood Flow, Oxygen Utilization, Glucose Utilization, and Peripheral Resistance, with Observations on the Effect of Tilting and Anxiety. J Clin Invest 28, 1163-1171.

Sollich, P., Krogh, A., 1996. Learning with ensembles: How over-fitting can be useful. Advances in Neural Information Processing Systems 8 8, 190-196.

Vafaee, M.S., Gjedde, A., 2000. Model of blood-brain transfer of oxygen explains nonlinear flow-metabolism coupling during stimulation of visual cortex. J Cereb Blood Flow Metab 20, 747-754.

Verweij, B.H., Amelink, G.J., Muizelaar, J.P., 2007. Current concepts of cerebral oxygen transport and energy metabolism after severe traumatic brain injury. Prog Brain Res 161, 111-124.

Wise, R.G., Harris, A.D., Stone, A.J., Murphy, K., 2013. Measurement of OEF and absolute CMRO2: MRI-based methods using interleaved and combined hypercapnia and hyperoxia. Neuroimage 83, 135-147.

494

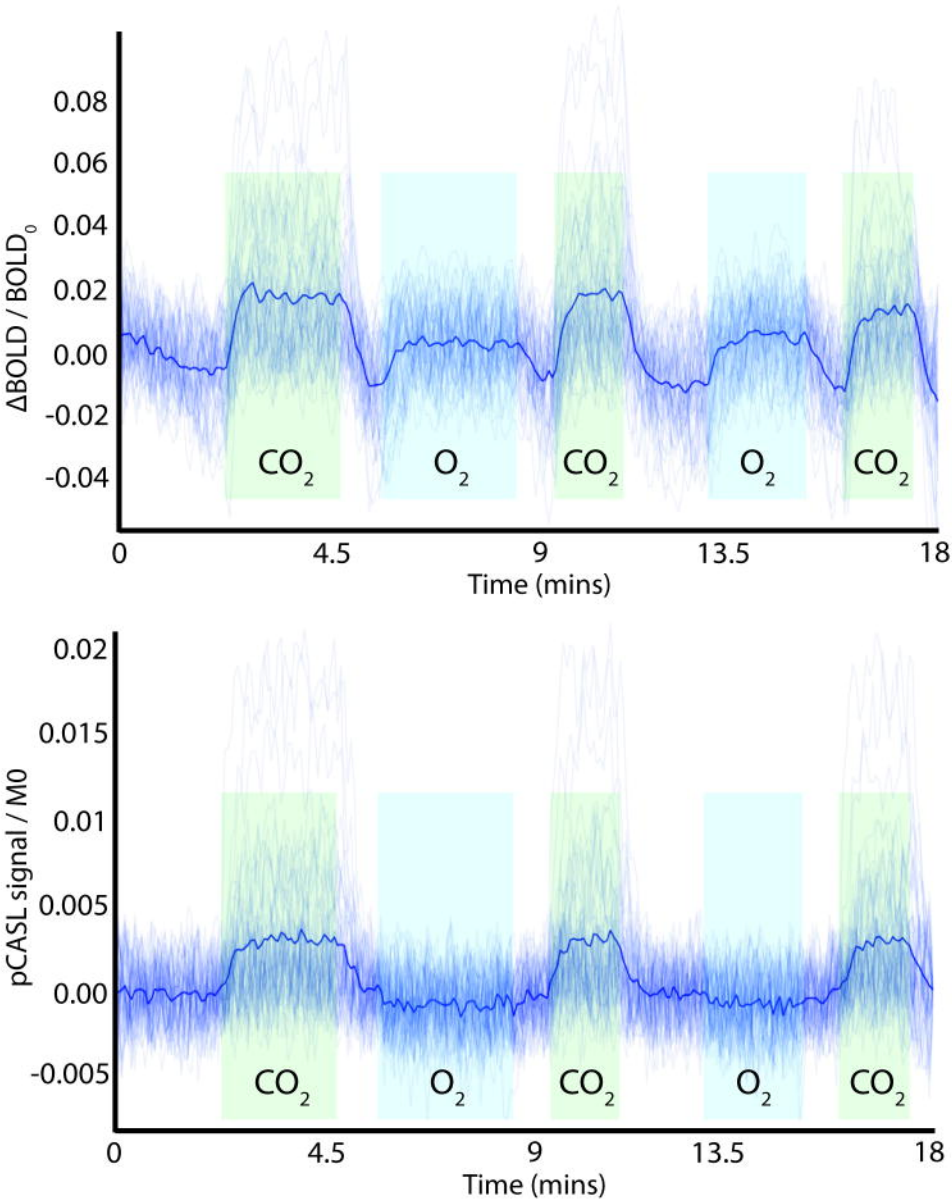

Figure 1. Example of simulated time-domain data (BOLD and ASL) with added noise and variation in physiological parameters, showing periods of hypercapnic (green) and hyperoxic (light blue) stimuli. The dark blue line represents the mean time-course over the example time series. Note the pCASL signal is normalised by the equilibrium magnetisation of arterial blood (M0) and has the baseline signal subtracted for display purposes.

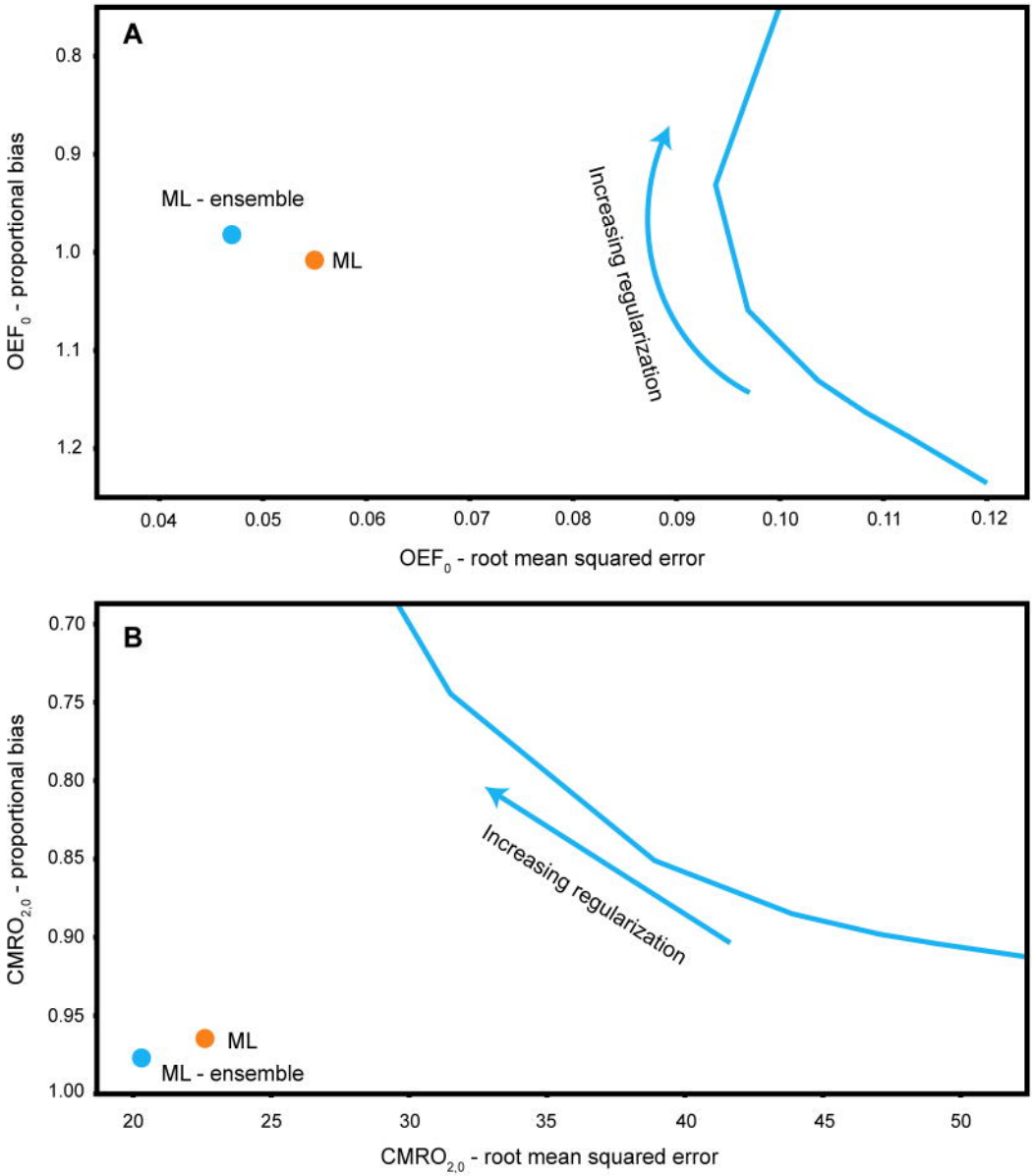

Figure 2. Root mean squared error and proportional bias in OEF_0 (A) and $CMRO_{2,0}$ (B) estimates for each analysis method fitting to simulated data (5000 simulations). Solid blue line plots the error and bias for increasing regularization weighting for the regularized non-linear least squares analysis.

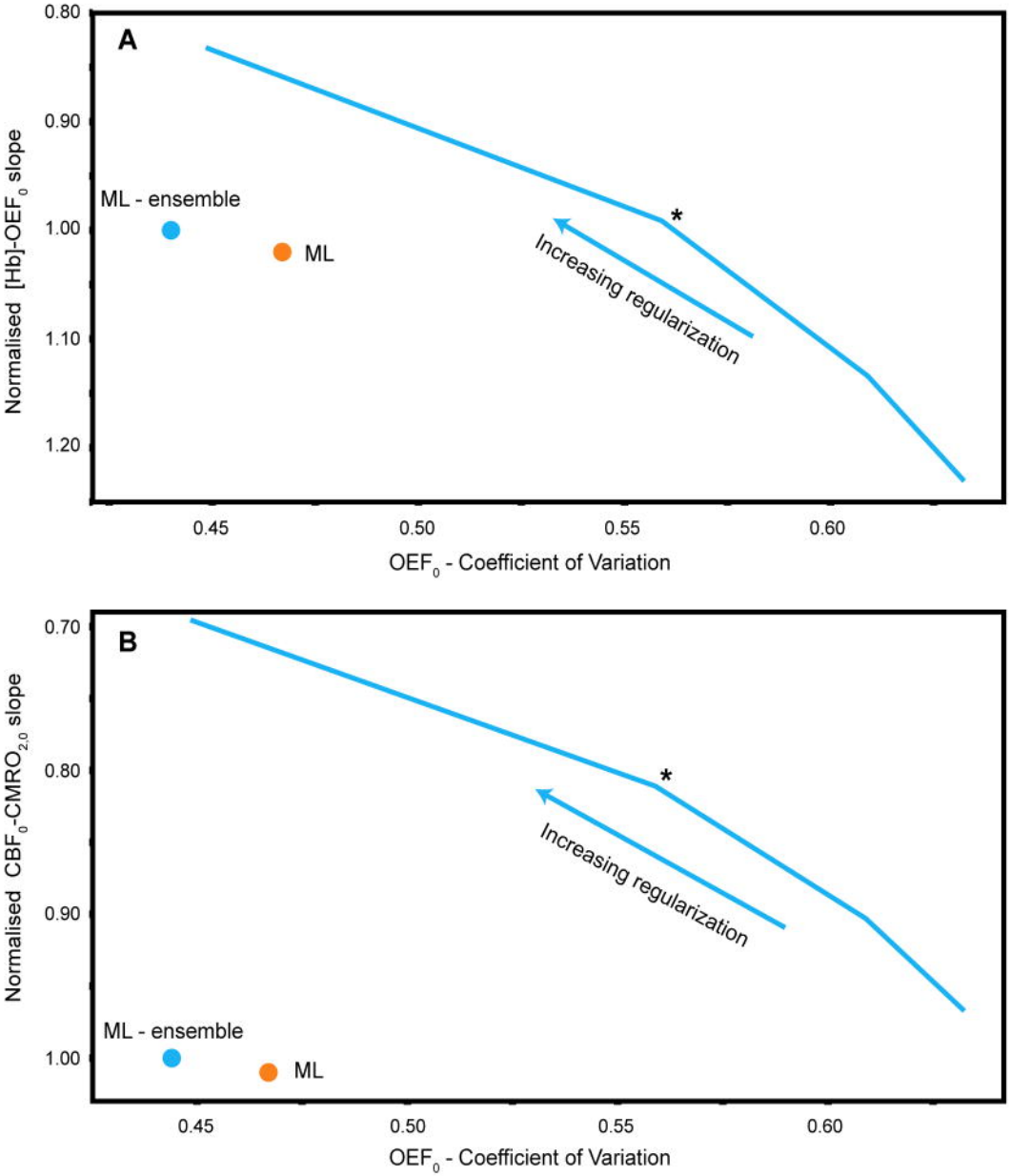

Figure 3. A. Coefficient of variation of grey matter OEF_0 estimates versus slope of [Hb]-OEF_0 relationship for each analysis method (rNLS fitting evaluated with increasing levels of regularization). The [Hb]-OEF0 slope has been normalised by the ML ensemble estimate of the [Hb]-OEF_0 slope. B. Coefficient of variation of grey matter OEF_0 estimates versus the slope of the CBF-$CMRO_2$ relationship, normalised by the ML (ensemble) estimate of the CBF-$CMRO_2$ slope. Solid blue line plots the coefficient of variation against the slope for increasing regularization weighting for the regularized non-linear least squares analysis. The asterisk indicates the chosen level of regularization for subsequent analysis/comparisons.

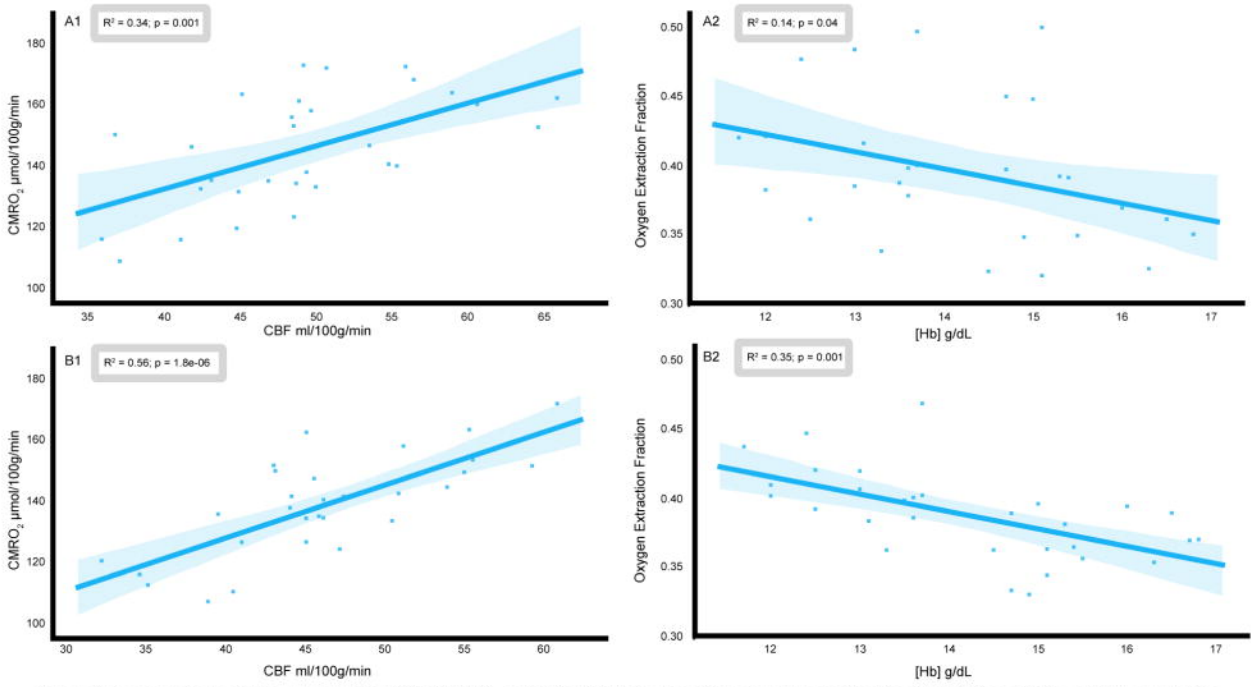

Figure 4. Scatter plots of the grey matter CBF-$CMRO_2$ and [Hb]-OEF relationships observed with the rNLS (A1 and A2) and ML ensemble (B1 and B2) methods across the 30 healthy volunteers.

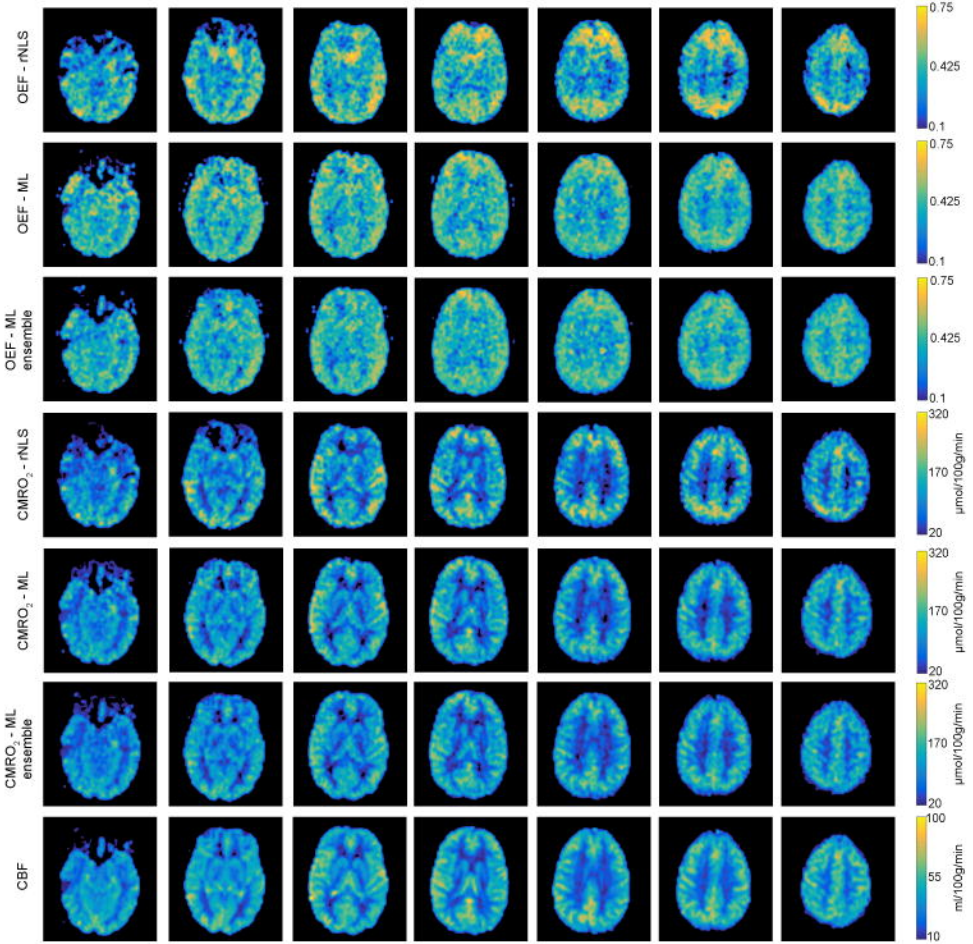

Figure 5. Example parameter maps (CBF0, OEF0, and CMRO2,0) from a single subject for each analysis method. Machine learning estimates of OEF0 are more uniform than regularized non-linear least squares estimates. Using an ensemble of MLP networks further reduces the spatial variation in OEF0 estimates, particularly in regions near air-tissue susceptibility interfaces.