Application of a ^{1}H Brain MRS Benchmark Dataset to Deep Learning for Out-of-Voxel Artifacts

Aaron T. Gudmundson[a,b], Christopher W. Davies-Jenkins[a,b], İpek Özdemir[a,b], Saipavitra Murali-Manohar[a,b], Helge J. Zöllner[a,b], Yulu Song[a,b], Kathleen E. Hupfeld[a,b], Alfons Schnitzler[c], Georg Oeltzschner[a,b], Craig E. L. Stark[d], and Richard A. E. Edden[a,b]

[a] Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine

[b] F. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute

[c] Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

[d] Department of Neurobiology and Behavior, University of California, Irvine, Irvine, CA.

Word Count: 5432

Number of Figures: 6

Number of Tables: 1

*Corresponding Author: Richard A. E. Edden, Ph.D.

Department of Radiology and Radiological Science, Johns Hopkins School of Medicine

600 N. Wolfe St, Park 338,

Baltimore, MD 21287

E-mail Address: edden@jhu.edu

Abstract:

Neural networks are potentially valuable for many of the challenges associated with MRS data. The purpose of this manuscript is to describe the AGNOSTIC dataset, which contains 259,200 synthetic ^{1}H MRS examples for training and testing neural networks. AGNOSTIC was created using 270 basis sets that were simulated across 18 field strengths and 15 echo times. The synthetic examples were produced to resemble *in vivo* brain data with combinations of metabolite, macromolecule, residual water signals, and noise. To demonstrate the utility, we apply AGNOSTIC to train two Convolutional Neural Networks (CNNs) to address out-of-voxel (OOV) echoes. A Detection Network was trained to identify the point-wise presence of OOV echoes, providing proof of concept for real-time detection. A Prediction Network was trained to reconstruct OOV echoes, allowing subtraction during post-processing. Complex OOV signals were mixed into 85% of synthetic examples to train two separate CNNs for the detection and prediction of OOV signals. AGNOSTIC is available through Dryad and all Python 3 code is available through GitHub. The Detection network was shown to perform well, identifying 95% of OOV echoes. Traditional modeling of these detected OOV signals was evaluated and may prove to be an effective method during linear-combination modeling. The Prediction Network greatly reduces OOV echoes within FIDs and achieved a median $\log_{10}$ normed-MSE of –1.79, an improvement of almost two orders of magnitude.

Keywords: Magnetic Resonance Spectroscopy, Synthetic Data, Simulation, Deep Learning, Out-of-voxel Artifacts, Human Brain

Abbreviations: ^{1}H, proton; 2HG, β-hydroxyglutarate; Ace, acetate; AGNOSTIC, adaptable generalized neural-network open-source spectroscopy training dataset of individual components; Ala, alanine; Asc, ascorbate; Asp, aspartate; ATP, adenosine triphosphate; bHB, β-hydroxybutyrate; Cho, choline-containing compounds; Cit, citrate; CNN, Convolutional Neural Networks; Cr, creatine; Cys, Cysteine; DL, deep learning; EA, Ethanolamine; EtOH, ethanol; FID, free induction decay; FWHM, full-width half-maximum; GABA, gamma-aminobutyric acid; Glc, glucose; Gln, glutamine; Glu, glutamate; Glx, sum of glutamate and glutamine; Gly, glycine; Glyce, glycerol; GM, gray matter; GPC, glycerophosphocholine; GSH, glutathione; H2O, water; HCar, homocarnosine; Hist, histamine; His, histidine; ISMRM, international society for magnetic resonance in medicine; Lac, lactate; LASER, localization by adiabatic selective refocusing; MEGA, Mescher-Garwood; mI, myo-inositol; MM, macromolecule; MRS, magnetic resonance spectroscopy; MSE, mean-squared error; NAA, N-acetylaspartate; NAAG, N-acetyl-aspartyl-glutamate; OOV, out-of-voxel; PCho, phosphocholine; PCr, phosphocreatine; PE, phosphoethanolamine; Phenyl, phenylalanine; PRESS, point resolved spectroscopy; ReLu, rectified linear unit; Ser, serine; sI, scyllo-inositol; sLASER, semi-adiabatic localization by adiabatic selective refocusing; SPECIAL, spin echo full intensity acquired localiezed; STEAM, stimulated echo acquisition mode; SNR, signal-to-noise ratio; T_2, spin-spin relaxation time; Tau, taurine; tCho, sum of choline-containing metabolites; tCr, sum of creatine and phosphocreatine; TE, echo-time; Thr; threonine; tNAA, sum of N-acetyl-aspartate and N-acetyl-aspartyl-glutamate; Trp, Tryptophan; Tyr, Tyrosine; Val, Valine

1. Introduction:

Proton (^{1}H) magnetic resonance spectroscopy (MRS) non-invasively measures levels of endogenous neurometabolites. MRS-visible metabolites are present at millimolar concentrations in the brain, yielding detectable signals with relatively low signal-to-noise ratio (SNR) which mutually overlap. *In vivo* spectra suffer from several artifacts that complicate modeling and interpretation of the data, including eddy current effects and out-of-voxel (OOV) echoes (Kreis, 2004). While there is some degree of standardization and consensus around pre-processing, modeling, and quantification of MRS data (Maudsley et al., 2021; Near et al., 2021; Öz et al., 2021; Wilson et al., 2019), this is an evolving field without a single ideal solution due to the complexity of the problem, and which therefore is likely to benefit from recent advances in machine learning.

Deep learning (DL) uses a network consisting of a series of computational layers to process information (Lecun et al., 2015). Iterative training allows features of the data to be identified and weighted to estimate a final function which predicts a desired output based on a given input (Goodfellow et al., 2016). Supervised learning involves training the network based on a pre-defined target, associating ground-truth parameters with each input. An extensive, balanced, and diverse dataset is preferred to increase the generalizability of the DL outcome. High-dimensional data, such as medical images or time series, are demonstrated to be the most beneficial set of data for several computer vision tasks, such as classification, registration, segmentation, reconstruction, and object detection (Gassenmaier, Küstner, et al., 2021; Lundervold & Lundervold, 2019).

DL has been developed for MRS data as a proof-of-concept in many applications, including metabolite quantification (Chandler et al., 2019; Hatami et al., 2018; H. H. Lee & Kim, 2019, 2020; Rizzo et al., 2023; Shamaei et al., 2023; Zhang & Shen, 2023), signal separation (Li et al., 2020), phase and frequency correction (Ma et al., 2022; Shamaei et al., 2023; Tapper et al., 2021), reconstruction of missing data (H. Lee et al., 2020), accelerated post-processing (Gurbani et al., 2019; Iqbal et al., 2021), denoising (Chen et al., 2023; Dziadosz et al., 2023; Lam et al., 2020), super-resolution (Gassenmaier, Afat, et al., 2021; Iqbal et al., 2019), artifact removal (Gurbani et al., 2018; Kyathanahally et al., 2018), and anomaly detection (Jang et al., 2021). Despite the potential, these methods have yet to be shown to generalize outside of small datasets with a single fixed acquisition protocol. Whereas 'classical' methods for post-processing are often driven by an

understanding of the problem to be solved, and therefore can often be applied broadly, deep learning methods cannot be assumed to function well outside of the specific datasets used for training and testing. Broadly applicable deep learning methods will only arise from broad training and testing. A key barrier is the lack of a generalized benchmark dataset for training and testing, to play the role that MNIST, ImageNet, and COCO have played in the field of Computer Vision (Fei-Fei et al., 2010; Li Deng, 2012; T.-Y. Lin et al., 2014). Such a dataset lowers the barrier to entry for neural network development in MRS and facilitates performance comparisons between models. The Synthetic Data Working Group of the MRS study group of the International Society for Magnetic Resonance in Medicine's Synthetic Data Working Group has recently highlighted the MRS community's need for such a resource. The ultimate goal of this work is to bridge the gap from the synthetic to the in vivo domain, including the additional domain-shift to clinical data.

OOV echoes, which represent a subset of the artifacts often referred to as 'spurious' or 'ghost' echoes (Kreis, 2004), are a substantial issue for *in vivo* MRS, and an under-studied potential DL application. MRS voxel localization is achieved via a combination of RF pulses and magnetic field gradients, with the intended coherence transfer pathway selected both by phase cycling and dephasing "crusher" gradient scheme (Bodenhausen, 2011). OOV signals arise from gradient echoes – signals from outside the shimmed voxel of interest are refocused by evolution in local field gradients that are either inherent (from air-tissue-bone interfaces) or arising from second-order shim terms (Starck et al., 2009). Therefore, brain regions close to air cavities (e.g., medial prefrontal cortex) or which require stronger shim gradients (e.g., thalamus, hippocampus, etc.) most commonly exhibit OOV artifacts (Starck et al., 2009). OOV echoes seldom occur at the time of the primary echo, so they manifest in the spectrum as broad peaks with strong first-order phase "ripple" that can obscure metabolite resonances. While acquisition strategies can mitigate OOV echoes to some extent, by careful consideration of crusher schemes or voxel orientation (Ernst & Chang, 1996; Landheer & Juchem, 2019; Song et al., 2023), post-processing strategies remain valuable where complete elimination is not possible.

This manuscript develops **A**daptable **G**eneralized **N**eural-Network **O**pen-source **S**pectroscopy **T**raining dataset of **I**ndividual **C**omponents (AGNOSTIC), a dataset consisting of 259,200 synthetic MRS examples. AGNOSTIC spans a range of field strengths, echo times, and clinical profiles, representing metabolite signals, macromolecule (MM) background signals,

136 residual water signals, and Gaussian noise as separate components. To date, DL applications to
137 MRS have relied upon narrow in-house-generated training datasets that limit the generalizability
138 of the solutions developed and comparisons between tools; AGNOSTIC is proposed as a
139 benchmark dataset to fill this gap. In order to demonstrate the utility of this resource, we then
140 illustrate a specific augmentation of the AGNOSTIC dataset to train neural networks for the
141 detection and prediction of OOV echoes.

2. Methods:

2.1. AGNOSTIC Synthetic Dataset

The parameter space that AGNOSTIC spans is deliberately broad, comprising: 18 field strengths; 15 echo times; broad distributions of metabolite, MM, and water amplitudes; and densely sampled time-domain to allow down-sampling. Calculations were carried out using an in-house and open-source Python 3 (Van Rossum & Drake, 2009) programming script using NumPy (Harris et al., 2020). The decision to use in-house software was motivated by needing the flexibility to simulate basis sets that could be manipulated on a spin-by-spin basis which could, for instance, allow for different spins within the same metabolite to have different relaxation rates (e.g., $Cr_{3.9}$ and $Cr_{3.0}$). The dataset is structured as a zipped NumPy archive file (.npz) and can be opened as a Python 3 dictionary object. This zipped NumPy file contains complex-valued NumPy arrays of time-domain (4096 timepoints) data corresponding to the metabolite, macromolecule, water, and noise components which can be combined in different ways depending on the application. For instance, a denoising model may want to target the combined metabolite, MM, and water signal without noise. Within the file, all the acquisition parameters (field strength, echo time, spectral width, etc.), simulation parameters (signal to noise, full-width half-max, concentrations, T_2 relaxation, etc.), and data augmentation options are specified as detailed below.

2.1.1. Basis Set Simulation:

Metabolite spectra are based upon density-matrix-simulated basis functions (Blum, 1981; Fano, 1957; Farrar, 1990; O. W. Sørensen et al., 1984). A total of 270 basis sets were created across 18 field strengths (1.4 T – 3.1 T in steps of 0.1 T) and 15 echo times (10 ms – 80 ms in steps of 5 ms). The Point RESolved Spectroscopy (PRESS) pulse sequence (Bottomley, 1982) was simulated using ideal pulses with TE1 = TE2. The simulated "acquisition window" was started immediately after the last pulse to generate points before the echo. Each metabolite basis was output as an N x 16684 NumPy array, where N is the number of spins for a given metabolite and 16684 is the fixed length of complex time points (300 points before the echo maximum, with an appropriate padding number of zeros and followed by the simulated pre-echo signal, and 16384 points after the echo). The simulated spectral width, centered on 4.7 ppm, was 63.62 ppm for all field strengths (e.g., 8 kHz at 3 T; 4 kHz at 1.5 T). By subsampling the intentionally long time-domain points in the basis set, we can achieve a series of different spectral widths within

the ranges commonly seen for *in vivo* experiments without the need to re-simulate the signal with different dwell times.

39 brain metabolite basis functions were simulated: Adenosine Triphosphate (ATP); Acetate (Ace); Alanine (Ala); Ascorbate (Asc); Aspartate (Asp); β-hydroxybutyrate (bHB); β-hydroxyglutarate (2HG); Citrate (Cit); Cysteine (Cys); Ethanolamine (EA); Ethanol (EtOH); Creatine (Cr); y-Amino-Butyric Acid (GABA); Glucose (Glc); Glutamine (Gln); Glutamate (Glu); Glycerophosphocholine (GPC); Glutathione (GSH); Glycerol (Glyce); Glycine (Gly); Water (H_2O); Homocarnosine (HCar); Histamine (Hist); Histidine (His); Lactate (Lac); Myo-Inositol (mI); N-Acetyl-Aspartate (NAA); N-Acetyl-Aspartate-Glutamate (NAAG); Phenylalanine (Phenyl); Phosphocholine (PCho); Phosphocreatine (PCr); Phosphoethanolamine (PE); Scyllo-Inositol (sI); Serine (Ser); Taurine (Tau); Threonine (Thr); Tryptophan (Trp); Tyrosine (Tyr); and Valine (Val). GABA was separately simulated using two different spin-system enumerations (Govindaraju et al., 2000; Near et al., 2012). Both α-glucose and β-glucose were simulated.

2.1.2. Assembly of Metabolite Component:

Individual metabolite basis functions were linearly combined to give a metabolite spectral component, weighted by metabolite concentrations sampled from distributions defined by our recent meta-analysis (Gudmundson et al., 2023), including both healthy and clinical cohort ranges. From the full basis sets, 22 metabolites were selected which had defined concentration ranges available in a recent meta-analysis that collated results from nearly 500 MRS papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Gudmundson et al., 2023; Moher et al., 2009; Page et al., 2021). One isomer of GABA (either the definition from (Govindaraju et al., 2000) or (Near et al., 2012)) and Glucose (α or β) were randomly chosen with equal probability for each example. Concentrations were selected with equal probability from a range defined by ±2.5 standard deviations from the meta-analysis mean of each cohort (Gudmundson et al., 2023) and are provided in supplemental tables 1 and 2.

T_2* relaxation decay of time-domain data was simulated with an exponential and Gaussian component to produce a Voigt lineshape (Marshall et al., 1997) in the frequency domain. The exponential component represents the pure T_2 arising from dipole-dipole interactions, paramagnetic interaction, etc., while the Gaussian component represents the

transverse dephasing from diffusion and exchange of spins in an inhomogeneous field (Koch et al., 2009; Marshall et al., 1997; Michaeli et al., 2002; Yablonskiy & Haacke, 1994). While pure T_2 is understood to be field-independent (Bloembergen et al., 1948; Carr & Purcell, 1954; Held et al., 1973; Michaeli et al., 2002), the dominant Gaussian decay (Marshall et al., 1997) increases with increasing static field strength and is attributed to greater microscopic (Michaeli et al., 2002) and macroscopic (Juchem et al., 2021; Tkáč et al., 2001) susceptibility gradients. Here, the pure Lorentzian T_2 component is based upon the relaxation times at 1.5 T from a relaxation meta-regression (Gudmundson et al., 2023), which are assumed to be the least impacted by susceptibility gradients that scale with B_0 (Bloembergen et al., 1948; De Graaf et al., 2006; Michaeli et al., 2002). Once the Lorentzian T_2 component was applied, the additional T_2* contributions were modeled by applying appropriate amounts of Gaussian broadening, to achieve a frequency-domain full-width half-maximum (*FWHM*) linewidth of the NAA singlet between 3 Hz and 18 Hz with a uniform distribution. A small amount of jitter (between 20 s^{-2} and 100 s^{-2}) was added to the Gaussian decay rate so that each metabolite would undergo a similar, but not identical, amount of Gaussian decay to better replicate the variability observed for *in vivo* data.

2.1.3. Macromolecular Component:

Fourteen MM signals were modeled at: 0.92 ppm; 1.21 ppm; 1.39 ppm; 1.67 ppm; 2.04 ppm; 2.26 ppm; 2.56 ppm; 2.70 ppm; 2.99 ppm; 3.21 ppm; 3.62 ppm; 3.75 ppm; 3.86 ppm; and 4.03 ppm (Cudalbu et al., 2021; Giapitzakis et al., 2018). MM chemical shifts were jittered by ± 0.03 ppm to both account for observed differences in MM designations and provide further dataset augmentation. Each MM signal was simulated as a singlet with exponential decay rate sampled uniformly from a range specified by literature of MM T_2 time constants (Murali-Manohar et al., 2020) and additional Gaussian decay to reach published linewidths (Giapitzakis et al., 2018; Murali-Manohar et al., 2020). MM amplitudes were sampled uniformly from within published ranges (Giapitzakis et al., 2018; Murali-Manohar et al., 2020).

2.1.4. Noise Component:

Noise was generated from a normal distribution, with independent random real and imaginary points. The noise was scaled such that the signal-to-noise ratio of the NAA singlet (SNR_{NAA} was defined, following Experts' Consensus (Öz et al., 2021), by *NAA height divided by the standard deviation of the noise*) was uniformly sampled between 5 and 80. The noise amplitude values are also stored within the archive file.

2.1.5. Residual Water Component:

The residual water basis signal was simulated as a composite signal (of up to five components). In order to simulate varying degrees of water suppression, the residual water signal was modeled by between 0 and 5 unique Voigt-shaped signals with variable ppm locations, phases, and amplitudes, based on the approach of (L. Lin et al., 2019). The ranges for these parameters are listed in Table 1. The final water component was scaled to be between 1× and 20× the maximum value of the frequency-domain metabolite spectrum. The water components used, along with their corresponding parametrizations, are stored within the NumPy archive file.

2.1.6. Frequency and Phase Shifts:

Within the NumPy archive file, a frequency shift, zero-order phase shift, and first-order phase shift are specified for each entry in the dataset, but not applied to the time-domain components. Frequency shifts were sampled uniformly from the range −0.313 ppm to +0.313 ppm. Zero-order phase shifts were sampled uniformly from the range −180 degrees to +180 degrees. First-order phase shifts were sampled uniformly from the range −19.5 degrees to +19.5 degrees per ppm. Users may choose to omit phase and frequency shifts, use the provided shifts, or specify their own.

2.2. Exemplar Application to AGNOSTIC: Machine Learning for Out-Of-Voxel Artifacts:

The primary motivation for the AGNOSTIC dataset is as a training resource for the development of processing, modeling, and analysis tools for MRS. Synthetic spectra with known ground truths are valuable in a range of applications, from the development and validation of traditional linear combination modeling algorithms to training DL models.

In order to demonstrate the utility of the dataset, an exemplar application is presented, in which the AGNOSTIC dataset is supplemented by simulated artifacts (in this case out-of-voxel OOV echoes) and used to train DL models to detect and predict the artifact signals. The AGNOSTIC dataset was developed as building blocks which can be combined to train a variety of different models. A strength of this dataset is that custom user-defined components can be utilized. We demonstrate this point here by building an OOV dataset to train and evaluate a DL model to identify and suppress OOV artifacts.

2.2.1. Simulation of Out-Of-Voxel Echoes:

OOV artifacts were defined as complex time-domain signals with a time point (τ_{OOV}), width (W_{OOV}), frequency (ω_{OOV}), phase (Φ_{OOV}) and amplitude (a_{OOV}) as shown in Figure 1. τ_{OOV} describes the timepoint of the top of the OOV echo and was sampled randomly from a uniform distribution between 10 ms and 400 ms. W_{OOV} describes the Gaussian decay rate and was sampled randomly from a uniform distribution between 500 s^{-2} and 8000 s^{-2}, resulting in a FWHM echo duration between 18 ms and 74 ms. ω_{OOV} describes the offset in the frequency domain, and was sampled randomly from a uniform distribution in order to produce OOVs that occur between 1 ppm and 4 ppm. a_{OOV} was sampled randomly from a uniform distribution to produce OOV echoes with an amplitude between 0.1% and 20% of the maximum time domain point. Φ_{OOV} was sampled uniformly between 0 degrees to 360 degrees.

$$Out\ of\ Voxel\ Echo\ =\ a_{OOV}\left(e^{-W_{OOV}(t-\tau_{OOV})^2}\right)\left(e^{-i\omega t}\right)\left(e^{-i\Phi}\right) \quad [1]$$

2.2.2. Integration of OOV Echoes into AGNOSTIC for the Training Dataset:

To build the OOV echoes dataset, we combined metabolite, water, MM, and noise components from the AGNOSTIC dataset. We then added OOV signals to 85% of the dataset and a complex zeros array in the remaining 15%. In total there were 180,000 examples used for network training, 1,800 examples used for validation, and 7,200 examples used for testing. Finally, we applied the included frequency and phase shifts specified within the AGNOSTIC dataset. The network input consisted of the combined metabolite, water, MM, noise, and OOV signals as a complex time-domain signal. This input was normalized so that the absolute maximum among the real and imaginary values was 1. Finally, training data were converted to a TensorFlow Dataset (Abadi et al., 2015).

2.2.3. Detection Network:

The first exemplar network is designed to detect OOV echoes within time-domain data by identifying the points in the spectra that have been contaminated by OOV echoes. This Detection Network is a fully Convolutional Neural Network (CNN) designed using TensorFlow2 with Keras (Chollet & others, 2015) in a Python 3 environment. The network consists of contracting encoding layers and expanding decoding layers with a total of 1.543 million parameters, as shown in Figure 2. Each layer was initialized (kernel_initializer) with

"he_normal" (He et al., 2015). Each convolutional layer (except the output layer which uses a sigmoid activation) includes batch normalization and a leaky rectified linear unit (ReLu) activation function (Maas et al., 2013)Click or tap here to enter text.. A kernel size of 3 (3 x 2 before collapsing the real/imaginary dimension and 3 x 1 afterward) was used for each convolutional layer. The network is designed to receive a time-domain input signal and return a binary mask of the same size as the input with ones placed in OOV-detected regions and zeros elsewhere. A ground-truth binary mask was determined as the 5% level of the maximum amplitude of the Gaussian OOV envelope located at the central peak. For training, the input and output of this network is a 60 x 2048 x 2 x 1 tensor, where 60 is the batch size, 2048 is the number of time points, 2 is the real/imaginary dimension, and 1 is the channel dimension.

The Dice coefficient (Carass et al., 2020; Dice, 1945; T. Sørensen, 1948) of the overlap between the network output and the correct binary OOV location vector was used as a training loss function, calculated as 2x the intersection divided by the union plus 1; where 1 was used to avoid division by 0. The Adam (Kingma & Ba, 2015) optimizer was used with a fixed learning rate of 0.0003. Success on the validation set was evaluated every 7,200 steps, at which time the network weights were saved if the validation loss improved. The final model that was selected had the smallest validation loss after 72 epochs. Training took approximately 2.5 hours and was performed on an 8 GB NVIDIA GeForce RTX 3070 GPU. A clustering algorithm was applied to the final network output, which zeroed any group of time points in which the network detected OOV echo that was smaller than 5 consecutive time points, to dampen spurious output. A cluster size of 5 was selected empirically to ensure detection of the narrow echoes while eliminating any false positives.

2.2.4. Modeling:

Modeling of the OOV echoes was performed as an optimization problem and solved with SciPy (Virtanen et al., 2020) minimization routines. Here, the non-gradient Powell (Powell, 1964, 1994) optimizer was used to determine the five OOV parameters (τ_{OOV}, W_{OOV}, ω_{OOV}, Φ_{OOV}, and a_{OOV}), minimizing the mean squared error (MSE) between the model and the data within the time window identified by the Detection Network. Initial values for τ_{OOV}, W_{OOV}, and the a_{OOV} are inferred from the Detection Network's output center timepoint, the detection duration, and the standard deviation of the target signal within the detected region.

Optimization was performed as three sequential optimization steps performed one after another. The first optimization is used to determine τ_{OOV}, W_{OOV}, and the a_{OOV} by minimizing the MSE between the absolute values of the model and the data (i.e., removing frequency and phase from the model) in the time domain. The second optimization determines ω_{OOV} by minimizing between the absolute values of the model and the data in the frequency domain. The third optimization refines the values determined by optimizations 1 and 2 and determines Φ_{OOV} by complex optimization in the time domain.

2.2.5. Prediction Network:

The second exemplar network is designed to predict the OOV echoes found within time-domain data. This prediction network is also a fully CNN designed using TensorFlow2 with Keras in a Python 3 environment, with the same architecture as the Detection Network (as shown in Figure 2). As such, the input and output of this network is also 60 x 2048 x 2 x 1 tensor, where 60 is the batch size, 2048 is the number of time points, 2 is the real/imaginary dimension, and 1 is the channel dimension. The network is designed to receive a time-domain input signal containing a combination of the ground-truth time-domain signal and the OOV artifact and return a time-domain output signal that only contains the OOV signal, amplified 10x. This amplification serves to focus the training on the OOV echo by non-uniformly (due to the OOV echo's non-linear decay) concentrating the network towards the center-most points of the OOV echoes to effectively center and reconstruct the predicted OOV echo on the τ_{OOV} with the correct W_{OOV}.

For training, a weighted mean squared error (weighting the timepoints within the ground-truth OOV mask uniformly by 10) was used as a loss function with an ADAM (Kingma & Ba, 2015) optimizer and a fixed learning rate of 0.0003. Success on the validation set was evaluated every 7,200 steps at which time the network weights were saved if the validation loss improved. The final model that was selected had the smallest validation loss after 72 epochs. Training took approximately 2.5 hours and was performed on an 8 GB NVIDIA GeForce RTX 3070 GPU.

2.2.6 Evaluating the Performance of Networks and Modeling:

In the final testing set, OOV artifacts were present in 6,137 of the total 7,200 examples (85.2%). The Detection Network was evaluated using the Dice coefficient (Carass et al., 2020; Dice, 1945; Powell, 1964), the overlap between the ground-truth binary OOV mask and the

cluster-thresholded network output. As well as computing global success, the dependence of detection success on various attributes of the OOV echo and the underlying spectrum were also investigated.

Both modeling and the prediction network return a pure OOV signal, and in both cases, the MSE between the prediction/model and the ground-truth OOV echo is used for evaluation. If the ground-truth echo datapoints are E_i and the model or echo prediction is M_i, we calculate the fractional remaining OOV amplitude as:

$$\text{Fractional OOV Remaining} = \frac{\sum|M_i - E_i|^2}{\sum|E_i|^2} \quad [2]$$

where the bars represent the complex amplitude. The sum is taken over the ground-truth range of the OOV echo. In order to visualize a wide range of success and failure, we take the $\log_{10}$ of this quantity for plotting (i.e., a $\log_{10}$ value of 1 is no change, anything positive is a manipulation that is worse than doing nothing, and a negative value show the order of magnitude of improvement). Note that E_i is the ground-truth echo signal, not the signal from which the echo is being removed which also contains metabolite, macromolecule, and noise components.

The timing of the OOV was found to be a key parameter determining the success of detection and prediction, and as a result, the evaluation metrics were calculated for the following time-bins (based on the known value of t_{OOV}): 10-20 ms; 20-40 ms; 40-60 ms; 60-80 ms; 80-120 ms; 120-200 ms; 200-300 ms; 300-400 ms.

2.2.7 *In Vivo* Proof-of-Principle

As a proof-of-principle demonstration of this exemplar use of the AGNOSTIC dataset, the network was applied to 256 transients of *in vivo* data, selected because they contain prominent OOV echoes and were excluded during quality assessment in a recent study (Zöllner et al., 2023). These data were collected on a 2.89 T Siemens scanner using the MEGA-PRESS (Mescher et al., 1996, 1998) pulse sequence with a TE of 68 ms and TR of 1.75 s, and a spectral width of 2.4 kHz. Note that this challenges the generality of the training because the network has never seen data acquired at 2.89 T, nor at 2.4 kHz spectral width, nor at TE 68 ms, nor with MEGA-Editing, nor with actual real RF pulses. Raw data from a 25 x 25 x 25 mm^3 voxel in the

cerebellum were loaded and coil combined in Osprey (Oeltzschner et al., 2020). Time-domain data were saved as a MATLAB (The MathWorks Inc., 2022) .mat file and loaded as a Python 3 object using SciPy. The data were normalized (as above with training data) to be used as input for the neural networks.

One challenge of *in vivo* data (and the reason that this network demonstration focuses substantially on synthetic data) is that no ground truth is available. Therefore, the degree of success in removing OOV echo signals from time-domain data D_i is:

$$\text{Fractional Reduction in standard deviation} = 1 - \frac{\sigma(D_i - M_i)}{\sigma(D_i)} \quad [3]$$

where s denotes the standard deviation. Note that, in contrast to the metric used for synthetic data in Equation 1, only D_i is available, not the ground truth E_i, which substantially changes the ceiling of success. It is still expected that substantial signal variance remains after OOV removal, since D_i contains metabolite signals and noise. The range over which this standard deviation is calculated is the 50% level of the normalized histogram of the detection network's output across the 256 transients. Note that this metric is an imperfect response to the absence of ground-truth knowledge for in vivo data, predicated on the assumption that subtracting out OOV signal reduces the standard deviation of the time-domain signal.

3. Results:

3.1. AGNOSTIC Synthetic Dataset:

The AGNOSTIC dataset contains 259,200 examples, consisting of 960 examples from each of the eighteen field strengths and fifteen echo times (i.e., 960x18x15=259,200). A representative set of ten spectra are shown in Figure 3, illustrating the diversity of field strengths, TEs, SNR, and linewidth within the dataset.

One challenge with making this dataset available is its size — 75 GB — but we do make it freely available on Dryad (DOI: 10.7280/D1RX1T). The basis sets from which these are constructed are more manageable — 9 GB — and can also be accessed through Dryad (DOI: 10.7280/D1RX1T). Code for generating the AGNOSTIC dataset locally is available at: https://github.com/agudmundson/agnostic.

3.2. Exemplar Application to AGNOSTIC: Machine Learning for Out-Of-Voxel Artifacts:

3.2.1. Detection Network:

Of the 6,137 examples where OOV artifacts were present, the Detection Network correctly identified 5,827 (94.9%) with a median Dice score of 0.974 (0.941–0.985 interquartile range) and missed 310 (5.05%) with a Dice score of 0.00. In the 1063 examples that did not include OOV artifacts, the network correctly ignored 912 (85.8%) and falsely detected OOV echoes in 151 (14.2%). Figure 4 shows the Detection Network's output for a synthetic OOV-corrupted example.

Analysis of the factors that determined success indicated that the time at which OOV signals occur is most critical. Therefore, OOV echoes were further broken down into eight time-bins, and the Dice score plotted in Figure 5. The median Dice scores — 0.165, 0.858, 0.892, 0.934, 0.960, 0.974, 0.978, and 0.978 — are poor in the first bin and improve thereafter. Note that these bins are not spaced equally to emphasize poor performance extremely early. The number of examples in each bin is 161, 289, 282, 329, 622, 1256, 1565, and 1633, respectively.

3.2.2. Modeling

The modeling optimization converged in 5,824 of the 5,827 examples where the detection network detected OOV artifacts and provided initial values. Across this subset of the examples, the modeling achieved a median $\log_{10}$(fractional OOV remaining) of −2.19 (−2.90 – −1.19 inter-

quartile range), i.e., a median reduction of more than two orders of magnitude. Figure 5 shows the resulting model for a synthetic OOV-corrupted example.

These values — broken down into 8 time-bins — are shown in Figure 5. The median $\log_{10}$(fractional OOV remaining) decreases across the time bins: 1.663, 1.324, 0.680, 0.223, −1.586, −2.276, −2.491, and −2.567.

3.2.3. Prediction Network:

In the 6,137 examples where OOV artifacts were present, the prediction network achieved a median $\log_{10}$ normed-MSE of −1.79 (−2.21 – −1.11 inter-quartile range). In the 5,824 examples where OOV artifacts were successfully modeled, the prediction network achieved a median $\log_{10}$ normed-MSE of −1.85 (−2.24 – −1.24 inter-quartile range). Figure 5 shows the Prediction Network's output for a synthetic OOV-corrupted example.

OOVs were further broken down into 8 time-bins (Figure 5) early — the number of examples in each bin is 86, 226, 261, 312, 592, 1208, 1538, 1601. The median $\log_{10}$(fractional OOV remaining) decreases across the time bins: −0.207, −0.583, −0.862, −1.250, −1.577, −1.878, −2.005, and −2.052.

3.2.4. *In Vivo* Proof-of-Principle:

The Detection Network identified an OOV in 243 of 256 transients (94.9%). In these 243 OOV-detected transients, the modeling achieved a median reduction in standard deviation of 71.0 % (60.2 -75.3% inter-quartile range). The Prediction Network achieved a median reduction in standard deviation of 69.65% (66.33 %/72.7 % inter-quartile range) in this subset. In the full set of 256 transients, the Prediction Network achieved a median 69.4 % (65.3 – 72.6 % inter-quartile range) reduction in standard deviation. The standard deviation of the noise floor was found to account for a median of 10.3% (9.35–11.6 % inter-quartile range) of the standard deviation of signal within the time window for the 256 averages. A representative *in vivo* example is shown in Figure 6.

4. Discussion:

AGNOSTIC is a benchmark MRS dataset for training and evaluating performance across various models. In order to make these synthetic data representative of in vivo brain MRS datasets, a total of 22 brain metabolites and 14 MM peaks were simulated within 270 basis sets, spanning field strengths from 1.4 T to 3.1 T and TEs from 10 to 80 ms. Parameterized water residual and noise were included. SNR and linewidths were assigned at random, independent of B_0 or TE. The broad span of the dataset is key in training networks that generalize. While AGNOSTIC is broad in these dimensions, it does only represent simulated data for PRESS (Bottomley, 1982) acquisitions, and may benefit from expansion to include other pulse sequences, such as STEAM (Frahm et al., 1987), SPECIAL (Mekle et al., 2009; Mlynarik et al., 2006), LASER (Garwood & DelaBarre, 2001), and semi-LASER (Scheenen, Heerschap, et al., 2008; Scheenen, Klomp, et al., 2008), and edited versions including MEGA (Mescher et al., 1996, 1998) and Hadamard-encoded (Chan et al., 2016, 2019; Oeltzschner et al., 2019; Saleh et al., 2016) schemes. AGNOSTIC is limited by simulations that used ideal pulses, a calculated trade-off to emphasize generalizability across field strength, echo time, and spectral width, and thus fail to capture effects associated with spatially heterogeneous coupling evolution. The extent to which these limitations matter will depend on the applications that AGNOSTIC synthetic data are being used for.

The Detection network was highly successful, identifying 94.9% of the testing set where OOV artifacts were present. The precise value of this success metric is obviously impacted by the parameters of the OOVs – a later minimum OOV time would tend to increase performance, and earlier would degrade it. It is noteworthy that, although the training datasets never contained more than one OOV echo, the detection and prediction networks were able to handle more than one OOV echo in vivo data, presumably because CNNs operate locally within the FID. It is also encouraging that the networks generalized well to the *in vivo* data (Figure 6), which was collected with unseen acquisition parameters, i.e., edited MEGA-PRESS (Mescher et al., 1996, 1998) data acquired at 2.89 T with a TE of 68 ms, and 2.4 kHz spectral width. While it is reasonable to believe that networks trained using AGNOSTIC will generalize well with *in vivo* clinical data, future work will need to evaluate performance for clinical applications.

In the exemplar OOV application, the success of the networks depended heavily on the timing of the OOV signal. The earliest OOV echoes were most challenging, unsurprisingly since such signals are broad Gaussian resonances that are indistinguishable from within-voxel MM and

baseline signals. Indeed, the only feature that differentiates OOV signals from other broad components of the model is timing. It is conceptually helpful to consider this in the Fourier domain, even though all network processing is performed in the time domain. In the frequency domain, a mismatch between the echo-top and the acquisition start is represented as a first-order phase error of the signal associated with that echo. Where insufficient first-order phase exists to be represented within the linewidth of the signal in question, (which in the time domain corresponds to substantial truncation of the lefthand side of the echo), the network struggles to identify OOV signals.

In the context of this study, modeling and prediction are treated as two alternative approaches to OOV characterization. For early OOV signals, the modeling approach tended to mis-attribute non-artifact signal as OOV signal, a result that the metric scored as worse than no intervention. The median performance of the Prediction network, even for very early OOV signals, was close to zero. Both modeling and prediction performance improve as the OOV moves later in the acquired signal, with modeling improving faster than the network, and performing better than prediction beyond 120 ms. This strong performance of the model at least in part reflects the exact match between the generative model of the synthetic OOV artifacts and the model that is being used to extract them. More moderate performance might be expected for real in vivo examples – but the same may also be true for networks which have been trained with the same synthetic data and may have learned specifically to identify OOV signals that have a Gaussian kernel.

One key difference between most DL applications and applications in MRS, is the strict requirement to preserve amplitude fidelity in network outputs. A common approach to artifacts in DL is to return an artifact-free version of the network input. In contrast, the approach taken here is to return the artifact, which has the following benefits: it avoids networks over-learning the formulaic pattern of typical spectra; it reduces the impact of the lack of sequence diversity within the AGNOSTIC dataset; and it is less likely to impact the amplitudes of metabolite signals.

The ultimate goal of this work is to extract metabolite levels from MRS data that are not impacted by OOV artifacts. This problem can be addressed at several points: either by not acquiring data that contain OOV artifacts; by removing OOV artifacts post-acquisition; and by incorporating appropriate OOV model components into quantification model so that the impact of OOV is minimized. While the work presented here focuses primarily on the second context, it raises important potential applications in the other contexts. One motivator for developing the Detection network is the possibility of real-time deployment during sequence acquisition to trigger

sequence changes when OOV artifacts are detected. The modeling applied here was time-restricted to a given window and ignored other components of the spectrum, but demonstrates potential for future integration within a full linear-combination model.

5. Conclusion:

In conclusion, we have presented the AGNOSTIC benchmark dataset which can be used for training and testing brain-specific ^{1}H MRS deep learning models. This large synthetic dataset is open-source and encompasses a range of field strengths, TEs, and dwell times to ensure networks are robust to a variety of *in vivo* data acquisitions protocols. Using this dataset, we have demonstrated an exemplar use case to develop CNNs to detect and predict out-of-voxel artifacts.

Acknowledgments:

This work has been supported by The Henry L. Guenther Foundation, Sonderforschungsbereich (SFB) 974 (TP B07) of the German Research foundation, and the National Institute of Health, grants T32 AG00096, R00 AG062230, R21 EB033516, R01 EB016089, R01 EB023963, K00AG068440, P30 AG066519, R21 AG053040, R01 AG076942, P30 AG066519 and P41 EB031771.

CRediT authorship contribution statement:

Aaron T. Gudmundson: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing-original draft, Writing-Review & Editing. **Christopher W. Davies-Jenkins:** Data Curation, Resources, Writing-original draft, Writing-Review & Editing. **İpek Özdemir:** Data Curation, Writing-original draft, Resources, Writing-Review & Editing. **Saipavitra Murali-Manohar:** Data Curation, Writing-original draft, Writing-Review & Editing. **Helge J. Zöllner:** Data Curation, Resources, Writing-original draft, Writing-Review & Editing. **Yulu Song:** Data Curation, Resources, Writing-original draft, Writing-Review & Editing. **Kathleen E. Hupfeld:** Writing-original draft, Writing-Review & Editing. **Alfons Schnitzler:** Data Curation, Resources, Writing-Review & Editing. **Georg Oeltzschner:** Conceptualization, Supervision, Writing-original draft, Writing-Review & Editing. **Craig Stark:** Conceptualization, Funding acquisition, Project Administration, Resources, Supervision, Writing-original draft, Writing-Review & Editing. **Richard A.E. Edden:** Conceptualization, Funding acquisition, Project Administration, Resources, Supervision, Visualization, Writing-original draft, Writing-Review & Editing.

References:

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). *Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems*. https://doi.org/10.5281/zenodo.4724125

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *ArXiv Preprint ArXiv:1803.08375*.

Bloembergen, N., Purcell, E. M., & Pound, R. V. (1948). Relaxation effects in nuclear magnetic resonance absorption. *Physical Review*, *73*(7), 679–712. https://doi.org/10.1103/PhysRev.73.679

Blum, K. (1981). *Density Matrix Theory and Applications* (1st ed.). Springer US. https://doi.org/10.1007/978-1-4615-6808-7

Bodenhausen, G. (2011). Reflections of pathways: A short perspective on 'Selection of coherence transfer pathways in NMR pulse experiments.' *Journal of Magnetic Resonance*, *213*(2), 295–297. https://doi.org/10.1016/j.jmr.2011.08.004

Bottomley, P. A. (1982). *Selective volume method for performing localized NMR spectroscopy*. *19*. https://patents.google.com/patent/US4480228A/en

Carass, A., Roy, S., Gherman, A., Reinhold, J. C., Jesson, A., Arbel, T., Maier, O., Handels, H., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Pham, D. L., Crainiceanu, C. M., Calabresi, P. A., Prince, J. L., Roncal, W. R. G., Shinohara, R. T., & Oguz, I. (2020). Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis. *Scientific Reports*, *10*(1), 1–19. https://doi.org/10.1038/s41598-020-64803-w

Carr, H. Y., & Purcell, E. M. (1954). *Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments*t*.

Chan, K. L., Edden, R. A. E., & Barker, P. B. (2019). *Simultaneous editing of GABA and GSH with Hadamard - encoded MR spectroscopic imaging*. *July 2018*, 21–32. https://doi.org/10.1002/mrm.27702

Chan, K. L., Puts, N. A. J., Schär, M., Barker, P. B., & Edden, R. A. E. (2016). HERMES: Hadamard encoding and reconstruction of MEGA-edited spectroscopy. *Magnetic Resonance in Medicine*, *76*(1), 11–19. https://doi.org/10.1002/mrm.26233

Chandler, M., Jenkins, C., Shermer, S. M., & Langbein, F. C. (2019). *MRSNet: Metabolite Quantification from Edited Magnetic Resonance Spectra With Convolutional Neural Networks*. 1–12. http://arxiv.org/abs/1909.03836

Chen, D., Hu, W., Liu, H., Zhou, Y., Qiu, T., Huang, Y., Wang, Z., Lin, M., Lin, L., Wu, Z., Wang, J., Chen, H., Chen, X., Yan, G., Guo, D., Lin, J., & Qu, X. (2023). Magnetic Resonance Spectroscopy Deep Learning Denoising Using Few In Vivo Data. *IEEE Transactions on Computational Imaging*, 1–12. https://doi.org/10.1109/TCI.2023.3267623

Chollet, F., & others. (2015). *Keras*.

578 Cudalbu, C., Behar, K. L., Bhattacharyya, P. K., Bogner, W., Borbath, T., de Graaf, R. A., Gruetter, R.,
579 Henning, A., Juchem, C., Kreis, R., Lee, P., Lei, H., Marjańska, M., Mekle, R., Murali-Manohar, S.,
580 Považan, M., Rackayová, V., Simicic, D., Slotboom, J., … Mlynárik, V. (2021). Contribution of
581 macromolecules to brain 1H MR spectra: Experts' consensus recommendations. *NMR in*
582 *Biomedicine*, *34*(5), 1–24. https://doi.org/10.1002/nbm.4393

583 De Graaf, R. A., Brown, P. B., McIntyre, S., Nixon, T. W., Behar, K. L., & Rothman, D. L. (2006). High
584 magnetic field water and metabolite proton T1 and T 2 relaxation in rat brain in vivo. *Magnetic*
585 *Resonance in Medicine*, *56*(2), 386–394. https://doi.org/10.1002/mrm.20946

586 Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species Author (s): Lee R .
587 Dice Published by : Ecological Society of America Stable URL :
588 http://www.jstor.org/stable/1932409. *Ecology*, *26*(3), 297–302.

589 Dziadosz, M., Rizzo, R., Kyathanahally, S. P., & Kreis, R. (2023). Denoising single MR spectra by deep
590 learning: Miracle or mirage? *Magnetic Resonance in Medicine*. https://doi.org/10.1002/mrm.29762

591 Ernst, T., & Chang, L. (1996). Elimination of artifacts in short echo time1H MR spectroscopy of the frontal
592 lobe. *Magnetic Resonance in Medicine*, *36*(3), 462–468. https://doi.org/10.1002/mrm.1910360320

593 Fano, U. (1957). Description of States in Quantum Mechanics by Density Matrix and Operator
594 Techniques. *Reviews of Modern Physics*, *29*(1), 74–93. https://doi.org/10.1103/RevModPhys.29.74

595 Farrar, T. (1990). Density matrices in NMR spectroscopy: Part I. *Concepts in Magnetic Resonance*, *2*, 1–
596 12.

597 Fei-Fei, L., Deng, J., & Li, K. (2010). ImageNet: Constructing a large-scale image database. *Journal of*
598 *Vision*, *9*(8), 1037–1037. https://doi.org/10.1167/9.8.1037

599 Frahm, J., Merboldt, K. D., & Hänicke, W. (1987). Localized proton spectroscopy using stimulated echoes.
600 *Journal of Magnetic Resonance (1969)*, *72*(3), 502–508. https://doi.org/10.1016/0022-
601 2364(87)90154-5

602 Garwood, M., & DelaBarre, L. (2001). The return of the frequency sweep: Designing adiabatic pulses for
603 contemporary NMR. *Journal of Magnetic Resonance*. https://doi.org/10.1006/jmre.2001.2340

604 Gassenmaier, S., Afat, S., Nickel, D., Kannengiesser, S., Herrmann, J., Hoffmann, R., & Othman, A. E.
605 (2021). Application of a Novel Iterative Denoising and Image Enhancement Technique in T1-
606 Weighted Precontrast and Postcontrast Gradient Echo Imaging of the Abdomen. *Investigative*
607 *Radiology*, *56*(5), 328–334. https://doi.org/10.1097/RLI.0000000000000746

608 Gassenmaier, S., Küstner, T., Nickel, D., Herrmann, J., Hoffmann, R., Almansour, H., Afat, S., Nikolaou, K.,
609 & Othman, A. E. (2021). Deep Learning Applications in Magnetic Resonance Imaging: Has the
610 Future Become Present? *Diagnostics*, *11*(12), 2181. https://doi.org/10.3390/diagnostics11122181

611 Giapitzakis, I. A., Avdievich, N., & Henning, A. (2018). Characterization of macromolecular baseline of
612 human brain using metabolite cycled semi-LASER at 9.4T. *Magnetic Resonance in Medicine*, *80*(2),
613 462–473. https://doi.org/10.1002/mrm.27070

614 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.
615 http://www.deeplearningbook.org/

616 Govindaraju, V., Young, K., & Maudsley, A. A. (2000). Proton NMR chemical shifts and coupling constants
617 for brain metabolites. *NMR in Biomedicine*, *13*(3), 129–153. https://doi.org/10.1002/1099-
618 1492(200005)13:3<129::AID-NBM619>3.0.CO;2-V

619 Gudmundson, A. T., Koo, A., Virovka, A., Amirault, A. L., Soo, M., Cho, J. H., Oeltzschner, G., Edden, R. A.
620 E., & Stark, C. E. L. (2023). Meta-analysis and open-source database for in vivo brain Magnetic
621 Resonance spectroscopy in health and disease. *Analytical Biochemistry*, *676*, 115227.
622 https://doi.org/10.1016/j.ab.2023.115227

623 Gurbani, S. S., Schreibmann, E., Maudsley, A. A., Cordova, J. S., Soher, B. J., Poptani, H., Verma, G.,
624 Barker, P. B., Shim, H., & Cooper, L. A. D. (2018). A convolutional neural network to filter artifacts in
625 spectroscopic MRI. *Magnetic Resonance in Medicine*, *80*(5), 1765–1775.
626 https://doi.org/10.1002/mrm.27166

627 Gurbani, S. S., Sheriff, S., Maudsley, A. A., Shim, H., & Cooper, L. A. D. (2019). Incorporation of a spectral
628 model in a convolutional neural network for accelerated spectral fitting. *Magnetic Resonance in*
629 *Medicine*, *81*(5), 3346–3357. https://doi.org/10.1002/mrm.27641

630 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E.,
631 Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M.,
632 Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming
633 with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

634 Hatami, N., Sdika, M., & Ratiney, H. (2018). Magnetic resonance spectroscopy quantification using deep
635 learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*
636 *Intelligence and Lecture Notes in Bioinformatics)*, *11070 LNCS*, 467–475.
637 https://doi.org/10.1007/978-3-030-00928-1_53

638 He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level
639 Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision*
640 *(ICCV)*, *2015 Inter*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

641 Held, G., Noack, F., Pollak, V., & Melton, B. (1973). Protonenspinrelaxation und Wasserbeweglichkeit in
642 Muskelgewebe / Proton Spin Relaxation and Mobility of Water in Muscle Tissue. *Zeitschrift Für*
643 *Naturforschung C*, *28*(1–2), 59–62. https://doi.org/10.1515/znc-1973-1-209

644 Iqbal, Z., Nguyen, D., Hangel, G., Motyka, S., Bogner, W., & Jiang, S. (2019). Super-Resolution 1H
645 Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning. *Frontiers in Oncology*, *9*.
646 https://doi.org/10.3389/fonc.2019.01010

647 Iqbal, Z., Nguyen, D., Thomas, M. A., & Jiang, S. (2021). Deep learning can accelerate and quantify
648 simulated localized correlated spectroscopy. *Scientific Reports*, *11*(1), 8727.
649 https://doi.org/10.1038/s41598-021-88158-y

650 Jang, J., Lee, H. H., Park, J.-A., & Kim, H. (2021). Unsupervised anomaly detection using generative
651 adversarial networks in 1H-MRS of the brain. *Journal of Magnetic Resonance*, *325*, 106936.
652 https://doi.org/10.1016/j.jmr.2021.106936

653 Juchem, C., Cudalbu, C., Graaf, R. A., Gruetter, R., Henning, A., Hetherington, H. P., & Boer, V. O. (2021).
654 B 0 shimming for in vivo magnetic resonance spectroscopy: Experts' consensus recommendations.
655 *NMR in Biomedicine*, *34*(5), 1–20. https://doi.org/10.1002/nbm.4350

656 Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International*
657 *Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

658 Koch, K. M., Rothman, D. L., & de Graaf, R. A. (2009). Optimization of static magnetic field homogeneity
659 in the human and animal brain in vivo. In *Progress in Nuclear Magnetic Resonance Spectroscopy*
660 (Vol. 54, Issue 2, pp. 69–96). https://doi.org/10.1016/j.pnmrs.2008.04.001

661 Kreis, R. (2004). Issues of spectral quality in clinical1H-magnetic resonance spectroscopy and a gallery of
662 artifacts. *NMR in Biomedicine*, *17*(6), 361–381. https://doi.org/10.1002/nbm.891

663 Kyathanahally, S. P., Döring, A., & Kreis, R. (2018). Deep learning approaches for detection and removal
664 of ghosting artifacts in MR spectroscopy. *Magnetic Resonance in Medicine*, *80*(3), 851–863.
665 https://doi.org/10.1002/mrm.27096

666 Lam, F., Li, Y., & Peng, X. (2020). Constrained Magnetic Resonance Spectroscopic Imaging by Learning
667 Nonlinear Low-Dimensional Models. *IEEE Transactions on Medical Imaging*, *39*(3), 545–555.
668 https://doi.org/10.1109/TMI.2019.2930586

669 Landheer, K., & Juchem, C. (2019). Dephasing optimization through coherence order pathway selection
670 (DOTCOPS) for improved crusher schemes in MR spectroscopy. *Magnetic Resonance in Medicine*,
671 *81*(4), 2209–2222. https://doi.org/10.1002/mrm.27587

672 Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.
673 https://doi.org/10.1038/nature14539

674 Lee, H. H., & Kim, H. (2019). Intact metabolite spectrum mining by deep learning in proton magnetic
675 resonance spectroscopy of the brain. *Magnetic Resonance in Medicine*, *82*(1), 33–48.
676 https://doi.org/10.1002/mrm.27727

677 Lee, H. H., & Kim, H. (2020). Deep learning-based target metabolite isolation and big data-driven
678 measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain.
679 *Magnetic Resonance in Medicine*, *84*(4), 1689–1706. https://doi.org/10.1002/mrm.28234

680 Lee, H., Lee, H. H., & Kim, H. (2020). Reconstruction of spectra from truncated free induction decays by
681 deep learning in proton magnetic resonance spectroscopy. *Magnetic Resonance in Medicine*, *84*(2),
682 559–568. https://doi.org/10.1002/mrm.28164

683 Li Deng. (2012). The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best
684 of the Web]. *IEEE Signal Processing Magazine*, *29*(6), 141–142.
685 https://doi.org/10.1109/MSP.2012.2211477

686 Li, Y., Wang, Z., & Lam, F. (2020). Separation of Metabolite and Macromolecule Signals for 1 H-Mrsi
687 Using Learned Nonlinear Models. *Proceedings - International Symposium on Biomedical Imaging*,
688 *2020-April*, 1725–1728. https://doi.org/10.1109/ISBI45749.2020.9098365

689 Lin, L., Považan, M., Berrington, A., Chen, Z., & Barker, P. B. (2019). Water removal in MR spectroscopic
690 imaging with L2 regularization. *Magnetic Resonance in Medicine*, *82*(4), 1278–1287.
691 https://doi.org/10.1002/mrm.27824

692 Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L.,
693 & Dollár, P. (2014). *Microsoft COCO: Common Objects in Context*. http://arxiv.org/abs/1405.0312

694 Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on
695 MRI. *Zeitschrift Für Medizinische Physik*, *29*(2), 102–127.
696 https://doi.org/10.1016/j.zemedi.2018.11.002

697 Ma, D. J., Le, H. A. M., Ye, Y., Laine, A. F., Lieberman, J. A., Rothman, D. L., Small, S. A., & Guo, J. (2022).
698 MR spectroscopy frequency and phase correction using convolutional neural networks. *Magnetic*
699 *Resonance in Medicine*, *87*(4), 1700–1710. https://doi.org/10.1002/mrm.29103

700 Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic
701 models. *In ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, *28*.

702 Marshall, I., Higinbotham, J., Bruce, S., & Freise, A. (1997). Use of Voigt lineshape for quantification of in
703 vivo 1H spectra. *Magnetic Resonance in Medicine*, *37*(5), 651–657.
704 https://doi.org/10.1002/mrm.1910370504

705 Maudsley, A. A., Andronesi, O. C., Barker, P. B., Bizzi, A., Bogner, W., Henning, A., Nelson, S. J., Posse, S.,
706 Shungu, D. C., & Soher, B. J. (2021). Advanced magnetic resonance spectroscopic neuroimaging:
707 Experts' consensus recommendations. *NMR in Biomedicine*, *34*(5), 1–22.
708 https://doi.org/10.1002/nbm.4309

709 Mekle, R., Mlynárik, V., Gambarota, G., Hergt, M., Krueger, G., & Gruetter, R. (2009). MR spectroscopy of
710 the human brain with enhanced signal intensity at ultrashort echo times on a clinical platform at 3T
711 and 7T. *Magnetic Resonance in Medicine*, *61*(6), 1279–1285. https://doi.org/10.1002/mrm.21961

712 Mescher, M., Merkle, H., Kirsch, J., Garwood, M., & Gruetter, R. (1998). Simultaneous in vivo spectral
713 editing and water suppression. *NMR in Biomedicine*, *11*(6), 266–272.
714 https://doi.org/10.1002/(SICI)1099-1492(199810)11:6<266::AID-NBM530>3.0.CO;2-J

715 Mescher, M., Tannus, A., O'Neil Johnson, M., & Garwood, M. (1996). Solvent suppression using selective
716 echo dephasing. *Journal of Magnetic Resonance - Series A*, *123*(2), 226–229.
717 https://doi.org/10.1006/jmra.1996.0242

718 Michaeli, S., Garwood, M., Zhu, X.-H., DelaBarre, L., Andersen, P., Adriany, G., Merkle, H., Ugurbil, K., &
719 Chen, W. (2002). ProtonT2 relaxation study of water, N-acetylaspartate, and creatine in human
720 brain using Hahn and Carr-Purcell spin echoes at 4T and 7T. *Magnetic Resonance in Medicine*,
721 *47*(4), 629–633. https://doi.org/10.1002/mrm.10135

722 Mlynarik, V., Gambarota, G., Frenkel, H., & Gruetter, R. (2006). Localized short-echo-time proton MR
723 spectroscopy with full signal-intensity acquisition. *MAGNETIC RESONANCE IN MEDICINE*, *56*(5),
724 965–970. https://doi.org/10.1002/mrm.21043

725 Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic
726 reviews and meta-analyses: The PRISMA statement. *BMJ (Online)*, *339*(7716), 332–336.
727 https://doi.org/10.1136/bmj.b2535

728 Murali-Manohar, S., Borbath, T., Wright, A. M., Soher, B., Mekle, R., & Henning, A. (2020). T2 relaxation
729 times of macromolecules and metabolites in the human brain at 9.4 T. *Magnetic Resonance in*
730 *Medicine*, *84*(2), 542–558. https://doi.org/10.1002/mrm.28174

731 Near, J., Harris, A. D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., Slotboom, J., Wilson, M., & Gasparovic,
732 C. (2021). Preprocessing, analysis and quantification in single-voxel magnetic resonance
733 spectroscopy: experts' consensus recommendations. *NMR in Biomedicine*, *34*(5), 1–23.
734 https://doi.org/10.1002/nbm.4257

735 Near, J., Leung, I., Claridge, T., Cowen, P., & Jezzard, P. (2012). Chemical shifts and coupling constants of
736 the GABA spin system. *Proc. Intl. Soc. Mag. Reson. Med.*, *20*(1993).

737 Oeltzschner, G., Saleh, M. G., Rimbault, D., Mikkelsen, M., Chan, K. L., Puts, N. A. J., & Edden, R. A. E.
738 (2019). Advanced Hadamard-encoded editing of seven low-concentration brain metabolites:
739 Principles of HERCULES. *NeuroImage*, *185*(September 2018), 181–190.
740 https://doi.org/10.1016/j.neuroimage.2018.10.002

741 Oeltzschner, G., Zöllner, H. J., Hui, S. C. N., Mikkelsen, M., Saleh, M. G., Tapper, S., & Edden, R. A. E.
742 (2020). Osprey: Open-source processing, reconstruction & estimation of magnetic resonance
743 spectroscopy data. *Journal of Neuroscience Methods*, *343*(June), 108827.
744 https://doi.org/10.1016/j.jneumeth.2020.108827

745 Öz, G., Deelchand, D. K., Wijnen, J. P., Mlynárik, V., Xin, L., Mekle, R., Noeske, R., Scheenen, T. W. J.,
746 Tkáč, I., Andronesi, O., Barker, P. B., Bartha, R., Berrington, A., Boer, V., Cudalbu, C., Emir, U. E.,
747 Ernst, T., Fillmer, A., Heerschap, A., … Wilson, M. (2021). Advanced single voxel 1H magnetic
748 resonance spectroscopy techniques in humans: Experts' consensus recommendations. *NMR in*
749 *Biomedicine*, *34*(5), 1–18. https://doi.org/10.1002/nbm.4236

750 Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff,
751 J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M.,
752 Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., … Mckenzie, J. E. (2021). PRISMA 2020
753 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews.
754 *The BMJ*, *372*. https://doi.org/10.1136/bmj.n160

755 Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables
756 without calculating derivatives. *The Computer Journal*, *7*(2), 155–162.
757 https://doi.org/10.1093/comjnl/7.2.155

758 Powell, M. J. D. (1994). A Direct Search Optimization Method That Models the Objective and Constraint
759 Functions by Linear Interpolation. In S. Gomez & J.-P. Hennart (Eds.), *Advances in Optimization and*

Numerical Analysis (pp. 51–67). Springer Netherlands. https://doi.org/10.1007/978-94-015-8330-5_4

Rizzo, R., Dziadosz, M., Kyathanahally, S. P., Shamaei, A., & Kreis, R. (2023). Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias. *Magnetic Resonance in Medicine*, *89*(5), 1707–1727. https://doi.org/10.1002/mrm.29561

Saleh, M. G., Oeltzschner, G., Chan, K. L., Puts, N. A. J., Mikkelsen, M., Schär, M., Harris, A. D., & Edden, R. A. E. (2016). Simultaneous edited MRS of GABA and glutathione. *NeuroImage*, *142*, 576–582. https://doi.org/10.1016/j.neuroimage.2016.07.056

Scheenen, T. W. J., Heerschap, A., & Klomp, D. W. J. (2008). Towards 1H-MRSI of the human brain at 7T with slice-selective adiabatic refocusing pulses. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *21*(1–2), 95–101. https://doi.org/10.1007/s10334-007-0094-y

Scheenen, T. W. J., Klomp, D. W. J., Wijnen, J. P., & Heerschap, A. (2008). Short echo time1H-MRSI of the human brain at 3T with minimal chemical shift displacement errors using adiabatic refocusing pulses. *Magnetic Resonance in Medicine*, *59*(1), 1–6. https://doi.org/10.1002/mrm.21302

Shamaei, A., Starcukova, J., Pavlova, I., & Starcuk, Z. (2023). Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *Magnetic Resonance in Medicine*, *89*(3), 1221–1236. https://doi.org/10.1002/mrm.29498

Song, Y., Zöllner, H. J., Hui, S. C. N., Hupfeld, K. E., Oeltzschner, G., & Edden, R. A. E. (2023). Impact of gradient scheme and non-linear shimming on out-of-voxel echo artifacts in edited MRS. *NMR in Biomedicine*, *36*(2). https://doi.org/10.1002/nbm.4839

Sørensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G., & Ernst, R. R. (1984). Product operator formalism for the description of NMR pulse experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *16*, 163–192. https://doi.org/10.1016/0079-6565(84)80005-9

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab*, *5*(4), 1–34.

Starck, G., Carlsson, A., Ljungberg, M., & Forssell-Aronsson, E. (2009). k-space analysis of point-resolved spectroscopy (PRESS) with regard to spurious echoes in in vivo (1)H MRS. *NMR IN BIOMEDICINE*, *22*(2), 137–147. https://doi.org/10.1002/nbm.1289

Tapper, S., Mikkelsen, M., Dewey, B. E., Zöllner, H. J., Hui, S. C. N., Oeltzschner, G., & Edden, R. A. E. (2021). Frequency and phase correction of J-difference edited MR spectra using deep learning. *Magnetic Resonance in Medicine*, *85*(4), 1755–1765. https://doi.org/10.1002/mrm.28525

The MathWorks Inc. (2022). *MATLAB version: 9.13.0 (R2022b)*. The MathWorks Inc. https://www.mathworks.com

Tkáč, I., Andersen, P., Adriany, G., Merkle, H., Uğurbil, K., & Gruetter, R. (2001). In vivo ^{1} H NMR spectroscopy of the human brain at 7 T. *Magnetic Resonance in Medicine*, *46*(3), 451–456. https://doi.org/10.1002/mrm.1213

798 Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

799 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E.,
800 Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J.,
801 Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0:
802 fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.
803 https://doi.org/10.1038/s41592-019-0686-2

804 Wilson, M., Andronesi, O., Barker, P. B., Bartha, R., Bizzi, A., Bolan, P. J., Brindle, K. M., Choi, I., Cudalbu,
805 C., Dydak, U., Emir, U. E., Gonzalez, R. G., Gruber, S., Gruetter, R., Gupta, R. K., Heerschap, A.,
806 Henning, A., Hetherington, H. P., Huppi, P. S., … Howe, F. A. (2019). Methodological consensus on
807 clinical proton MRS of the brain: Review and recommendations. *Magnetic Resonance in Medicine*,
808 *82*(2), 527–550. https://doi.org/10.1002/mrm.27742

809 Yablonskiy, D. A., & Haacke, E. M. (1994). Theory of NMR signal behavior in magnetically
810 inhomogeneous tissues: The static dephasing regime. *Magnetic Resonance in Medicine*, *32*(6),
811 749–763. https://doi.org/10.1002/mrm.1910320610

812 Zhang, Y., & Shen, J. (2023). Quantification of spatially localized MRS by a novel deep learning approach
813 without spectral fitting. *Magnetic Resonance in Medicine*. https://doi.org/10.1002/mrm.29711

814 Zöllner, H. J., Thiel, T. A., Füllenbach, N.-D., Jördens, M. S., Ahn, S., Wilms, L. M., Ljimani, A., Häussinger,
815 D., Butz, M., Wittsack, H.-J., Schnitzler, A., & Oeltzschner, G. (2023). J-difference GABA-edited MRS
816 reveals altered cerebello-thalamo-cortical metabolism in patients with hepatic encephalopathy.
817 *Metabolic Brain Disease*, *38*(4), 1221–1238. https://doi.org/10.1007/s11011-023-01174-x

818

819 **Figures:**

820

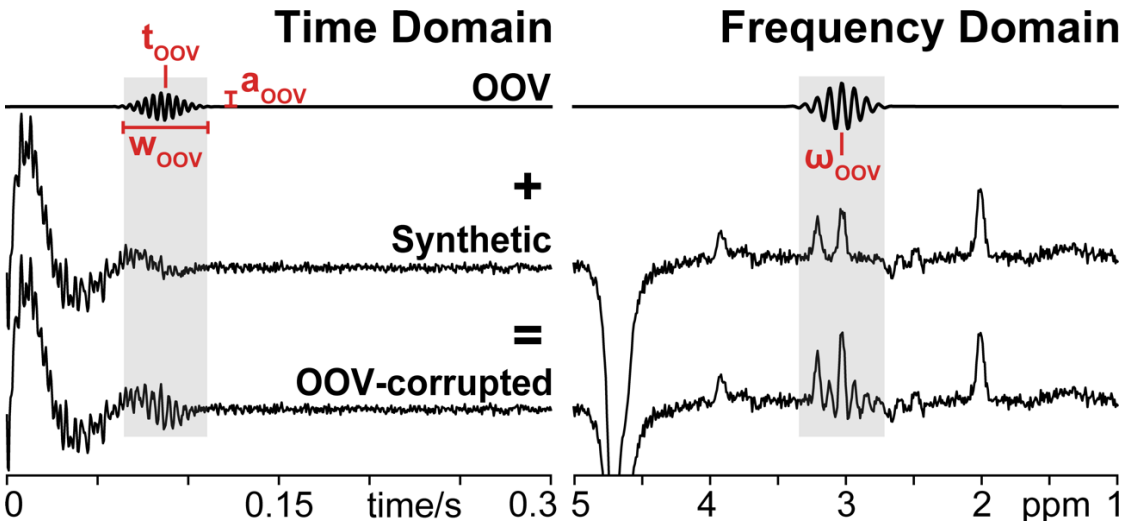


Figure 1. Simulation of OOV echoes and OOV-corrupted synthetic data: OOV echoes were simulated as complex time-domain signals with a center timepoint (τ_{OOV}), width (W_{OOV}), frequency (ω_{OOV}), phase (Φ_{OOV}), amplitude (a_{OOV}). OOV echoes were added to 85% of synthetic data to create datasets for training and evaluation.

821

822

823

824

825

826

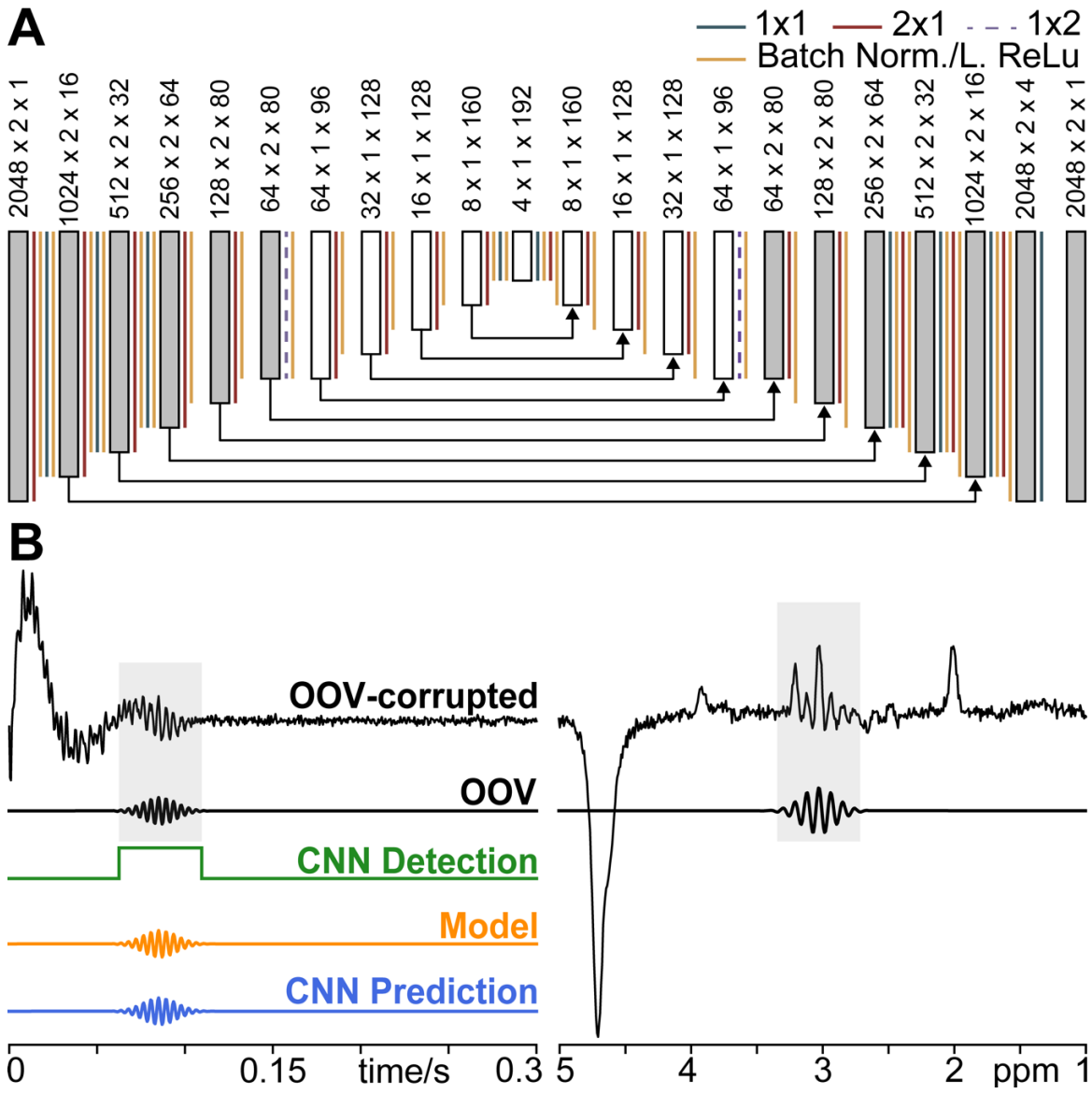


Figure 2. Convolutional Neural Network Architecture, Input, and Output: A) Fully convolutional neural network architecture used for both the Detection and Prediction Network. Convolutional strides, batch normalization, and Leaky ReLu activation functions are denoted by a colored line. Dark gray blocks represent complex data with the 2nd dimension representing real and imaginary components, while white blocks represent the network abstracted single dimension. Arrows show residual connections. Note, inputs and outputs are all time-domain signals; Frequency-domain is shown for convenient visualization. B) OOV-corrupted synthetic example and the isolated OOV. The complex OOV-corrupted data was used as the Detection and Prediction Network input. The target Output is the isolated OOV.

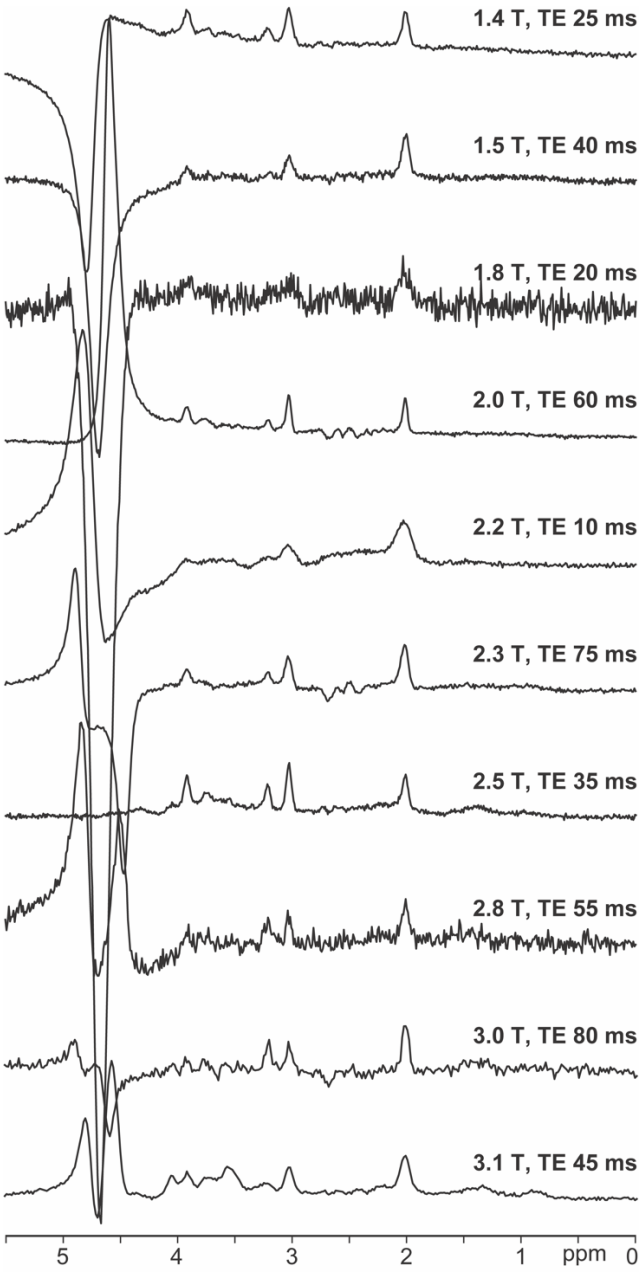


Figure 3. AGNOSTIC synthetic dataset. 10 representative spectra from the AGNOSTIC dataset. The 10 examples show the diversity of field strength, TE, linewidths, and residual water signal present among the data. Note, examples are shown here in the frequency-domain to better illustrate the heterogeneity, but the dataset provides time-domain examples.

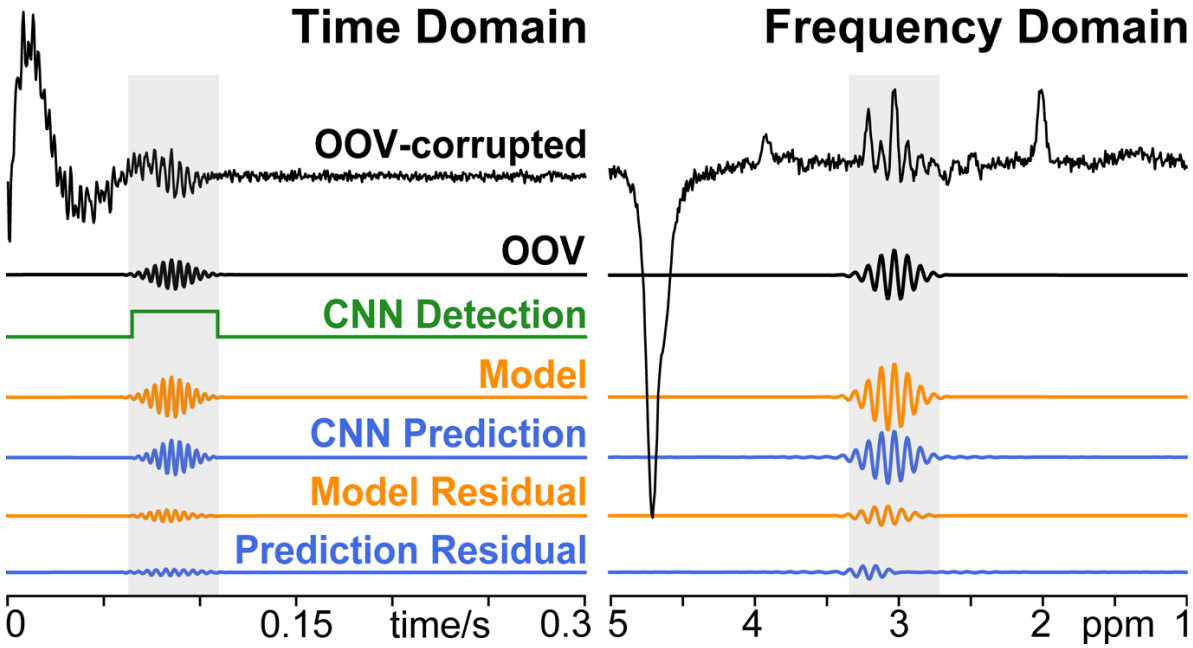


Figure 4. OOV-corrupted example: OOV-corrupted synthetic example and the isolated OOV. Results from Detection Network (green), Model (orange), and Prediction Network (blue) are shown below the ground truth OOV-corrupted and OOV. OOV residuals are shown for the Model (orange) and Prediction Network (blue) demonstrating remaining signal after subtraction. Note, frequency-domain is shown for convenient visualization, but the Detection Network, Modeling, and Prediction Network all operate on time-domain signals.

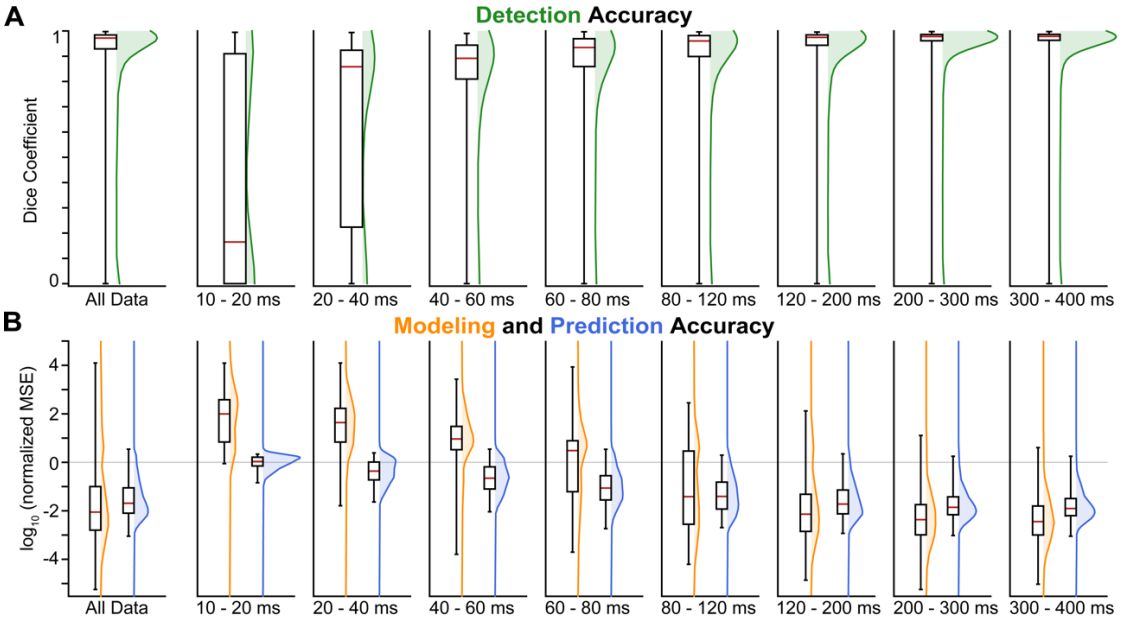


Figure 5. Evaluation of Detection Network, Modeling, and Prediction Network. A testing set with 7200 (2400 examples with 3 different OOV echoes) unseen examples was used to evaluate the A) Detection Network and B) Modeling and Prediction Network. Performance across the whole test set is shown on the left-hand side. Performance across the binned center timepoint (τOOV) is shown across the right-hand side.

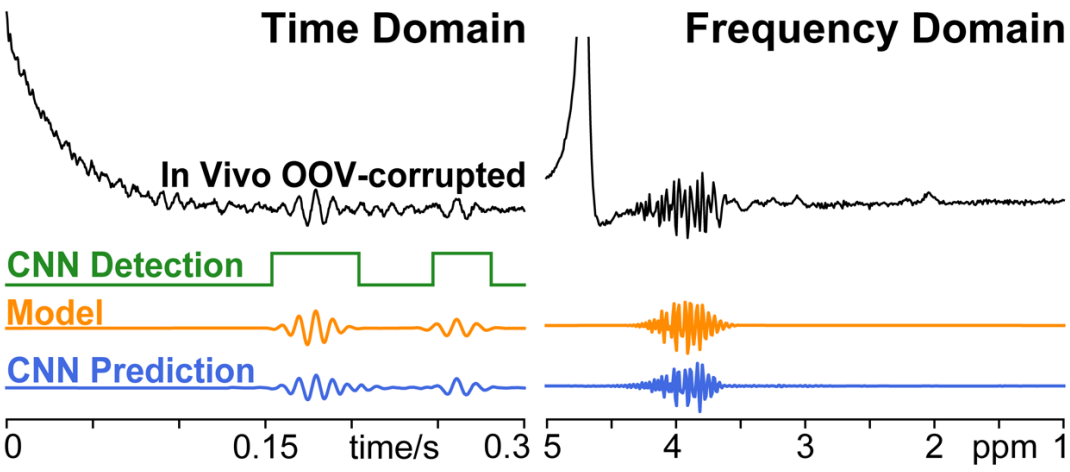


Figure 6. *In vivo* MEGA-PRESS OOV-corrupted example. Results from Detection Network (green), Model (orange), and Prediction Network (blue) are shown below. Detection and Prediction CNNs identified and reconstructed the OOV echo, despite having never seen data acquired with 2.89 T, 2.4 kHz spectral width, 68 ms, editing, nor real RF pulses. Note, frequency-domain is shown for convenient visualization, but the Detection Network, Modeling, and Prediction Network all operate on time-domain signals.

858

859 **Tables:**

860

861 Table 1. Parametrization of the residual water signal components within AGNOSTIC.

	Location / ppm		Phase / deg		Amplitude	
Component	Low	High	Low	High	Low	High
1	4.679	4.711	−10	10	1.00	1.00
2	4.599	4.641	15	45	.35	.55
3	4.759	4.801	−60	−30	.35	.55
4	4.449	4.541	−70	45	.10	.25
5	4.859	4.901	105	135	.10	.25

862

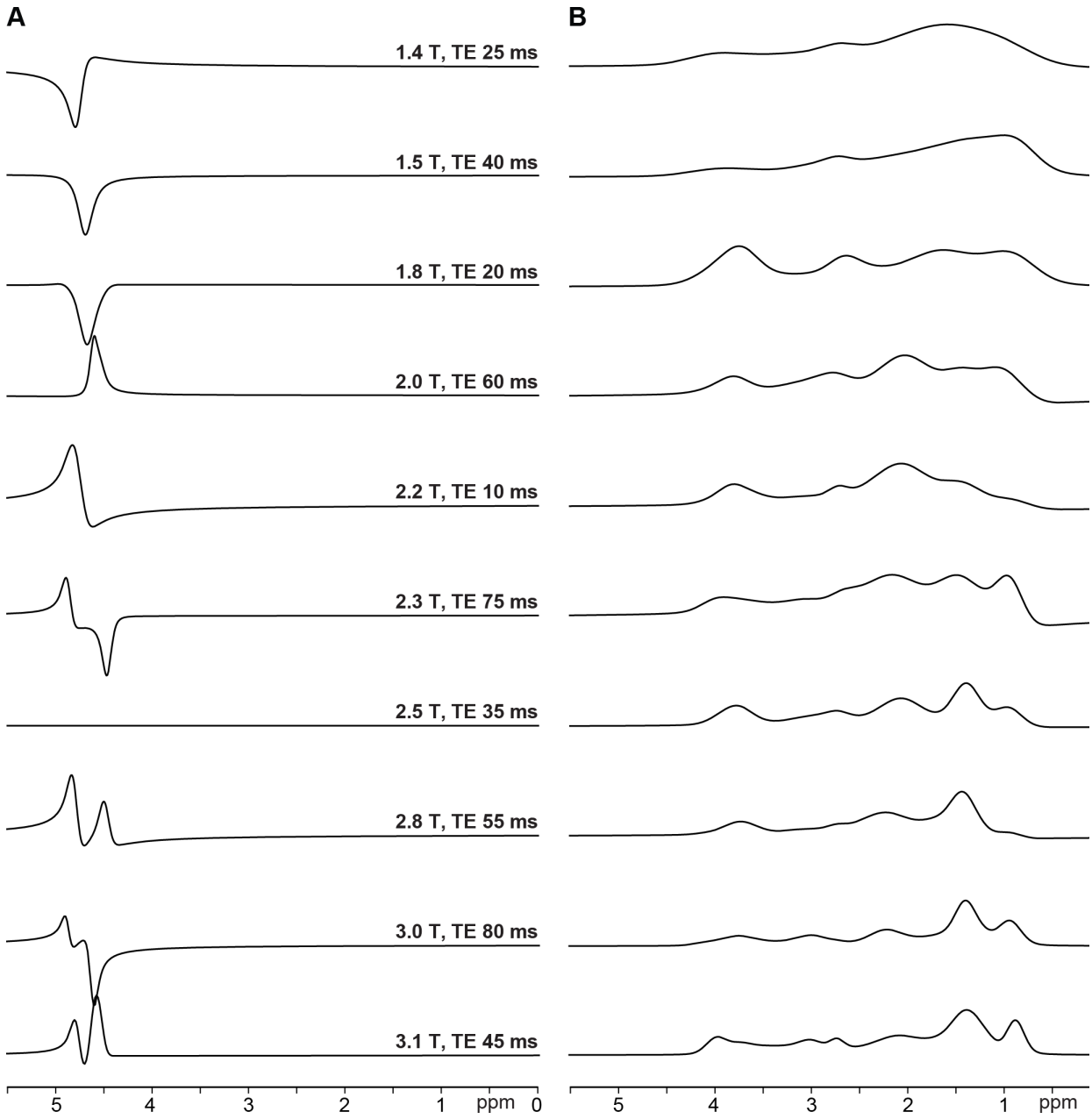
863

864

865

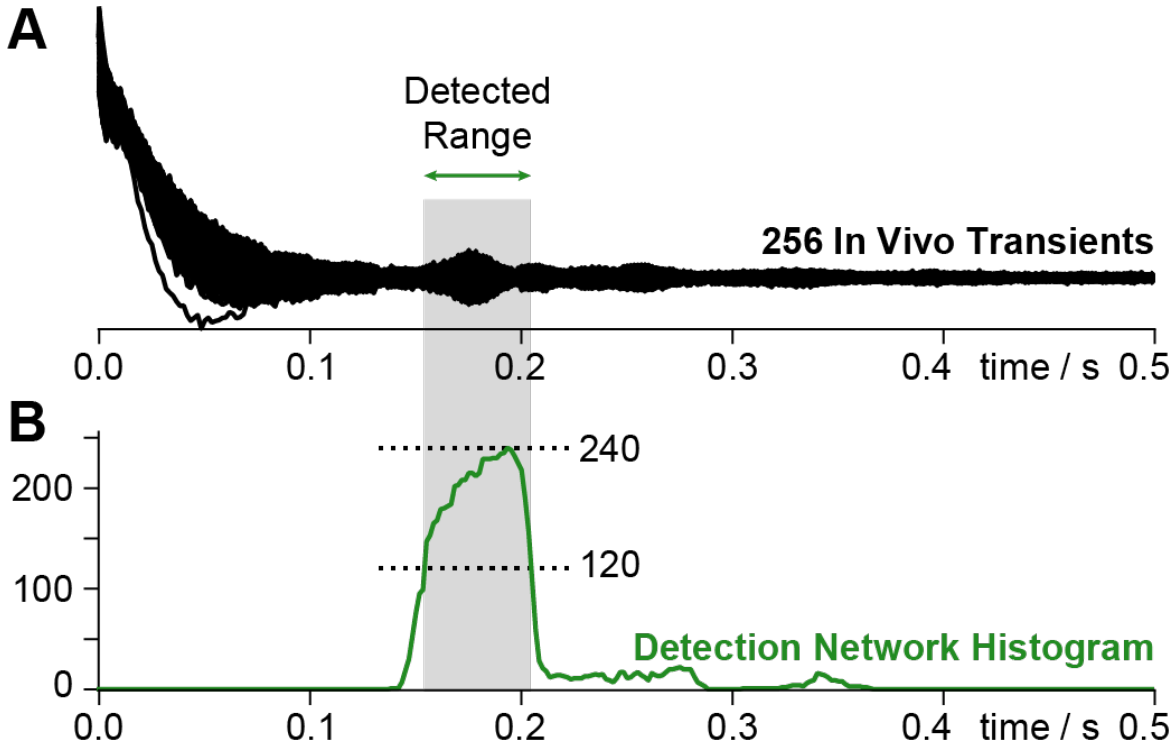
866 **Supplemental Material:**

867



868

869 **Supplemental Figure 1:** Ten representative examples of **A)** residual water and **B)** macromolecule
870 components. Examples match the full spectra shown in Figure 3; each spectrum is scaled independently
871 for visualization.



Supplemental Figure 2: Time-domain window (gray) used to calculate the fractional reduction in standard deviation for the *in vivo* transients. **A)** Each of the 256 MEGA-PRESS transients (128 Edit-on and 128 Edit-Off) overlaid. **B)** Histogram (green) showing the total number of detections by the Detection Network across each timepoint. This window was established algorithmically by using 50% of the maximum count as a threshold for the window.

	Synthetic Range (mM)	
Metabolite	**Low**	**High**
Acetate	0.00	0.00
Alanine*	0.47	0.77
Ascorbate	0.36	1.53
Aspartate	0.00	4.66
Adenosine Triphosphate	0.00	0.00
β-Hydroxybutyrate	0.00	0.00
β-Hydroxyglutarate	0.00	0.00
Citrate	0.00	0.00
Creatine	1.41	10.50
Cysteine	0.00	0.00
Ethanol Amine	0.00	0.00
Ethyl Alcohol	0.00	0.00
γ-Amino Butyric Acid	0.52	1.99
Glucose*	0.94	1.53
Glutamine	0.26	3.64
Glutamate	3.88	13.17
Glycerophosphocholine	0.05	5.00
Glutathione	0.16	2.41
Glycine*	0.94	1.53
Glycerol	0.00	0.00
Histamine	0.00	0.00
Histidine	0.00	0.00
Homocarnosine	0.00	0.00
Lactate	0.00	1.44
Macromolecule 0.92	1.00	30.00
Macromolecule 1.21	1.00	8.00
Macromolecule 1.39	1.00	35.00
Macromolecule 1.67	1.00	15.00
Macromolecule 2.04	1.00	35.00
Macromolecule 2.26	1.00	20.00
Macromolecule 2.56	1.00	5.00
Macromolecule 2.70	1.00	7.00
Macromolecule 2.99	1.00	10.00
Macromolecule 3.21	1.00	7.00
Macromolecule 3.62	1.00	5.00
Macromolecule 3.75	1.00	10.00
Macromolecule 3.86	1.00	4.00
Macromolecule 4.03	1.00	7.00
Myo-inositol†	2.08	14.00
N-Acetylaspartate†	5.38	18.00
N-Acetylaspartylglutamate†	0.26	2.26
Phosphocholine*	0.01	2.00
Phosphocreatine*	3.38	6.44
Phosphoethanolamine*	1.41	2.30
Phosphoethyl Alcohol	0.00	0.00
Phenylalanine	0.00	0.00
Scyllo-inositol	0.00	0.39
Serine	0.00	0.00
Taurine	0.00	2.89
Threonine	0.00	0.00
Tryptophan	0.00	0.00
Tyrosine	0.00	0.00
Valine	0.00	0.00

Supplemental Table 1: Concentration ranges for the healthy brain, used to generate synthetic spectra. These mM values were based upon a meta-analysis preliminary to (Gudmundson et al., 2023), with some values (marked *) supplemented from the Fit Challenge ranges (Marjańska et al., 2021) and other ranges (marked †) extended to offer greater flexibility. Concentrations were sampled uniformly between the low and high values to generate the synthetic spectra.

Disease / Metabolite	Synthetic Range Low	High
Seizure		
Creatine	0.918	1.012
Phosphocreatine	0.918	1.012
Glycerophosphocholine	0.731	1.147
Phosphocholine	0.731	1.147
γ-Amino Butyric Acid	0.930	1.173
Glutamate	0.787	1.247
Glutamine	0.787	1.247
Glutathione	0.887	1.243
Myo-inositol	0.802	1.134
N-Acetylaspartate	0.751	1.002
N-Acetylaspartylglutamate	0.751	1.002
Stroke		
Creatine	0.684	1.146
Phosphocreatine	0.684	1.146
Glycerophosphocholine	0.855	1.527
Phosphocholine	0.855	1.527
Glutamate	0.874	1.140
Glutamine	0.874	1.140
Lactate	1.000	6.922
Myo-inositol	0.827	1.265
N-Acetylaspartate	0.727	1.074
N-Acetylaspartylglutamate	0.727	1.074
Traumatic Brain Injury		
Aspartate	0.785	0.910
Creatine	0.814	1.162
Phosphocreatine	0.814	1.162
Glycerophosphocholine	0.930	1.057
Phosphocholine	0.930	1.057
γ-Amino Butyric Acid	0.860	0.984
Glutamate	0.824	1.214
Glutamine	0.824	1.214
Myo-inositol	0.737	1.315
N-Acetylaspartate	0.795	1.011
N-Acetylaspartylglutamate	0.795	1.011
Type-1 Diabetes		
Aspartate	0.895	1.496
Creatine	0.977	1.039
Phosphocreatine	0.977	1.039
Glycerophosphocholine	1.034	1.140
Phosphocholine	1.034	1.140
Glutamate	0.895	1.216
Glutamine	0.956	1.353
Glutathione	0.872	1.435
Myo-inositol	0.893	1.092
N-Acetylaspartate	0.947	1.008
N-Acetylaspartylglutamate	0.947	1.008
Scyllo-inositol	0.501	0.992
Taurine	0.754	1.322

Metabolite	Synthetic Range Low	High
Cancer		
Creatine	0.256	1.340
Phosphocreatine	0.256	1.340
Glycerophosphocholine	1.139	1.949
Phosphocholine	1.139	1.949
Glutamate	0.780	1.320
Glutamine	0.780	1.320
Lactate	1.00	9.99
Myo-inositol	0.829	1.519
N-Acetylaspartate	0.509	0.956
N-Acetylaspartylglutamate	0.509	0.956
Chronic Pain		
Glycerophosphocholine	0.943	1.285
Phosphocholine	0.943	1.285
γ-Amino Butyric Acid	0.896	1.168
Glutamate	0.790	1.121
Glutamine	0.790	1.121
Myo-inositol	0.942	1.049
N-Acetylaspartate	0.775	1.280
N-Acetylaspartylglutamate	0.775	1.280
Migraine		
Aspartate	0.434	1.409
Creatine	0.921	1.011
Phosphocreatine	0.921	1.011
Glycerophosphocholine	0.959	1.137
Phosphocholine	0.959	1.137
Glutamate	0.841	1.119
Glutamine	0.841	1.119
Myo-inositol	0.866	1.032
N-Acetylaspartate	0.755	1.067
N-Acetylaspartylglutamate	0.755	1.067
Fibromyalgia		
Creatine	0.760	1.429
Phosphocreatine	0.760	1.429
Glycerophosphocholine	0.840	1.236
Phosphocholine	0.840	1.236
γ-Amino Butyric Acid	0.724	0.937
Glutamate	1.005	1.104
Glutamine	0.711	1.107
Myo-inositol	0.844	1.232
N-Acetylaspartate	0.847	1.061
N-Acetylaspartylglutamate	0.847	1.061

…table continued on next page

Disease / Metabolite	Synthetic Range Low	High
Post-Traumatic Stress Disorder		
Creatine	0.940	1.235
Phosphocreatine	0.940	1.235
Glycerophosphocholine	0.843	1.284
Phosphocholine	0.843	1.283
γ-Amino Butyric Acid	0.982	1.059
Glutamate	0.892	1.134
Glutamine	0.892	1.134
Myo-inositol	0.939	1.198
N-Acetylaspartate	0.969	1.156
N-Acetylaspartylglutamate	0.969	1.156
Obsessive Compulsive Disorder		
Creatine	0.890	1.320
Phosphocreatine	0.890	1.320
Glycerophosphocholine	0.784	1.223
Phosphocholine	0.784	1.223
Glutamate	0.868	1.243
Glutamine	0.868	1.243
Myo-inositol	0.743	1.437
N-Acetylaspartate	0.846	1.100
N-Acetylaspartylglutamate	0.846	1.100
Depression		
Creatine	0.938	1.021
Phosphocreatine	0.938	1.021
Glycerophosphocholine	0.741	1.158
Phosphocholine	0.741	1.158
γ-Amino Butyric Acid	0.769	1.400
Glutamate	0.872	1.119
Glutamine	0.894	1.177
Glutathione	0.822	1.082
Myo-inositol	0.874	1.239
N-Acetylaspartate	0.864	1.080
N-Acetylaspartylglutamate	0.864	1.080
Addiction		
Creatine	0.775	1.161
Phosphocreatine	0.755	1.161
Glycerophosphocholine	0.788	1.202
Phosphocholine	0.788	1.202
γ-Amino Butyric Acid	0.669	1.289
Glutamate	0.807	1.229
Glutamine	0.807	1.229
Glycine	0.969	1.335
Glutathione	0.935	1.442
Myo-inositol	0.820	1.135
N-Acetylaspartate	0.761	1.195
N-Acetylaspartylglutamate	0.761	1.195

Disease / Metabolite	Synthetic Range Low	High
Schizophrenia		
Creatine	0.948	1.045
Phosphocreatine	0.948	1.045
Glycerophosphocholine	0.946	1.157
Phosphocholine	0.946	1.157
γ-Amino Butyric Acid	0.732	1.261
Glutamate	0.857	1.164
Glutamine	0.857	1.164
Myo-inositol	0.806	1.239
N-Acetylaspartate	0.910	1.103
N-Acetylaspartylglutamate	0.910	1.103
Psychosis		
Creatine	0.983	1.059
Phosphocreatine	0.983	1.059
Glycerophosphocholine	0.892	1.127
Phosphocholine	0.892	1.127
γ-Amino Butyric Acid	0.725	1.176
Glutamate	0.813	1.172
Glutamine	0.813	1.172
Glycine	1.131	1.423
Glutathione	0.917	1.034
Myo-inositol	0.892	1.090
N-Acetylaspartate	0.910	1.048
N-Acetylaspartylglutamate	0.910	1.048
Personality Disorder		
Creatine	0.961	1.110
Phosphocreatine	0.961	1.110
Glycerophosphocholine	0.925	1.007
Phosphocholine	0.925	1.007
Glutamate	0.949	1.207
Glutamine	0.949	1.207
Glutathione	0.917	1.034
Myo-inositol	0.989	1.081
N-Acetylaspartate	0.880	0.997
N-Acetylaspartylglutamate	0.880	0.997
Bipolar Disorder		
Creatine	0.900	1.061
Phosphocreatine	0.900	1.061
Glycerophosphocholine	0.854	1.269
Phosphocholine	0.854	1.269
Glutamate	0.907	1.115
Glutamine	0.907	1.115
Glutathione	0.957	1.150
Myo-inositol	0.812	1.209
N-Acetylaspartate	0.863	1.109
N-Acetylaspartylglutamate	0.863	1.109

…table continued on next page

Disease / Metabolite	Synthetic Range Low	High
Multiple Sclerosis		
Glycerophosphocholine	0.880	1.077
Phosphocholine	0.880	1.077
γ-Amino Butyric Acid	0.851	1.017
Glutamate	0.887	1.030
Glutamine	0.887	1.030
Glutathione	0.844	1.069
Myo-inositol	0.892	1.078
N-Acetylaspartate	0.924	1.044
N-Acetylaspartylglutamate	0.924	1.044
Parkinson's Disease		
Creatine	0.850	1.100
Phosphocreatine	0.850	1.100
Glycerophosphocholine	0.780	1.201
Phosphocholine	0.780	1.201
γ-Amino Butyric Acid	0.679	1.390
Glutamate	0.887	1.224
Glutamine	0.887	1.224
Myo-inositol	0.810	1.190
N-Acetylaspartate	0.756	1.240
N-Acetylaspartylglutamate	0.756	1.240
Essential Tremor		
Creatine	0.924	1.053
Phosphocreatine	0.924	1.053
Glycerophosphocholine	0.851	1.044
Phosphocholine	0.851	1.044
γ-Amino Butyric Acid	0.802	1.218
Glutamate	1.050	1.434
Glutamine	1.050	1.434
N-Acetylaspartate	0.919	1.136
N-Acetylaspartylglutamate	0.919	1.136
Dementia		
Ascorbate	1.132	1.231
Aspartate	1.028	1.168
Creatine	1.010	1.028
Phosphocreatine	1.010	1.028
Glycerophosphocholine	0.850	1.150
Phosphocholine	0.850	1.150
γ-Amino Butyric Acid	0.513	1.183
Glutamate	0.771	1.139
Glutamine	0.955	1.172
Myo-inositol	0.801	1.397
N-Acetylaspartate	0.723	1.038
N-Acetylaspartylglutamate	0.723	1.038
Scyllo-inositol	0.476	1.312
Taurine	0.882	1.013
APOE4		
Aspartate	1.028	1.168
Glycerophosphocholine	0.965	1.019
Phosphocholine	0.965	1.019
γ-Amino Butyric Acid	0.513	1.183
Glucose	0.971	1.028
Glutamate	0.836	1.126
Glutamine	0.909	1.232
Glutathione	0.834	1.103
Myo-inositol	0.959	1.092
N-Acetylaspartate	0.895	1.063
N-Acetylaspartylglutamate	0.895	1.063

Supplemental Table 2: Clinical population scaling factors used to generate synthetic spectra. In each case the simulated concentration for a given clinical spectrum was determined by a uniformly sampled concentration drawn from the ranges shown in Supplemental Table 1, multiplied by a scaling factor determined by a uniformly sampled scalar from these ranges provided in Supplemental Table 2.

	Synthetic Range (ms)	
Metabolite	**Low**	**High**
Acetate	0.00	0.00
Alanine*	100.00	250.00
Ascorbate	100.00	250.00
Aspartate	120.15	204.55
Adenosine Triphosphate	0.00	0.00
β-Hydroxybutyrate	0.00	0.00
β-Hydroxyglutarate	0.00	0.00
Citrate	0.00	0.00
Creatine 3.03	164.08	242.70
Creatine 3.91	135.18	213.80
Creatine 6.65	164.08	242.70
Cysteine	0.00	0.00
Ethanolamine	0.00	0.00
Ethyl Alcohol	0.00	0.00
γ-Amino Butyric Acid	77.37	161.77
Glucose	100.00	250.00
Glutamine	103.96	184.89
Glutamate	140.96	219.58
Glycerophosphocholine	198.77	278.54
Glutathione	108.59	188.36
Glycine	121.31	204.55
Glycerol	0.00	0.00
Histamine	0.00	0.00
Histidine	0.00	0.00
Homocarnosine	0.00	0.00
Lactate	142.12	226.52
Macromolecule 0.92	20.00	60.00
Macromolecule 1.21	20.00	60.00
Macromolecule 1.39	20.00	60.00
Macromolecule 1.67	20.00	60.00
Macromolecule 2.04	20.00	60.00
Macromolecule 2.26	20.00	60.00
Macromolecule 2.56	20.00	60.00
Macromolecule 2.70	20.00	60.00
Macromolecule 2.99	20.00	60.00
Macromolecule 3.21	20.00	60.00
Macromolecule 3.62	20.00	60.00
Macromolecule 3.75	20.00	60.00
Macromolecule 3.86	20.00	60.00
Macromolecule 4.03	20.00	60.00
Myo-inositol†	139.80	219.58
N-Acetylaspartate	242.70	320.17
N-Acetylaspartylglutamate	132.87	216.11
Phosphocholine	100	250
Phosphocreatine 3.03	130	210
Phosphocreatine 3.93	100	180
Phosphocreatine 6.58	130	210
Phosphocreatine 7.30	130	210
Phosphoethanolamine	100	250
Phosphoethyl Alcohol	0.00	0.00
Phenylalanine	0.00	0.00
Scyllo-inositol	100	250
Serine	0.00	0.00
Taurine	151.37	231.14
Threonine	0.00	0.00
Tryptophan	0.00	0.00
Tyrosine	0.00	0.00
Valine	0.00	0.00

Supplemental Table 3: T_2 Relaxation time ranges in milliseconds for the healthy brain derived from 1.5 T multiple meta-regression preliminary to (Gudmundson et al., 2023). Relaxation times were sampled uniformly between the low and high values.

B_0 (T)	Every Point	Every 2nd Point	Every 3rd Point	Every 4th Point	Every 5th Point	Every 6th Point	Every 7th Point	Every 8th Point
1.4	3733.33	1866.67	1244.44	933.33	746.67	622.22	533.33	466.67
1.5	4000.00	2000.00	1333.33	1000.00	800.00	666.67	571.43	500.00
1.6	4266.67	2133.33	1422.22	1066.67	853.33	711.11	609.52	533.33
1.7	4533.33	2266.67	1511.11	1133.33	906.67	755.56	647.62	566.67
1.8	4800.00	2400.00	1600.00	1200.00	960.00	800.00	685.71	600.00
1.9	5066.67	2533.33	1688.89	1266.67	1013.33	844.44	723.81	633.33
2.0	5333.33	2666.67	1777.78	1333.33	1066.67	888.89	761.90	666.67
2.1	5600.00	2800.00	1866.67	1400.00	1120.00	933.33	800.00	700.00
2.2	5866.67	2933.33	1955.56	1466.67	1173.33	977.78	838.10	733.33
2.3	6133.33	3066.67	2044.44	1533.33	1226.67	1022.22	876.19	766.67
2.4	6400.00	3200.00	2133.33	1600.00	1280.00	1066.67	914.29	800.00
2.5	6666.67	3333.33	2222.22	1666.67	1333.33	1111.11	952.38	833.33
2.6	6933.33	3466.67	2311.11	1733.33	1386.67	1155.56	990.48	866.67
2.7	7200.00	3600.00	2400.00	1800.00	1440.00	1200.00	1028.57	900.00
2.8	7466.67	3733.33	2488.89	1866.67	1493.33	1244.44	1066.67	933.33
2.9	7733.33	3866.67	2577.78	1933.33	1546.67	1288.89	1104.76	966.67
3.0	8000.00	4000.00	2666.67	2000.00	1600.00	1333.33	1142.86	1000.00
3.1	8266.67	4133.33	2755.56	2066.67	1653.33	1377.78	1180.95	1033.33

Supplemental Table 4: Field strengths (*Tesla*) and possible spectral widths (*Hertz*) available using the AGNOSTIC basis sets. These combinations are achievable by subsampling the time-domain from every timepoint to every 8th timepoint and allows for maintaining a minimum of 2048 timepoints. Each of these combinations is available for the 15 echo times, from 10 ms to 80 ms, in steps of 5 ms.

References:

Gudmundson, A. T., Koo, A., Virovka, A., Amirault, A. L., Soo, M., Cho, J. H., Oeltzschner, G., Edden, R. A. E., & Stark, C. E. L. (2023). Meta-analysis and open-source database for in vivo brain Magnetic Resonance spectroscopy in health and disease. Analytical Biochemistry, 676, 115227. https://doi.org/10.1016/j.ab.2023.115227

Marjańska, M., Deelchand, D. K., Kreis, R., Alger, J. R., Bolan, P. J., Borbath, T., Boumezbeur, F., Fernandes, C. C., Coello, E., Nagraja, B. H., Považan, M., Ratiney, H., Sima, D., Starčuková, J., Soher, B. J., Wilson, M., & van Asten, J. J. A. (2022). Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. Magnetic Resonance in Medicine, 87(1), 11–32. https://doi.org/10.1002/mrm.28942