Improving Sensitivity to Functional Responses without a Loss of Spatiotemporal Precision in Human Brain Imaging

Logan T. Dowdle, Luca Vizioli, Steen Moeller, Cheryl Olman, Geoff Ghose, Essa Yacoub, Kamil Ugurbil

Abstract

As the neuroimaging field moves towards detecting smaller effects at higher spatial resolutions, and faster sampling rates, there is increased attention given to the deleterious contribution of unstructured, thermal noise. Here, we critically evaluate the performance of a recently developed reconstruction method, termed NORDIC, for suppressing thermal noise using datasets acquired with various field strengths, voxel sizes, sampling rates, and task designs.

Following minimal preprocessing, statistical activation (t-values) of NORDIC processed data was compared to the results obtained with alternative denoising methods. Additionally, the consistency of unbiased estimations of task responses at the single-voxel, single run level, using a finite impulse response model were evaluated. To examine the potential impact on effective image resolution, the overall smoothness of the data processed with different methods was examined. Finally, to determine if NORDIC alters or removes important temporal information, an exhaustive K-fold cross validation approach was employed, using unbiased task responses to predict held out timeseries, quantified using R^2.

After NORDIC, the t-values are increased, an improvement comparable to what could be achieved by 1.5 voxels smoothing, and task events are clearly visible and have less cross-run error. These advantages are achieved without significant compromises in spatial and temporal resolution. Cross-validated R^2s based on the unbiased models show that NORDIC is not distorting the temporal structure of the data and is the best predictor of non-denoised time courses. The results demonstrate that 1 run of NORDIC data is equivalent to using 2 to 3 original runs, and performs equally well across a diverse array of functional imaging protocols.

Significance Statement

For functional neuroimaging, the increasing availability of higher field strengths and ever higher spatiotemporal resolutions has led to concomitant increase in concerns about the deleterious effects of thermal noise. Historically this noise source was suppressed using methods that reduce spatial precision such as image blurring or averaging over a large number of trials or sessions, which necessitates large data collection efforts. Here, we critically evaluate the performance of a recently developed reconstruction method, termed NORDIC. Across datasets varying in field strength, voxel sizes, sampling rates, and task designs, NORDIC produces substantial gains in data quality. Both conventional t-statistics derived from general linear models and coefficients of determination for predicting unseen data are improved, without notably increasing image smoothness.

Introduction

The growing use of high (3 to 7 Tesla) and ultrahigh field (UHF, defined as ≥7 Tesla) magnetic fields for functional magnetic resonance imaging (fMRI) of brain activity has led to increases in the available signal-to noise-ratio (SNR) and subsequently corresponding interest in smaller voxel sizes and/or shorter repetition times. Early fMRI experiments tended to sample units of tissue that were on the order of 30 µL in voxel volume, a volume that potentially contains millions of neurons. In contrast, contemporary UHF high resolution fMRI studies, enabled by the significantly higher signal-to-noise (SNR) and functional contrast-to-noise ratios (fCNR) available at UHF (1, 2), have attained resolutions that have ~0.5 µL voxel volumes (3-5) (e.g. ~0.8 mm isotropic voxel dimensions). These developments, together with the early demonstration that neurovascular coupling, the mechanism underlying the fMRI method, has specificity at the level of mesoscopic scale organizations of the brain (6, 7), have led to a burgeoning activity in fMRI studies of cortical columns and layers (reviews (8, 9)) and other fine scale organizations (e.g. (10, 11)), including the imaging of orientation domains together with ocular dominance columns for the first time in the human brain (12, 13).

However, the small voxel volumes in such high-resolution fMRI studies have pushed the temporal SNR (tSNR) of fMRI time series into a low SNR regime where thermal noise of the MR measurement begins to dominate over signal fluctuations induced by physiological processes (often referred to as “physiological noise”) (14, 15). Similarly, the use of highly accelerated fMRI approaches, introduced for rapid coverage of large volumes at high spatial resolution using UHF (16), has increasingly become the method of choice for data acquisition. These methods, popularized by the Human Connectome Project (17, 18), also push gradient echo (GE)-based fMRI data towards the thermal noise-dominated regime as repetition time and, consequently, flip angles and the signal magnitude detected in each image decreases (17, 18). The spatially non-uniform noise amplification introduced by parallel imaging reconstructions (i.e., the g-factor noise (19)) further exacerbates the thermal noise penalty (19, 20). The resulting low SNR regime leads to difficulties in estimating the fine

scale detail of the hemodynamic response, a critical goal given the variability of the hemodynamic response across large (21-23) and small (24) regions of the brain.

The thermal noise associated with the MR measurement is not impacted by the various denoising approaches intended to suppress the contributions of structured, i.e. non-white noise in an fMRI time series, emanating from physiological processes (e.g., (25-30)), low-frequency signal drift, or motion. Spatial filtering (i.e. “smoothing”), on the other hand, does reduce the thermal noise contribution and hence improves the temporal signal to noise ratio (tSNR) (31). However, it results in substantial losses in spatial precision (31). Similarly, combining data from several different subjects reduces noise in general *via* group averaging; however, this option is not a desirable approach for high resolution studies because it inevitably incurs some degree of spatially non-uniform complex blurring nor is it valid for studies focused on single subject responses or inter-subject variability. Notably these single subject statistical approaches, such as multivoxel pattern analyses (MVPA) (32) or encoding models (33-35), typically leverage separate runs in order to determine cross validated accuracy. In SNR-starved regimes, these methods become difficult to implement, thereby limiting the range of possible scientific questions that can be addressed.

With the growing focus on more spatially and temporally precise fMRI measurements, we present a detailed evaluation of the achievable gains, and the resulting spatiotemporal precision, of a new denoising method, Noise Reduction with Distribution Corrected (NORDIC) PCA (36, 37). NORDIC suppresses Gaussian distributed noise associated with the MR detection process in repetitively acquired images, reducing thermal noise contributions throughout the image. The aforementioned non-white noise sources, such as physiological effects, signal drift or head motion remain intact, as these structured noise components can be distinguished from Gaussian distributed noise. We compare NORDIC to a number of other noise suppression approaches using a variety of different datasets which vary in voxel size (0.8 to 2mm), repetition time (0.35 to 2.1s), field strength (3 and 7 Tesla), and use both block and event related task designs. The results obtained on 8 data sets (obtained from 3 subjects) demonstrate that the NORDIC method consistently leads to substantial

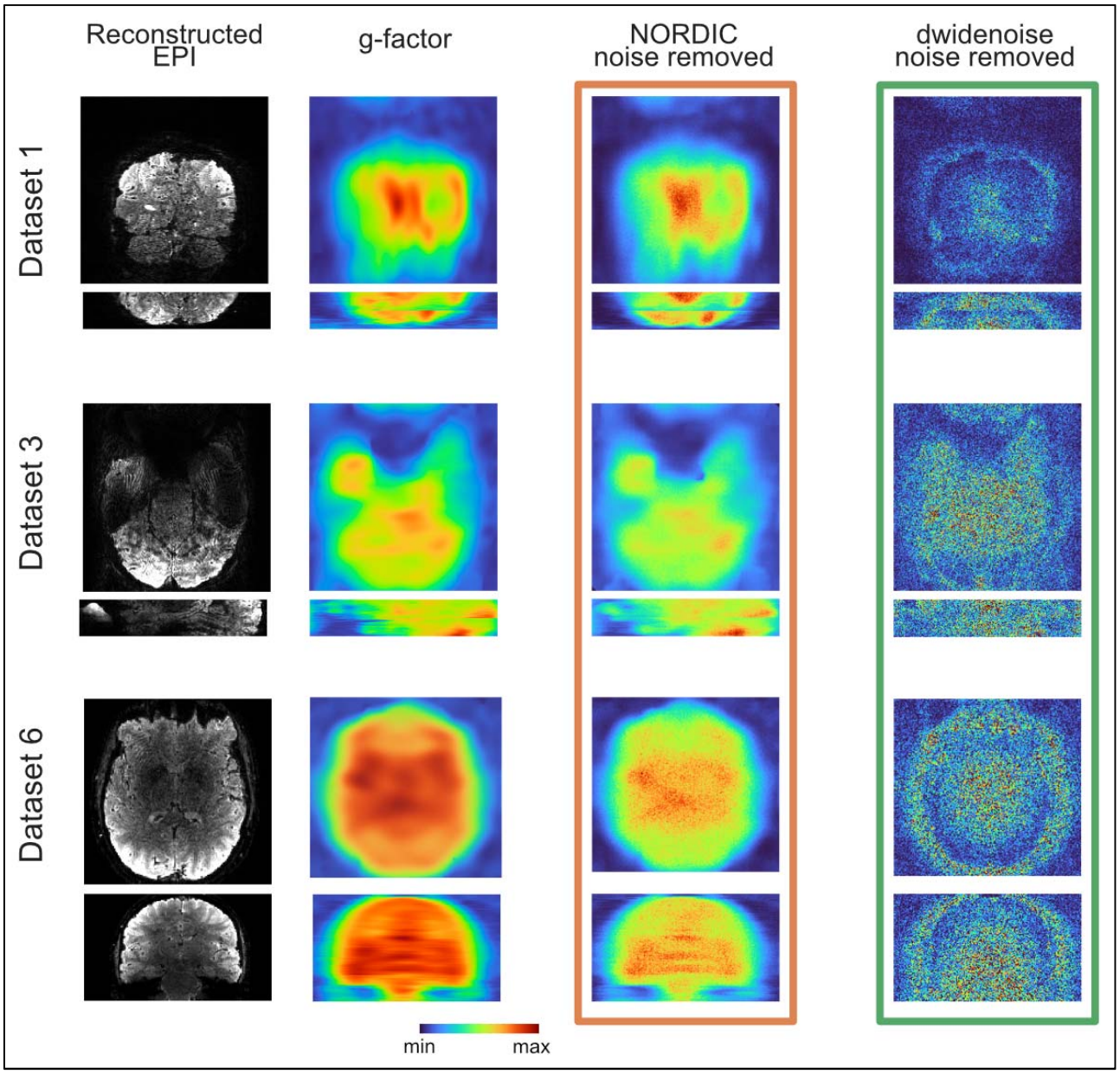
statistical gains under a conventional generalized least-squares (GLSQ) framework, produces better single-run single-voxel hemodynamic response function (HRF) estimates and does so without introducing any notable spatial or temporal blurring, thereby preserving the intrinsic information present in an fMRI timeseries.

Results

For this manuscript, 8 datasets are considered. Datasets (DS) 1, 2, 5, 6, and 7 were block designs which used a flashing checkerboard (8Hz) positioned either in a center position (‘target’) or in a surround with the center cut out (‘surround’). Representative datasets (of 8) were selected to present in figures. DS3 used an event related design in which intact and phase scrambled full color faces (14 male, 14 female) were presented. DS4 used an event related design with grayscale images of faces (20 male, 20 female), presenting neutral expressions, and DS8 (used only for FFT analyses) employed a modified rotating wedge retinotopy paradigm (see Methods for details). Datasets varied in the number of individual runs from 6 to 9, for a total of 59 unique runs. DS6 and 7 were acquired at 3T, while the rest were obtained at 7T. Unless indicated otherwise, the summary metrics provided in the results refer to the mean across 7 datasets (DS1- 7; DS8 was only considered for FFT analyses), normalized if necessary (i.e., due to different voxel sizes).

Spatial Characteristics of Extracted Noise

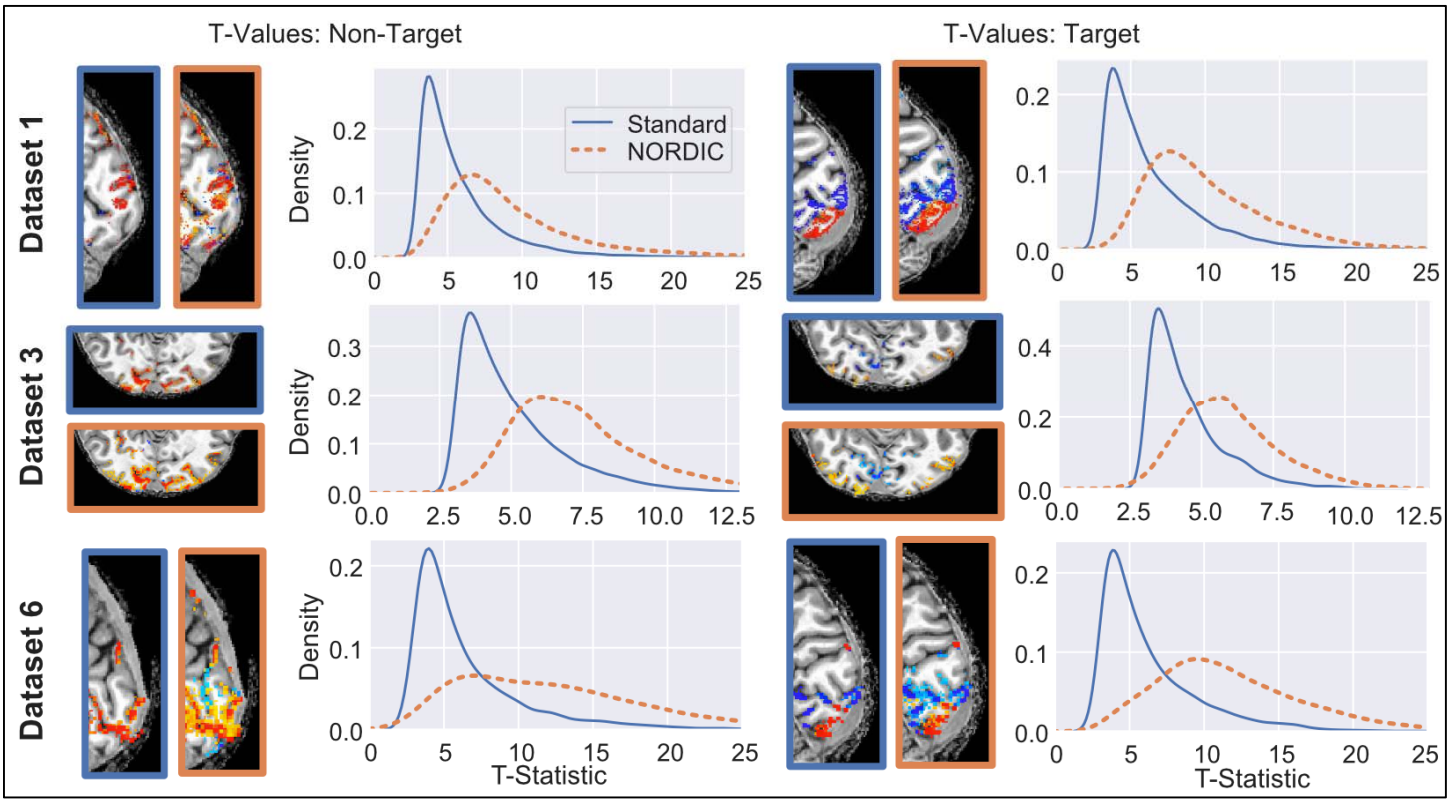
Following image reconstruction and denoising, we first examined the spatial characteristics of the noise amplification due to acceleration ('g-factor') as well as the noise removed from the data timeseries by NORDIC and dwidenoise. Figure 1 shows the images for 3 data sets comparing the temporal mean of the extracted noise from the first run with the maps of the g-factor produced from the raw k-space data. Anatomical boundaries are visible in the dwidenoise data, however the NORDIC processed data shows maps that are consistent with the g-factor maps.



Caption Figure 1. Comparing Residuals from Denoising Methods. The first column shows the selected views using the reconstructed EPI images. The next column shows the g-factor maps calculated from the raw k-space data. The last two columns show the temporal mean (absolute values) of the extracted noise from the first run for NORDIC (column 3) and dwidenoise (column 4).

Conventional General Linear Model (GLM) results

Figure 2 shows the distribution of t-values from the conventional GLM with a canonical HRF, which combined all individual runs within each of the datasets. Processing with NORDIC leads to an increase in the t-values, visible as a large rightward shift in their distribution relative to the data reconstructed with standard methods (hereafter Standard data). The t-values were extracted from an identical ROI, created based on the Standard data (See Methods). This increase in t-values is found within both the Target ROI (DS1,DS6: center > surround checkerboard; DS3: faces > scrambled) as well as the non-Target ROI (response to surround checkerboard and scrambled stimuli only). This effect is consistent across all 7 Datasets (Supplemental Figure 1); the mean of the one-sided t-statistic across datasets within the non-target ROI was 8.7±5.0 for NORDIC and 5.67±3.6 for the Standard Reconstruction. In the Target ROI these values were 9.9±4.5 for NORDIC and 6.16±3.6 for Standard.



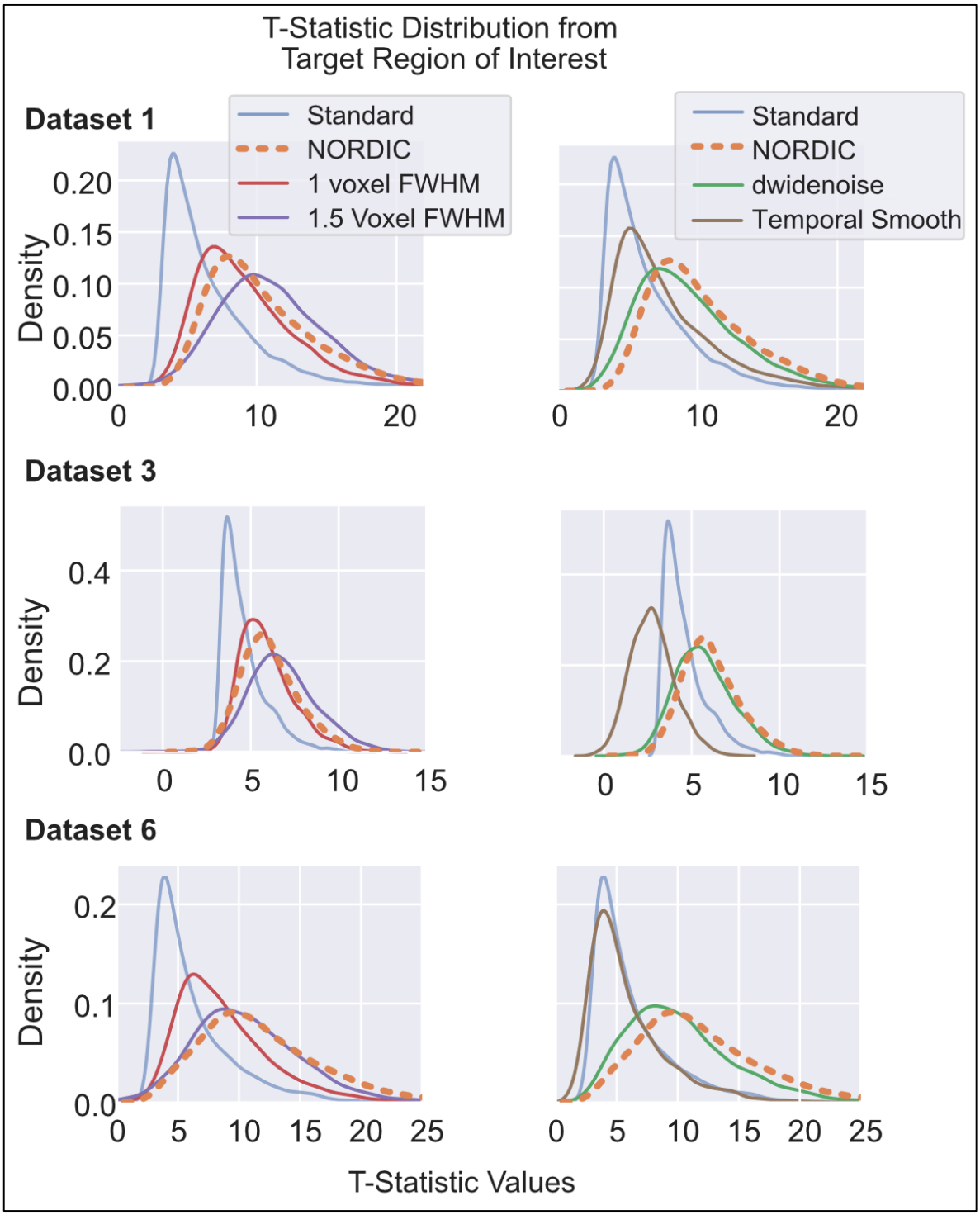
Caption Figure 2. The distributions of t-Statistics from the model using all runs of data are shown. Distributions show the t-statistics of voxels from Non-Target (left) and Target (right) ROIs, defined as those that displayed significant positive stimulus-evoked changes relative to baseline (Non-Target) or in the contrast between Target and Non-Target conditions (Target) in the Standard data. Functional maps of the corresponding contrast are shown for visual reference at a t-value threshold of 3.3, corresponding to voxel-wise $p<0.001$ (uncorrected). NORDIC and Standard reconstructed functional maps are identified by the color of the border of the two images shown for each dataset (blue=Standard, orange=NORDIC). Across all datasets and both ROIs, the distribution of t-statistics for NORDIC was higher, with a longer tail.

Comparison with Other Noise Reduction Methods.

In order to evaluate the relative performance of NORDIC compared to other methods that seek to reduce normally distributed noise, we ran identical general linear models (GLMs) on data that were spatially (1 and 1.5 voxels FWHM) or temporally (sliding window ~10s) smoothed following preprocessing. In addition, we used dwidenoise (38, 39), a method available in MRtrix3 (40) that attempts to suppress thermal noise using the properties of the Marchenko-Pastur distribution with local PCA (MPPCA), using the default settings on the Standard reconstruction. Here we only consider the Target ROI, which is based on the contrast center > surround (DS 1, 6) or faces > scrambled (DS3). T-Statistic distributions for DS1, 3, and 6 are shown in Figure 3. Results for all datasets are given in Supplementary Fig. S2. Distributions for NORDIC and Standard are identical to that given in Figure 2, with the mean across all datasets at 9.9±4.5 and 6.16±3.6, respectively. The mean for t-statistics for these other methods are as follows: 8.5±4.2 for 1 voxel of additional smoothing, 10.1±4.9 for 1.5 voxels of additional spatial smoothing, 9.2±4.5 for dwidenoise, and 6.6±4.4 for temporal smoothing (Supplemental Fig. S2). The reduction in noise afforded by the NORDIC reconstruction produces t-statistics that exceed 1 voxel of smoothing and are nearly equivalent to data smoothed by 1.5 voxels. However, the statistical gains from NORDIC processing do not come at the cost of blurring the spatial precision of the data, which is discussed in the next section.

Temporal smoothing appears to confer minimal benefits with respect to t-statistic distributions for the block designs used in DS1 and DS6 (Fig. 2); as can be expected, it begins to fail as a processing method when used on the fast event related design in DS3, yielding t-statistics that decrease and approach zero due to blending the events that are closely spaced in time. No such effect is seen in the NORDIC reconstruction, which is in fact right-shifted with no negative values. The

performance of dwidenoise approaches NORDIC, in terms of t-statistics but, as discussed later, has a complex spatial smoothing effect on the data.



Caption Figure 3. Distributions of t-statistics from alternative noise reduction methods within the Target ROI. Left column shows data from Standard, NORDIC and spatial smoothing with 1 and 1.5 voxel FWHM spatial smoothing. Right column compares the same Standard and NORDIC data against temporal smoothing and dwidenoise denoising. T-values were extracted from the Target ROI defined using the Standard data. The t-values obtained with NORDIC (Orange, dashed) processed data is comparable to the effects of an additional 1 or 1.5 voxels FWHM gaussian smoothing. While temporal smoothing (brown) did increase t-statistics for Dataset 1, note that for the fast event-related design (Dataset 3)

this led to a temporal blending of neighboring events, leading to positive and negative t-values, an effect not found in NORDIC data.

Effects of Denoising on Global Image Smoothness

AFNI's *3dFWHMx* was used on the data both prior to and subsequent to processing steps that corrected for slice timing and motion (labeled as Pre- and Post-Processing in Figure 4) and on the residuals from the conventional GLM (Figure 4) to estimate the smoothness introduced in the images by the various processing and denoising steps. For the pre- and post-processing examples, the data were detrended (including removal of the mean) to remove temporal drifts and variation in voxel intensities related to anatomy. These estimates of smoothness are based on the variance of differences between each voxel and its neighbors relative to the variance of voxels over the masked image (see Methods) and we refer to these estimates as global smoothness. These effective FWHM estimates are derived from geometric mean of the fit of a Gaussian + monoexponential decay model to each individual volume in the examined Datasets. Any spatial smoothing introduced by post-acquisition data manipulations shows up as an increase in FWHM relative to the Standard data.

The nominal resolution specified for image acquisition for these datasets was 0.8, 0.8, and 1.2mm isotropic, respectively. The FWFM measured in the Standard data before any processing were 0.887±0.002, 0.900±0.003, and 1.244±0.002, respectively, which is marginally higher than the nominal resolution specified for image acquisition. In examining the NORDIC datasets for the smoothness prior to any processing, we found a small increase in estimated smoothness associated with the NORDIC reconstruction (Table 1, Figure 4). The NORDIC data smoothness, relative to the Standard reconstruction values, corresponds to an average increase in estimated image smoothness of 5.13% across the 3 datasets shown in Figure 4. For dwidenoise, a much larger increase in smoothness was evident, with an estimated FWHM before processing corresponding to a 22% increase on average. Across all 7 Datasets (Supplemental Figure 4), NORDIC led to a 5.6% average increase in estimated

image smoothness, whereas dwidenoise led to a 16% increase in image smoothness, prior to motion correction and slice timing.

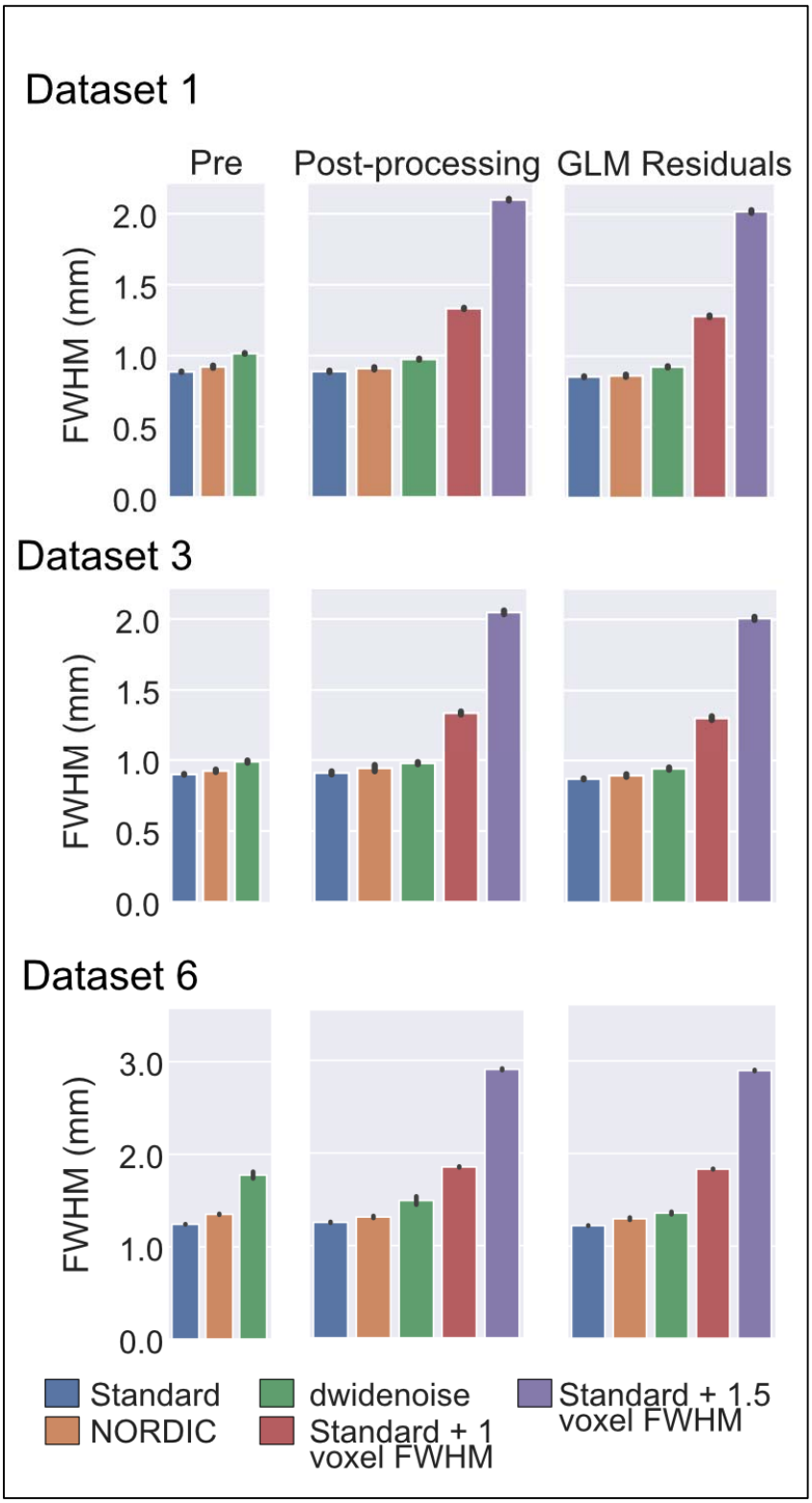
	Before Processing ('Pre')			Post Processing			Residuals		
	DS1	DS3	DS6	DS1	DS3	DS6	DS1	DS3	DS6
Standard	0.89±0.002	0.9±0.003	1.24±0.002	0.89±0.006	0.91±0.012	1.25±0.006	0.85±0.004	0.87±0.004	1.22±0.002
NORDIC	0.92±0.009	0.92±-0.01	1.35±0.006	0.91±0.009	0.94±0.024	1.31±0.012	0.86±0.009	0.89±0.01	1.3±0.014
dwidenoise	1.02±0.003	0.99±0.009	1.78±0.039	0.98±0.004	0.98±0.008	1.49±0.048	0.92±0.005	0.94±0.009	1.36±0.015
+1 Voxel FWHM	-	-	-	1.33±0.005	1.34±0.012	1.85±0.006	1.33±0.005	1.3±0.015	1.83±0.004
+1.5 Voxel FWHM	-	-	-	2.1±0.007	2.05±0.014	2.9±0.01	2.1±0.007	2.01±-0.012	2.9±0.006

Table 1. Estimated Smoothness in millimeters (FWHM). The first set of columns shows the estimated smoothness before any additional processing, within a brain mask made from the first run. The next set shows the estimated smoothness after motion correction and slice timing, with the bottom two rows reporting the effects of explicit, intentional smoothing. The final set of columns shows the estimated smoothness of the residuals from the conventional GLM. Across all processing timepoints the NORDIC data is minimally smoother than the Standard data. Values are mean across runs, plus/minus standard deviation.

Following the pre-processing steps, which were identical for all subsequent applications of "denoising", smoothness estimates remained similar for the 3 datasets reported here (Column 2 Table 1, Figure 4). The mean increase in estimated smoothness for all 7 datasets, relative to the Standard post-processed data, was estimated to be larger by 3.3% for NORDIC, 9.3% for dwidenoise, 51% for 1 additional voxel of smoothing, and 140% for 1.5 voxels of smoothing.

Following a conventional GLM, the mean increase in estimated smoothness of the residuals for all 7 datasets, relative to the Standard post-processed data, was 3.7% for NORDIC, 8.0% for dwidenoise, 52.7% for 1 additional voxel of smoothing, and 142.8% for 1.5 additional voxels of smoothing. Note that the estimated smoothness of the GLM residuals is often lower than the corresponding smoothness estimates on data that were merely detrended or pre-processed. This is likely because the GLM removed strong stimulus-induced spatial correlations that inflated the estimates of smoothness computed across the entire volume.

For all processing stages, the increase in estimated smoothness of NORDIC was significant (all $p \ll 0.001$), as was the increase in estimated smoothness due to dwidenoise (all $p \ll 0.001$). In addition, NORDIC was significantly less smooth at all stages, compared to dwidenoise processed data (all $p < 0.001$).



Caption Figure 4. Estimated spatial smoothness in mm (FWHM) at various processing stages for each method. Prior to processing (left), NORDIC results in an average increase in 5.1% in smoothness and dwidenoise results in an increase of 22.4% on average. After processing, but prior to the GLM (middle) this trend remains. Note that the image smoothness of

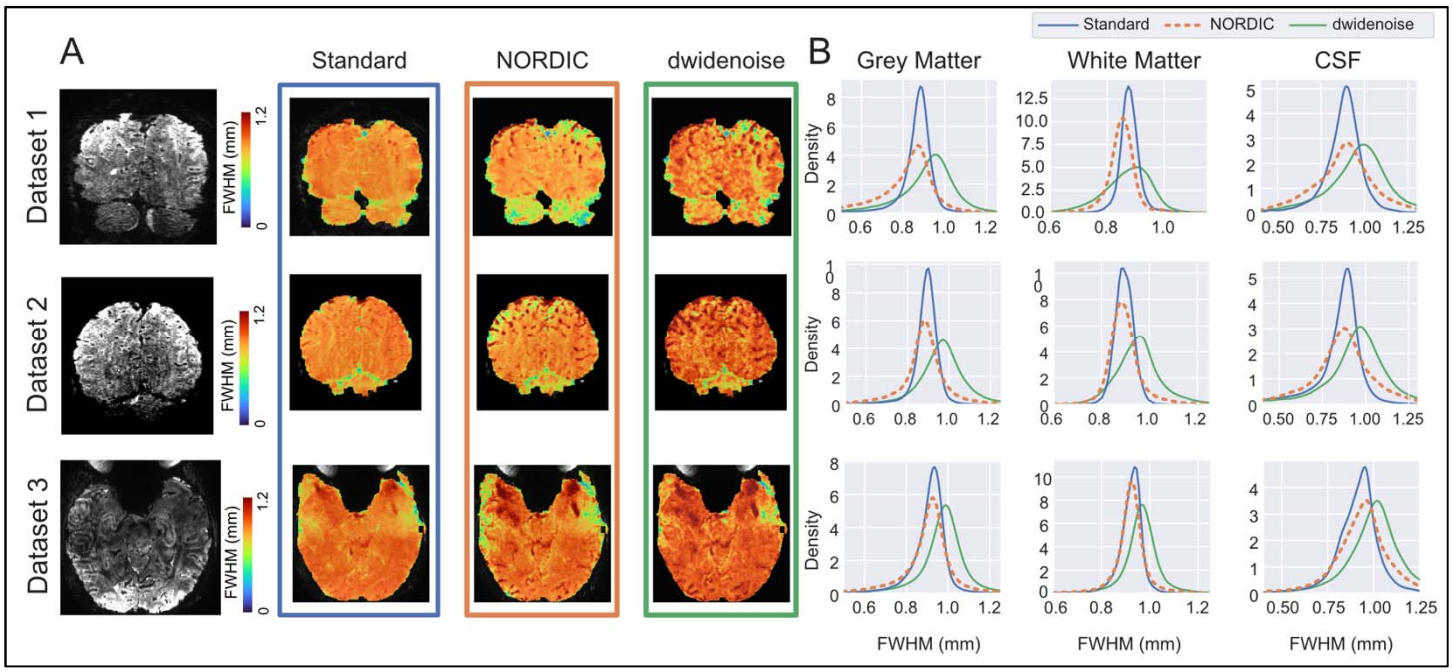
the Standard, NORDIC, and dwidenoise data are substantially below the level of the additional 1 or 1.5 voxels of additional smoothing. These trends remain the same for the residuals (right) after the conventional GLM.

Effects of Denoising on Local Image Smoothness

Local smoothness estimates derived from AFNI's *3dLocalACF* are presented in Figure 5. We focus on the 0.8mm high resolution Datasets (DS1 – DS3) for which spatial precision is most important. Local smoothness is derived from the *residuals* of the conventional GLM, by fitting the spatial decay of the correlation of each individual voxel's timeseries with its neighbors, within a spherical neighborhood. In other words, unlike the previously described "global smoothness" which is estimated within a given volume, this "local smoothness" estimate represents a value of how each voxel is correlated with neighboring voxels in the times series, using the mixed gaussian + monoexponential fit. Smoothness is reported as the FWHM, in mm, of this fit. The first column in Figure 5 shows the slice presented in the subsequent columns. Visual inspection of the next 3 panels shows that the local smoothness varies across the brain and tissue classes in the residuals of the task GLM; however, these effects are minimal in the Standard data. Following NORDIC processing, regions corresponding to white matter have similar or reduced spatial autocorrelation values, whereas gray matter is more variable across these presented datasets. Following dwidenoise, there is a general increase in the local FWHM estimates across the brain. The distributions of the local smoothness estimates for all voxels within each tissue class from segmentation are shown on in Figure 5B. Across the three high resolution datasets shown (0.8mm) , the mean FWHM with the gray matter mask was 0.87±0.02 for Standard, 0.87±0.05 for NORDIC and 0.96±0.04 for dwidenoise. The mean FWHM in white matter was 0.90±0.02 for Standard, 0.89±0.04 for NORDIC and 0.93±0.06 for dwidenoise. The mean FWHM in CSF was 0.88±0.04 for Standard, 0.91±0.04 for NORDIC and 0.98±0.03 for dwidenoise (see Supplemental Table 1 for individual dataset values).

These values are consistent with global smoothness estimates but provide a more nuanced view of what is contributing to overall image smoothness, as it is expected to vary by tissue class.

Across all datasets, relative to the Standard data, the mean local smoothness estimates for NORDIC were 3.6% smaller in gray matter, 12.2% smaller in white matter and 12.1% larger in CSF (all $p<0.001$). For dwidenoise, smoothness estimates were 2% greater in gray matter, 12.1% smaller in white matter and 13.4% larger in CSF (all $p<0.001$). For all tissue classes, NORDIC had significantly lower estimates of local smoothness relative to dwidenoise (all $p<0.001$).

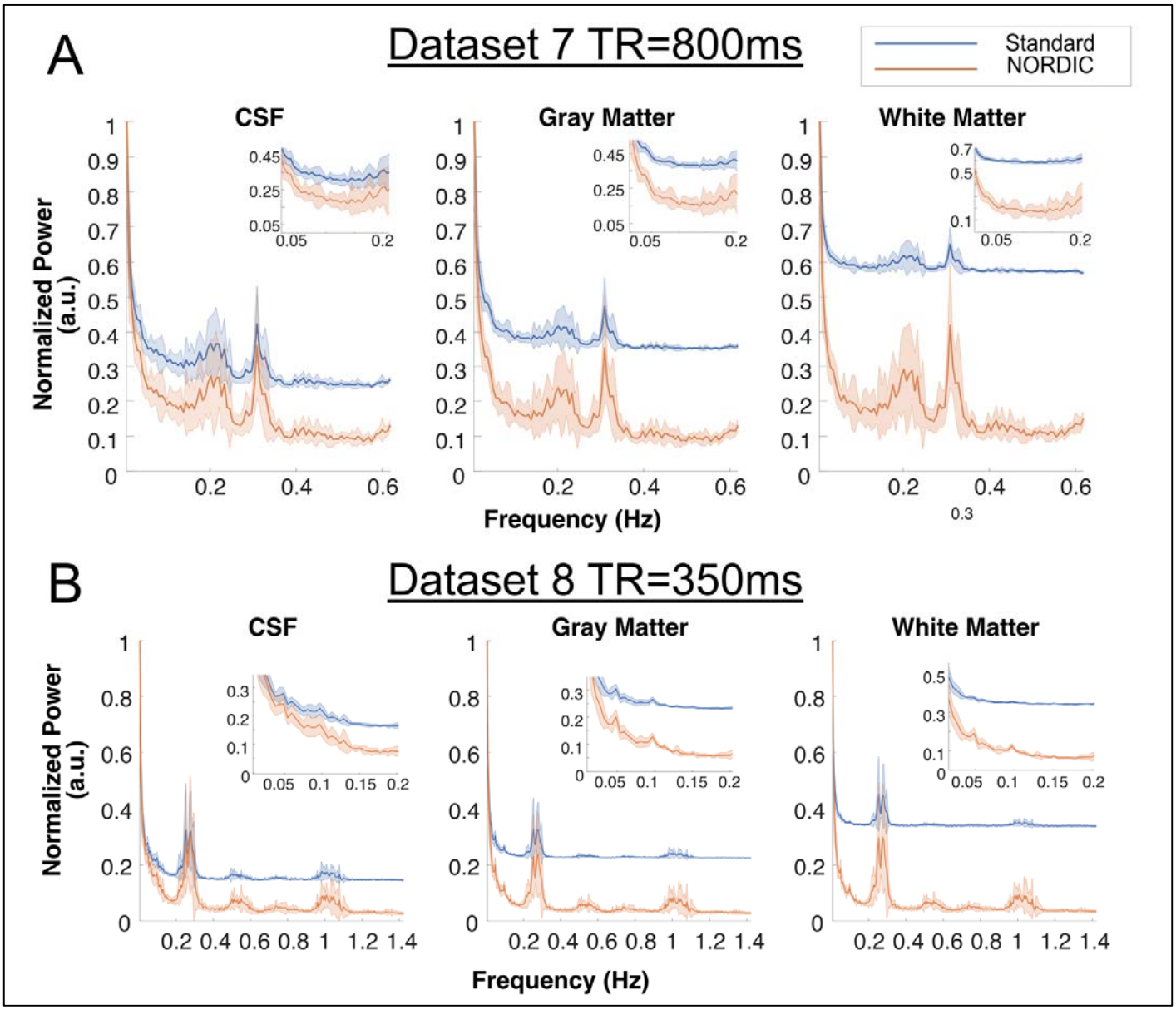


Caption Figure 5. Local smoothness estimated from GLM residuals, across all runs. **A) Selected slices and local smoothness estimates, in FWHM mm.** The leftmost panel shows the selected EPI slice. The next three panels show the estimated, voxel-wise (local) spatial smoothness for the three different processing methods, Standard (blue border), NORDIC (orange) and dwidenoise (green), with the scales identical between the different processing types. Note that the local spatial smoothness is often highest in dark areas of the EPI image, likely associated with veins. **B) Full distributions of voxel-wise local smoothness estimates within different tissue classes.** These kernel density estimates show the distributions of the local spatial smoothness estimates in tissue classes derived from a T1-weighted anatomical image for Standard (blue), NORDIC (orange) and dwidenoise (green). Local smoothness is somewhat decreased following NORDIC, except within the CSF mask. All datasets had a prescribed resolution of 0.8mm isotropic.

Evaluating temporal precision

Fourier Analysis: The normalized power spectra for DS7 and DS8 show that denoising by NORDIC results in a wide reduction in power across frequency bands, with the effects most apparent at higher frequencies where thermal noise is the dominant noise source (Figure 6). This effect is most visible in DS8, which was collected at a TR of 350ms. At this sampling rate, the frequency associated with cardiac noise (~1Hz) is clearly visible in the data and clearer after NORDIC processing. The task related frequency (0.05Hz) is also more pronounced after NORDIC for DS8. Though this effect is

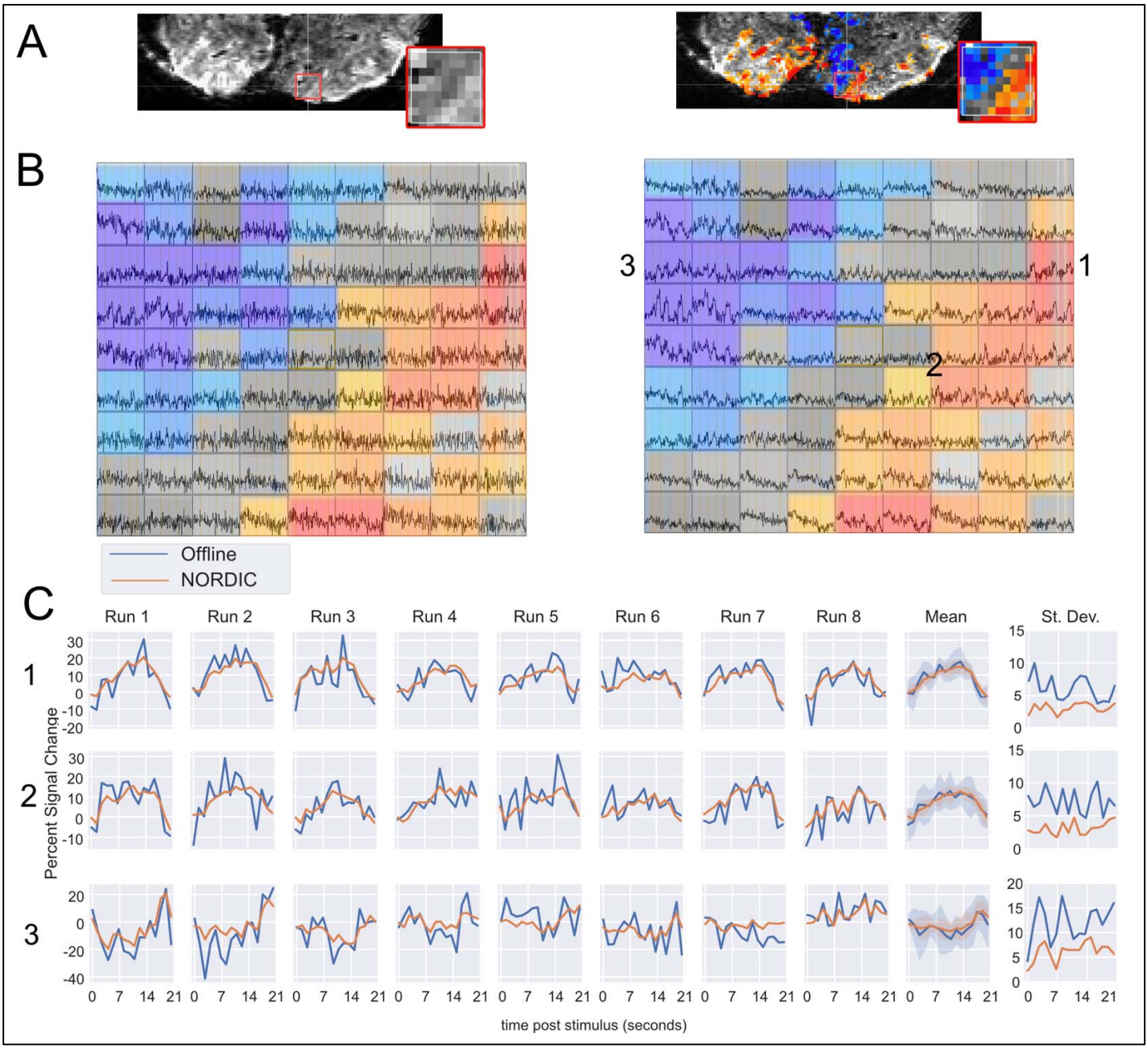
most visible in the gray matter partition, it is also visible in the white matter. This is likely a combined effect of both partial volume effects and imperfect ROI overlap between the anatomically derived tissue segmentation and the distorted EPI images.



Caption Figure 6. Frequency Plots from FFT analysis. **Panel A)** Normalized power spectra from DS7, a 3T, HCP-like acquisition with 800ms TR. NORDIC processing reduces power throughout most frequency bands, with effects such as a task harmonic (~0.2 Hz) and the respiratory band (~0.3 Hz) becoming more visible. Insets show power from 0.01 through 0.2. **Panel B)** Normalized power spectra from DS8, a rapidly sampled acquisition with 350ms TR. The effect of NORDIC in broadly reducing power remains pronounced throughout higher frequencies. Here the respiratory and cardiac signals are clear at ~0.3 and ~1 Hz respectively. In the gray matter, a clear peak at 0.5Hz, corresponding the task frequency is also clearer after NORDIC processing (see inset). Shading shows standard deviation across independent runs for both A and B.

Cross Validation: Following NORDIC processing, estimates of single-run and single voxel HRFs were markedly improved. The HRF estimates were produced by using an unbiased finite

impulse response (FIR) model, which constructs TR-by-TR estimates of the response to the events in the stimulus, estimated for each run independently. We find that following processing with NORDIC, the FIR time courses show more plausible hemodynamic-like responses and less error across runs. To see the origin of these improved effects, we consider 81 voxels selected from DS1 focusing on an area that features both target and non-target sensitive voxels. Figure 7A (top row) shows the underlying data and activation maps from the conventional GLM across all runs of the Standard processed data, highlighting in the inset containing the 81 voxels considered for Figs. 7B and 7C. Visual inspection of individual voxel time courses (with activation map overlaid) from a single run (the first run) of DS1 in Fig. 7B shows that the reduction in noise from the NORDIC method (right panel) does not lead to a spread of the activation (consistent with the negligible change in spatial smoothness) but instead reduces the noise level such that stimulus-coupled signal changes becomes more visible. The selected 81 voxels contain responses to the target (indicated by #1, #2), responses to target and surround (within grey boundary), and responses only to the surround (indicated by #3). Despite identical voxels being selected for the Standard data (left panel), stimulus-evoked responses are difficult to see. Though the spatial maps presented in Figure 7A (upper row) was derived from the full 8-runs, the task events are visible in the individual voxels of the single run after NORDIC processing (Fig. 7B, right panel) but generally not in the Standard. Section C shows response estimates from an unbiased finite impulse response (FIR) model for the selected voxels, showcasing improvements in single-run, single voxel FIR estimates.



Caption Figure 7. Example of activation, improved time courses and FIR estimates in run 1 of DS1, Panel A, upper row Different views of the area under consideration from Standard reconstruction**:** Left column shows the mean GE-EPI values for the area under consideration. Right shows the activation amplitude (-12 to 12 % signal change) in the selected slice for all runs, Center > Surround from the Standard data (map threshold from t-stat ≥ 3.3). Inset boxes in panel A show the 81 voxels considered in Panel B. **B) Time courses from first run for 81 voxels.** To visualize task responsive voxels, we shade them based on the contrast from all runs of Standard data, as seen in Panel A, right. The stimulus-evoked signal amplitude changes associated with the three surround and the three target stimulus epochs are clearly visible in the NORDIC processed timeseries of the corresponding voxels but are largely invisible in the Standard data due to high noise levels. **C) Single Run FIR Estimates for the Target Condition.** Responses to the target (center) are illustrated in selected voxels 1 and 2 for individual runs are shown. The final columns show the across-run average and standard deviation respectively. shading in the latter plot shows standard deviation from the mean, which is also plotted separately

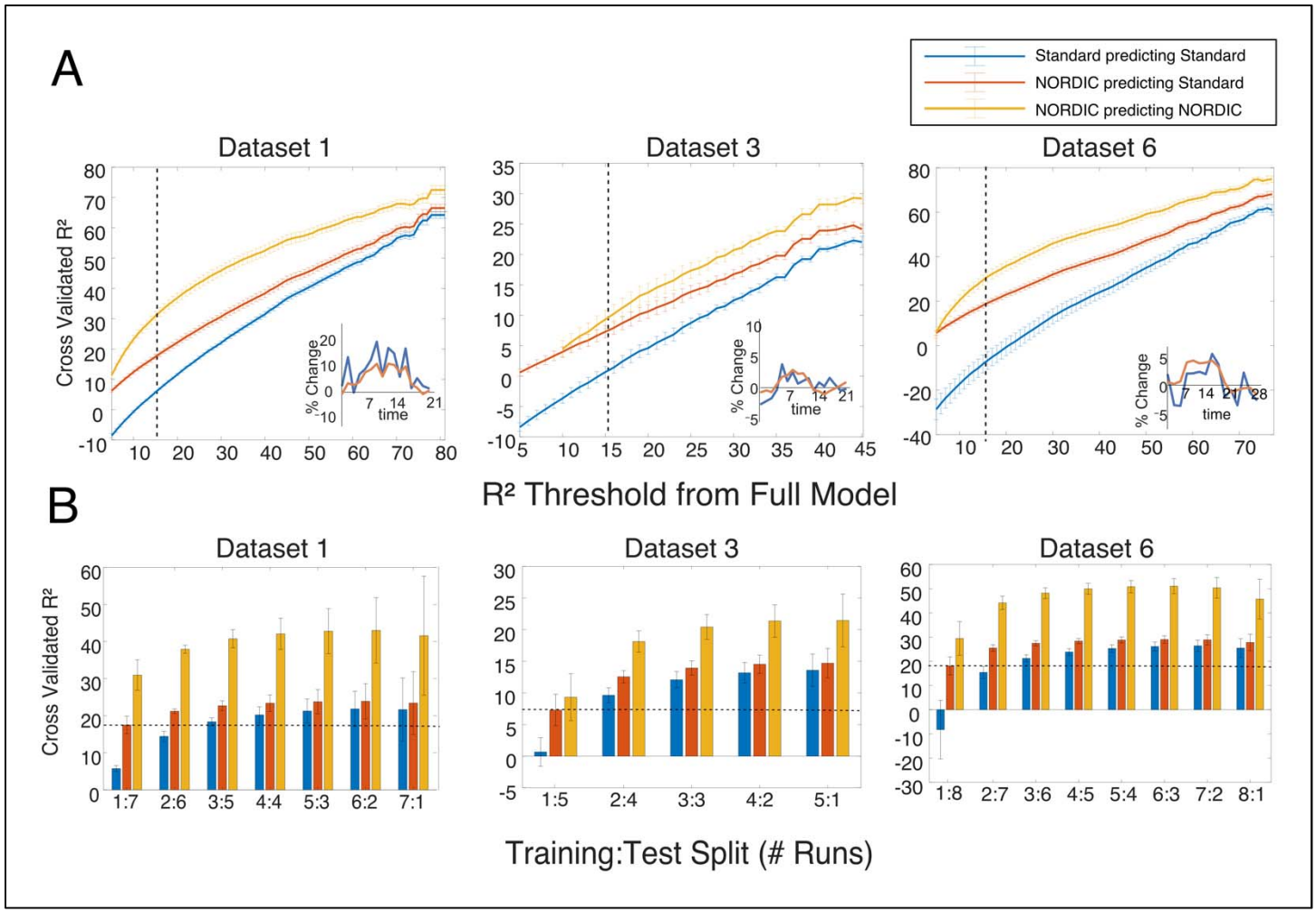
for clarity (note that the deviation associated with the Standard data far exceeds that of NORDIC). Voxel 3, which is sensitive to the surround condition, remains closer to the expected zero amplitude (i.e. non-responsive). This is particularly true for NORDIC processed data, which is associated with lower standard deviation.

To further investigate whether NORDIC processing does indeed preserve the inherent signal present in the fMRI data (i.e., only suppressing thermal noise), we considered an exhaustive K-fold cross validation scheme to determine how effectively FIR model results could predict held-out timeseries quantified with the coefficient of determination, R^2 (See Methods). In order to limit this analysis to putatively task related voxels we first produced a map of R^2 values, derived from a cross-validated (leave-one-out) model using all runs and a canonical HRF. One map for each dataset was produced, and the same map of voxels was used to evaluate each of the three prediction methods.

Panel A in Figure 8 shows the performance of a *single* run of Standard and NORDIC data in predicting the full timeseries of held out data. In general, the R^2 metric increases as voxels with better predicative accuracy in the full model are included, with all lines tending to increase from left to right. Notably, however, a single run of NORDIC is better able to predict the held out runs of Standard data compared to the Standard data itself. Further, this benefit is maintained even for voxels that had a high signal to noise ratio (i.e., far right of the graph). Exemplar estimates from the finite impulse response (FIR) model which produces an unbiased estimated of the HRF are shown for a single voxel from a single run in the lower right of each graph in Figure 8A. As expected for such high-resolution data, the estimates from the Standard data (blue) are noisy. However, following NORDIC processing, these single run estimates show clear HRF-like properties.

In order to quantify this improvement across all voxels and runs we considered a K-fold cross validation where $1 \leq K <$ Number of runs (Figure 8B). Using a threshold of voxels that were able to explain 15% of the variance in the full model (vertical lines in panel A), we can see that one run of NORDIC is able to predict a held out timeseries as well as 2 to 3 runs of Standard data (horizontal dashed line, Panel B), and two runs of NORDIC are nearly able to predict as well as any number of

Standard runs combined. Across all thresholds and folds, NORDIC processed data is also always able to better predict the timeseries of data that has undergone NORDIC processing - which would be the typical use case. These findings apply to all datasets considered (Supplemental Figure S5).



Caption Figure 8. NORDIC processing leads to higher cross-validation performance, predicting from a finite impulse response model. **Panel A)** Exhaustive cross validation performance when training on using only one run. Cross validated R^2 is shown for training on Standard data and predicting held-out Standard data (blue), training on NORDIC data and predicting held-out NORDIC data (yellow), and training on NORDIC data and predicting held-out Standard data (Orange). X-Axis indicates voxel inclusion threshold derived from leave-one-out cross-validated R^2 using a canonical HRF on Standard data. Insets show example single voxel single run FIR model estimates for Standard (blue) and NORDIC (Orange). NORDIC processing can produce estimates that better predict Standard data compared to Standard data itself. Error bars are standard error over permutations. Dashed lines show an R^2 threshold of 15% used in panel B. **Panel B)** K-Fold training was repeated for all Ks greater than 1 and less than the number of runs. Colors are as above; the number of runs included in the training vary across the X-axis, with bar height reflecting the R^2 obtained. Dashed line indicates the performance of training on one run of NORDIC data, which is equivalent in cross validation performance to using 2 or 3 runs of Standard data. Including more data allows Standard models to approach, but not reach 2 to 3 runs of NORDIC data. Error bars again show standard error across permutations.

Discussion

Most applications of the MRI method are highly SNR limited. As such, it is common to perform some sort of noise mitigation procedure in post-acquisition data processing in order to increase the SNR of the data. Although the ultimate goal is to do so without compromising any information, generally there is a trade-off, such as effective spatial and/or temporal resolution. For example, spatial smoothing, which is an extremely common strategy employed to increase the SNR, achieves SNR gains by averaging the data over spatial coordinates of the image, thus inducing blurring and reducing spatial precision. Denoising methods that try to characterize the components of the data and selectively remove some of them also potentially run the risk of misidentification, removing signal components that contain information as well as noise. Therefore, it is imperative that a careful and critical evaluation is performed when deploying such an approach, especially when substantial and potentially transformative gains are promised for the field of interest, as is the case with the application of NORDIC to fMRI. In this paper we present such an evaluation for NORDIC.

The recently developed NORDIC denoising method was evaluated on a variety of field strengths, voxel sizes, TRs, and stimulus designs. We find that NORDIC-processed data removes noise that matches g-factor maps (Fig 1), producing much higher t-values under a conventional modeling framework (Figs. 2, 3), consistent with the prior report (36). These fMRI t-statistics are approximately equivalent to those produced by smoothing the data using a kernel of 1.5 voxels; however, analysis of the NORDIC-processed data does not show any such increase in smoothness (Figs. 4,5). The frequency spectrum of data following NORDIC shows a widespread reduction, consistent with white-noise suppression (Fig. 6). These SNR gains are also reflected in markedly improved single run, single voxel FIR estimates, which in turn produce better predictions of held-out, non-denoised Standard data (Fig. 7,8), compared to the Standard data itself. Collectively these findings support the use of NORDIC for a wide variety of fMRI applications, ranging from HCP-like data obtained at 3T to cutting edge high-resolution fMRI data acquired at ultrahigh magnetic fields.

Conventional GLMs

The t-statistic is widely used in the fMRI literature to report the statistical veracity of signal increases and decreases as spatial maps. Consistent with our prior report (36), we find that NORDIC-processed data had substantially larger t-values compared to the Standard data alone (Figure 2). These gains are similar to or greater in magnitude than those reported by others using denoising methods to remove structured noise such as multi-echo denoising (41, 42), ICA-based denoising strategies such as ICA-AROMA (43) or using SNR efficient accelerated imaging sequences like SMS/MB (16) to collect more data in a given period of time (44). However, NORDIC is a complement rather than a replacement to these methods as it focuses on suppressing thermal noise. Of the other processing methods examined here, the performance of NORDIC with respect to t-statistics exceeds all but the 1.5 voxel smoothing (Figure 3), even in the data with relatively large voxels (i.e., 2mm isotropic resolution 3T HCP protocol). Of course, as previously mentioned, NORDIC accomplishes this without loss of spatial precision. NORDIC also out-performed the benefits one would get with temporal smoothing, even in cases where long duration 12s events were separated by long (i.e., 12s) inter-stimulus intervals. While NORDIC does produce a timeseries that is less corrupted by thermal noise, it does not average over a temporal window or compromise temporal information. This is most clear in DS3, which used a fast event related design. Following temporal smoothing, the t-values for the face condition in this design decreased. This reflects the mixing of neighboring events due to the short ISI of 2 seconds. The opposite effect is found following NORDIC processing, in other words t-values increased, and negative t-values were absent. This finding indicates that the temporal precision is maintained with NORDIC and that temporal blurring is not occurring.

These findings present the possibility of new avenues of research, ultrahigh spatial resolution studies, the use of smaller ROIs for ROI-based analysis, or examining single-trial response estimates (45), which represents an important but often SNR-starved analysis strategy.

Image Smoothness.

Despite the similarity of the distributions of the t-values between NORDIC and smoothed data, no corresponding increase in smoothness is observed for NORDIC (Figure 3). In fact, at its maximum, NORDIC only increased the estimated smoothness by 6.1%. This is smaller than the effect of conventional preprocessing methods, which are known to produce images with greater smoothness characteristics due to the need to interpolate values on a new image grid (46).

Here we considered estimates of global smoothness at all stages of processing. While typical smoothing estimates use the residuals of the data as an estimate of the overall smoothness of the noise, it is possible for these estimates to be oversized. For example, coherent areas of signal change could remain due to a mismatch between the canonical HRF and the subject's response. The patterns of smoothness reported here are consistent between processing stages, supporting the argument for minimal image smoothing due to the NORDIC method. This global measure of image smoothness is often used in the context of cluster correction, as it is thought to reflect the underlying smoothness of the acquired image or image timeseries (47). The spatial autocorrelation methods (both global and local) to estimate smoothness in this work are different from approaches that estimate a functional point spread function (48), which instead attempt to quantify the functional precision available in the maps of functional responses. For the latter, prior work found that NORDIC had no impact on the functional point spread of the BOLD signal (36). Together, these reports show that NORDIC is able to suppress thermal noise without excessive blurring or altering the inherent spatial structure of the functional image.

This finding is further corroborated when we examine the local smoothness of the residuals. This method considers the spatial fall-off of the local temporal correlations in the residuals from the GLM. NORDIC processing produced local smoothness estimates that were nearly equivalent or less than those of the standard data (Figure 5). The exception is within a CSF mask, which includes features such the superior sagittal sinus. In the EPI data, this would be associated with low raw signal

values, and thus a high proportion of thermal noise. When the thermal noise is suppressed by NORDIC, it unmasks higher local correlation due to the BOLD effect associated with pial veins as well as other physiological processes such as pulsations due to heart beat and respiration. These local smoothness findings are different from the global smoothness estimates, in that, on average, NORDIC processed data had less estimated smoothness in both gray and white matter. One possible source of this difference is likely due to the difference in analytical methods. For example, global smoothness estimates consider each volume independently, in effect examining the variance across space. In contrast, local smoothness considers the variance over the timeseries correlation within a local neighborhood. It is plausible that neighboring voxels could have highly correlated timeseries, despite large differences in signal magnitude (i.e., high spatial variance), such as at gray/white matter boundaries due to partial volume effects. It is also possible that the global smoothness after NORDIC seemed slightly increased in the GM due to effects similar to those described above for the CSF mask. Nevertheless, both global and local smoothness estimates provide strong evidence that NORDIC processing is not leading to meaningful increases in smoothness, in agreement with an examination of the functional point spread (36).

Temporal Precision.

The use of NORDIC produced fMRI voxel time courses in which responses to task events were more visible and not subject to any apparent smoothing (Figure 5). We first examined the normalized power spectrum of DS7 (800ms TR) and DS8 (350ms TR). For both datasets, task and physiologically related frequencies are clearer after NORDIC processing, relative to the Standard data. In general, there is a large reduction of power at nearly all frequencies, consistent with the reduction of normally distributed noise. This corresponds to an increased ability to identify and resolve frequencies associated with physiological noise in each individual voxel timeseries.

We then used the unbiased estimates of the hemodynamic response function (HRF), produced by finite impulse response (FIR) models, in order to simultaneously examine the denoising

performance of NORDIC and whether this resulted in any loss of temporal information. NORDIC produced FIR estimates that were associated with less cross-run variability (Figure 7). In many cases, particularly in high-resolution studies, these types of response estimates are produced by simultaneously modeling multiple runs or averaging the signal within an ROI. Here, however, we show that single-run estimates are reliable, even at the single-voxel level.

The primary concern is that these estimates are the result of suppressing both signal and noise. That is, the process of removing thermal noise has also removed signal and, ultimately, temporal information. In the NORDIC data, this is not the case, as shown by the fact that these FIR estimates accurately reconstruct data that were held-out from the model. This was performed in an exhaustive K-fold fashion, considering all combinations of $1 \leq K <$ Number of runs. Based on the coefficient of determination (R^2), not only was NORDIC data better able to predict held out NORDIC data, but that it was also better able to predict held-out Standard data (Figure 8). This is a critical feature in considering the performance of a denoising method and, of course, is not always achieved. For example, large amounts of spatial smoothing will lead to increased t-values, but the voxel-wise HRFs derived from smoothed data will no longer correspond to the precise spatial location of voxels in unsmoothed data. Across all datasets, the NORDIC data were better able to predict the Standard data as measured by the coefficient of determination, including datasets using acquisition methods for which the thermal noise contribution is lower (i.e., datasets with larger voxels) relative to physiological fluctuations.

The NORDIC processed data were also able to better predict NORDIC timeseries. While less critical than the above demonstration of signal preservation, this indicates that the effects of NORDIC are consistent from run to run. In this context, one (DS1, DS6) or two (DS3) runs of NORDIC have better cross validated performance using voxels that survive the 15% R^2 threshold (Figure 8) than any number runs of Standard data. As the typical fMRI experiment would employ similarly denoised data throughout all analyses, rather than testing against the standard data (as was done here for validation), these large SNR gains represent the expected benefit of using NORDIC. Since NORDIC

denoising is done for each run separately, the data from separate runs remains statistically independent. This, in conjunction with the large gains in cross validated performance may allow new analysis approaches, such as voxel-wise analyses, which previously required large regions of interest. Furthermore, these improvements could translate to shorter scanning times, with many added advantages, for example decreasing the possibility of motion and time burden for participants or patients. Additional work is also required to examine how NORDIC complements denoising strategies that remove structure noise, such as ICA-AROMA (43) or those that denoise on the basis of multiple runs, such as GLMDenoise (49).

Although a large variety of datasets were considered in this work, including different TRs, voxel sizes, event designs, stimulus categories, and field strengths, the present work only evaluated gradient echo BOLD functional imaging, by far the most commonly employed strategy for functional imaging. The principles of NORDIC are expected to work equally well with other approaches of functional mapping, such as spin echo (SE) based BOLD fMRI (e.g. (50)), or functional mapping based on non-BOLD contrast mechanisms such as blood flow changes (e.g. ASL (51) and VASO (52)). NORDIC will likely be more useful for these other functional imaging approaches since they inherently have poor sensitivity and at any given spatial resolution will be more limited by thermal noise associated with the MR measurement compared to GE BOLD fMRI.

Comparison with Alternative Methods

Widely used methods to remove thermal noise have only recently developed and evaluated for diffusion MRI. In functional imaging, the growing interest in higher and higher resolutions and the capability of collecting this data with reasonable sampling rates has led to increased attention to the thermal noise contribution. While thermal noise is not typically the dominate noise source in fMRI, with voxel volumes below approximately $3mm^3$ at 7T (14) thermal noise begins to dominate and therefor meaningfully corrupt signals of interest.

Choices for the reduction of thermal noise are limited, and functional neuroimaging has primarily depended on averaging or spatial smoothing. Averaging requires large time commitments and can be complicated by difficulty in aligning across multiple runs, sessions or participants, while spatial smoothing with gaussian kernels unavoidably leads to a loss in spatial precision which is often the expressed purpose of high-resolution fMRI. While more advanced smoothing methods have been developed, which constrain smoothing on the basis of anatomy (53, 54), these methods are associated with a tradeoff – for example averaging across cortical depth may allow for high resolution analyses across the surface, but necessitates the loss of depth dependent activity profiles.

For diffusion imaging, which is associated with low SNR even with larger voxels, the picture is somewhat different. Though the method considered in this manuscript, dwidenoise, was developed primarily to suppress thermal noise in diffusion imaging (38, 39), it has recently been used for resting state (55, 56) and evaluate for presurgical mapping (57). Despite this usage, there has only been a limited examination of the cost/benefit tradeoffs, with no publications examining higher (sub-millimeter) resolutions. Here we find that dwidenoise does offer improvements in typical task-based activation measures such as t-statistics, however, this appears to be at the cost of increases in estimated image smoothness. This is most apparent for the 0.8mm high-resolution datasets (DS 1, 2, 3), for which precision is most desired (Figure 2).

Conclusion

The NORDIC method is a promising new tool which can be applied to functional imaging data in order to decrease the contribution of thermal noise. This is of particular importance for ultra-high spatial resolution functional neuroimaging data targeting mesoscopic scale organizations at high field strengths, which is SNR starved even after extremely long data acquisitions. Data processed with NORDIC results in substantial gains in t-values without a meaningful increase in image smoothing. In addition, NORDIC data preserves the voxel-wise temporal information and is better able to predict

held out data. These findings suggest that NORDIC is an effective method to increase the functional contrast-to-noise ratio of fMRI, thereby improving HRF estimates and/or permitting reduced fMRI acquisition times – potentially enabling entirely new study designs and statistical approaches to data analysis.

Material and Methods

Stimuli and Datasets

For this manuscript, 8 datasets are considered. For all stimuli, participants viewed the images though a mirror attached to the head coil. Datasets 1, 2, 5, 6, and 7 (DS1, DS2, DS5, DS6, and DS7) are block designs which used a flashing checkerboard (8Hz) positioned either in a center position (‘target’) or in a surround with the center cut out (‘surround’), centered on a gray background. The center stimulus subtended approximately 6.5 degrees of visual angle, as did the width of the surround border. Stimuli were presented in a standard 12 seconds on 12 seconds off block design paradigm. At 7T, stimuli were presented on a Cambridge Research Systems BOLDscreen 32 LCD monitor positioned at the head of the scanner bed (resolution 1920 x 1080 at 120 Hz), whereas at 3T the stimuli were presented using a NEC NP4000 projector, using a projection screen placed at the end of the bore of the MR scanner (resolution 1024 x 768 at 60 Hz).

For Dataset 3 (DS3), we used an event related design in which full-color, intact and phase scrambled faces were presented. Stimuli were centered on a gray background. Stimuli were on screen for 2 seconds and separated by at least a 2 second interstimulus interval. For all runs, blank trials (2 per run) were included to jitter the stimulus presentation. For some runs, the ISI was randomly selected from a Poisson distribution, and ranged from 2 to 6 seconds. In addition, for 2 runs, the participant was instructed to consider changes in the fixation color and press a button when it turned red. In the remaining 4 runs, the participant pressed a button each time a face was on screen indicating the gender perceived. The effect of these factors is not considered in the present

work. All runs and reconstruction methods were subjected to identical processing and analysis steps (below).

For Dataset 4 (DS4) we used grayscale images of faces (20 male, 20 female), presenting neutral expressions. We manipulated the phase coherence of each face from 0% to 40% in steps of 10%, resulting in 200 images (5 visual conditions x 20 identities x 2 genders). We equated the amplitude spectrum across all images. Stimuli approximately subtended 9° of visual angle. Faces were cropped to remove external features, with edges smoothed into the gray background. Stimulus presentation began and ended with a 12 second fixation period, and had a duration of approximately 3 mins and 22 secs. Within each run, we showed 40 images (5 phase coherence levels x 4 identities x 2 genders) presented for 2 seconds, with a 2 second interstimulus interval (ISI) as well as 10% blank trials (i.e., 4 seconds of fixation) randomly interspersed amongst the 40 images, effectively jittering the ISI. Stimulus presentation was pseudorandomized across runs, with the only constraint being the non-occurrence of 2 consecutive presentations of the same phase coherence level.

For Dataset 8 (DS8), we used a modified rotating wedge retinotopy paradigm. The frequency of ring sweeps was approximately 0.05Hz. The subject maintained fixation on a central point throughout the task. The wedge extended from the central fixation point to the edge of the screen, with a width of 20 degrees.

For DS1- 7 , stimulus presentation was controlled using Psychophysics Toolbox (3.0.15)-based scripts on a Mac Pro Computer. For DS8, the stimulus was controlled using custom, inhouse developed software.

MRI Acquisition

All functional MRI data were collected with either a 7T Siemens Magnetom System with a single channel transmit and 32-channel receive NOVA head coil or a 3T Siemens Magnetom Prisma system using the Siemens 32-channel head coil. All functional images were obtained using T2*-weighted, multislice (SMS)/Multiband(MB) gradient echo, Echo Planar (GE-EPI) (16).

7T fMRI Data (DS1 to 5, 8). For Datasets 1 and 2 (DS1, DS2), imaging was restricted to the posterior occipital lobe, using a right to left phase encoding direction (42 slices, TR=1.35s, MB=2, GRAPPA=3, 6/8 Partial Fourier, 0.8mm isotropic, TE=26.4ms, Flip Angle=58.2°, Bandwidth=1190Hz/pixel, Echo Spacing=1.00 ms). Dataset 3 (DS3) images captured the occipital pole and ventral temporal lobe using an anterior to posterior phase encoding direction (42 slices, TR=1.4s, MB=2, GRAPPA= 3, 6/8 Partial Fourier, 0.8mm isotropic, TE= 27.4ms, Flip Angle =78°, Bandwidth= 1190Hz/pixel Echo Spacing=1.03 ms). For Dataset 4 and 5 (DS4, DS5), whole brain images were collected, using an anterior to posterior phase encoding direction, with parameters matched to the HCP 7T Protocol (85 slices, TR=1s, MB=5, GRAPPA=2, 7/8 Partial Fourier, 1.6mm isotropic, TE=22.2ms, Flip Angle=78°, Bandwidth=1925Hz/pixel, Echo Spacing=0.64 ms). Dataset 8 (DS8), images captures only the occipital pole using a right to left phase encoding direction (32 slices, TR=0.350s, MB=4, GRAPPA=2, 6/8 Partial Fourier, 1.3mm isotropic, TR=23ms, Flip Angle=32, Bandwidth=1865, Echo Spacing=0.586ms).

3T fMRI Data (DS6, 7). Dataset 6 (DS6) was acquired with a higher resolution than typically used at 3T studies, capturing most of the brain, excluding the cerebellum, with anterior to posterior phase encoding. (100 slices, TR= 2.1s, MB= 4, GRAPPA =2, 7/8 Partial Fourier, 1.2mm isotropic, TE= 32.6ms, Flip Angle= 78°, Bandwidth= 1595Hz/pixel, Echo Spacing=0.76 ms). Dataset 7 was a whole brain study, using an HCP-like 3T protocol (72 slices, TR= 0.8s, Multiband= 8, 2mm isotropic, TE =37ms, Flip Angle= 52°, Bandwidth =2290Hz/pixel, Echo Spacing=0.58 ms).

Anatomical Imaging: T1-weighted anatomical images were obtained for DS1-DS7 using an MPRAGE (58) sequence (192 slices; TR=1900 ms; FOV: 256 x 256 mm; flip angle= 9°; TE= 2.52 ms; spatial resolution 0.8 mm isotropic voxels) which were collected with a 3T Siemens Magnetom Prisma^{fit} system. The anatomical images for DS8 were acquired using a MP2RAGE (59) sequence (192 slices; TR=4300 ms; FOV: 240 x 225 mm; flip angle= 4°; TE= 2.27 ms; spatial resolution 0.8 mm isotropic voxels) collected with a 7T Siemens Magnetom

scanner. Anatomical images were used only to define regions of interest (ROIs, see Region of Interest Creation, below).

Initial MR Image Preprocessing. Two separate reconstruction methods were used in all subsequent analyses described below. Following data acquisition, the k-space data files for each receive channel produced by the SIEMENS system were saved. These were reconstructed offline (as opposed to using the scanners inbuilt reconstruction algorithms) using standard techniques implemented in-house to produce magnitude images with minimal processing, similar to the typical DICOM images produced by the default Siemens reconstruction. These minimally processed data are referred to as the “Standard” reconstruction, to emphasize the fact that this is a standard reconstruction of the magnitude images. The second image reconstruction, which is the primary consideration of this manuscript, is derived from the same raw k-space files, however, following the denoising steps that aim to remove thermal noise. The denoising method, known as NORDIC for Noise Reduction with Distribution Corrected (NORDIC) PCA, is described in prior work (36, 37). In brief, this method uses a patch based, PCA implementation to identify and discard components of the data that are indistinguishable from zero-mean, normally distributed (i.e., thermal) noise, using the magnitude and complex portions of the MRI signal as input. We refer to these data as “NORDIC” throughout the remainder of the manuscript. The use of an offline reconstruction for these data, rather than the typical scanner reconstruction assured that, other than the denoising step, all other reconstruction steps were identical.

In addition to the Standard and NORDIC reconstruction, we considered a third denoising approach. We applied dwidenoise (38, 39) version 3.0.0, with MRtrix3, to the Standard magnitude images. In brief, dwidenoise also aims to suppress normally distributed noise using a patched-based denoising approach. Noise components are estimated on the basis of Marčenko-Pastur principal component analysis (MPPCA) with the size of the patch depending primarily on timeseries length. Default settings were used. For DS1, DS2 and DS6, this led to a 5x5x5 patch size, whereas Datasets

DS3, DS4, DS5 and DS7 had a 7x7x7 patch size. The “Standard” data was considered the reference point for further analyses.

Prior to any additional processing, we examined the noise removed by NORDIC and dwidenoise. Specifically, the noise removed by NORDIC was calculated by taking the mean of the magnitude of the complex difference between the Standard and NORDIC data. The noise removed by dwidenoise residuals was calculated as the absolute value of the mean difference between the Standard and dwidenoise data. Maps of the g-factor were derived from the k-space data files.

Processing

All subsequent fMRI processing was performed using AFNI (60) and was identical for preparations of the data. For all Datasets and denoising methods, we used conventional processing steps, with settings chosen to minimize any loss of precision. First, slice timing was corrected by using the most accurate interpolation option (Fourier) using the first slice as the reference timepoint. Motion correction was then performed using the first volume of the first run from the Standard reconstruction as the registration target, with the ‘Fourier’ estimation and interpolation option chosen. Using the same target for motion correction across all reconstruction methods allows for subsequent voxel to voxel comparisons between the different methods. In order to compare the signal characteristics of the reconstructions to approaches which aim to reduce the impact of thermal noise, we created three additional comparator datasets from the Standard data. This included data smoothed with a FWHM gaussian kernel equivalent in size to one voxel (hereafter labeled “+1 voxel FWHM”), data smoothed with a FWHM gaussian kernel equivalent in size to 1.5 voxels (“+1.5 voxel FWHM”), and data temporally smoothed (“+temporal smoothing”) using a sliding window average approach with window sized between 9 and 10.5 s.

Data were then scaled to have a voxel-wise mean of 100. In order to evaluate the fMRI statistical performance of each data set we considered two general linear modeling frameworks.

Task Event Modeling

GLM One: Conventional Approach. The scaled data were passed through a generalized least squares (GLSQ) regression model using a conventional hemodynamic response estimate ("SPMG1"), with a beta estimate produced on a per run basis for events in each of the 9 (DS6), 8 (DS1, DS2, DS5, DS7), or 6 (DS3, DS4) runs. This was performed using AFNI's *3dREMLfit*, which calculates an autoregressive moving average (ARMA(1,1)) model to estimate the temporal autocorrelation on a voxel by voxel basis, thereby improving the accuracy of t-statistic estimates (61).

To aid in the creation of regions of interest (ROIs, see below) we used the Monte Carlo method as implemented in *3dClustSim*, using smoothness estimated from the residuals of the standard data to determine the cluster size that achieved a statistical significance of $p<0.05$ (family-wise error rate corrected). Specifically, this is the smoothness estimate derived from *3dFWHMx* with the '-acf' option. *3dClustSim* simulates 10,000 noise-only datasets with smoothness that is matched to the data to create a null distribution of cluster sizes. This was done only for the Standard data and these cluster thresholds were only used to create ROIs and not considered further.

GLM Two: Finite Impulse Response (FIR) Model. In order to fully investigate the temporal information present in each functional acquisition, we also used a finite impulse response (FIR) model using AFNI's *3dDeconvolve* function with TENT estimators. For all datasets this created a TR-by-TR estimate of the HRF for target and non-target conditions. The window for which the TENT estimates were created varied between datasets, ranging from 15 to 29.4 seconds, but was identical between processing methods. These estimates of the task-responsive hemodynamic response function (HRF) were produced on a run-by-run, voxel-by-voxel basis which represents a highly SNR-starved modeling regime, as these types of unconstrained HRF models typically require the estimates of HRFs over multiple runs. The events for DS3 were separated by 1-second steps. Thus, the data for the FIR model was simultaneously slice time corrected and up-sampled to a 1-second sampling rate using in-house code prior to processing in AFNI. This step was performed in an identical manner for

the Standard, NORDIC, and dwidenoise data, and was performed only for the FIR model. Though we show time courses for the HRF estimates for only the “Target” condition, the full model was used for cross validated prediction accuracy introduced below.

Region of Interest Creation

Anatomical ROIs: Anatomical images for each dataset were segmented into different tissues and skull-stripped using the Segment tool from SPM12 (62). These skull-stripped anatomical images were aligned (rigid-body) to the mean of the first run of the standard data, after it had undergone motion correction, using a local Pearson correlation estimator (63) (*3dAllineate)*. The first three tissue class images, corresponding to gray matter, white matter and a mixture of CSF and vasculature (hereafter just CSF) were also transformed into this epi-aligned space. The aligned tissue probability maps were then converted to binary masks, with a threshold of 0.95 probability for each class and then resampled to match each dataset’s EPI grid. These ROIs were then masked by the ‘epi’ mask generated during processing to minimize the amount of each anatomical ROI that includes voxels outside of the acquired brain data or overlapped with areas of near-complete signal dropout. These steps produced grey matter, white matter and CSF regions of interest (ROIs) which are aligned to each unique functional dataset. Note that no distortion correction was applied to the functional data in order to minimize any additional blurring.

Functional ROIs: Multiple ROIs were created from the Standard full GLSQ model’s corresponding contrasts of interest from each dataset. For all datasets (except DS4, DS8) we created a “Target ROI”, which included all clusters with more than 10 contiguous activated (defined as contact via faces, edges, or corners) voxels using a voxel threshold of $p<0.001$ for contrast of the target stimulus (center or faces) vs the alternative stimulus (surround or scrambled faces). For all datasets, 10 contiguous voxels at this threshold produced clusters that exceed the typical $p_{FWE}<0.05$ threshold. For DS4, which used face stimuli with variable phase coherence, the ROIs were generated from a separate localizer analysis of Standard data, using Faces greater than Scrambled Faces contrast. No

functional ROIs were created for DS8, as it was only used for the frequency spectrum analysis (see Temporal Precision Evaluation).

In a similar way, a “Non-Target” ROI was also created from all significant clusters greater than 10 contiguous voxels at a voxel-wise threshold of $p<0.001$ and positive signal associated with the alternative condition, that is, surround or scrambled faces, only. The use of these two different contrasts produced a complementary selection of voxels.

Collectively we produced 2 functional ROIs and 3 anatomically derived ROIs, per dataset. These were then used to summarize the distribution of the values from other voxel wise measures (e.g., t-statistics).

Spatial Precision Evaluation.

Global Smoothness: Spatial precision was estimated using global smoothness estimates produced from each dataset within a mask. This mask, automatically generated during processing by AFNI, is a contiguous volume produced by an interactively clipping process which excludes the background and very low intensity values (*3dAutomask*). This mask will be subsequently described as the “EPI mask”. These global smoothness estimates, produced by *3dFWHMx* using the ‘-acf’ option were conducted for three timepoints in processing and analysis. Specifically, we calculated global smoothness on each dataset prior to any processing, after processing and on the residuals from the conventional GLM. These FWHM estimates, in units of mm, are derived from the mixed gaussian + monoexponential decay model fit produced by the ‘-acf’ option and are thought to be more accurate estimates of the spatial smoothness of fMRI data, as this accounts for the possibility of long tail spatial autocorrelation (47). Significance was evaluated using paired t-tests between the 3 primary types of processing (Standard, NORDIC, dwidenoise).

Local Smoothness: In addition, we performed an analysis using *3dLocalACF*, which estimates a voxel-wise spatial autocorrelation function, using a spherical local neighborhood with a radius of 10 voxels. As this method is highly sensitive to trends within the data, this local smoothness

was estimated only on the residuals of the conventional GLM. This was done independently per run, and then averaged. Values were then summarized on the basis of gray matter, white matter and a "CSF" mask, which due to segmentation being performed only on a T1 weighted anatomical image also contains areas of probable veins (e.g., superior sagittal sinus).

Temporal Precision Evaluation.

Fourier Spectrum Analysis: Using DS7 and DS8, which were sampled at 800 and 350ms respectively, we performed a fast Fourier transform on each independent run of the data, doing this for the Standard and NORDIC data. The FFT was performed on the scaled data from preprocessing. We then took the mean across voxels within each tissue class (GM, WM, CSF), and then the mean and standard deviation across each independent run. We then examined the normalized frequency spectrum within each ROI up the Nyquist frequencies (DS7: 0.625Hz, DS8: 1.429Hz) to determine how NORDIC processing altered the frequency spectrum and if physiological frequency peaks remained in the data. For DS7, the CSF, gray matter and white matter ROIs contained 13314, 66664 and 42356 voxels respectively. For DS8, the CSF, gray matter and white matter ROIs contained 2367, 51797 and 19281 voxels respectively.

Cross validation: We used an exhaustive K-fold permutation approach to evaluate the accuracy of the estimated HRF time courses derived from the full FIR model ("GLM Two", see above) and to determine if the NORDIC denoising method altered their temporal structure. Specifically, we varied the number of runs, K, used to predict the remainder between 1 and the number of runs - 1. Prior to being entered into the model, we projected out polynomials (up to order = run duration in minutes - 1) to remove low frequency drift and masked the data using the EPI data mask derived from the Standard data.

In order to limit or analysis to voxels that were plausibly task responsive, we determined the overall leave-one-out cross validated coefficient of determination R^2 using all runs of each Standard

dataset in a model with a conventional HRF. This generated one map of R^2 across the whole brain for each Dataset. This map was used only to summarize the subsequent exhaustive FIR based cross-validation scheme.

In each fold, we used K runs to estimate the FIR model, constructing a series of betas for each stimulus, corresponding to the estimated BOLD response over time to each stimulus on a voxel-by voxel basis. These estimates were then multiplied by a design matrix for the held-out runs (number of runs - K) to generate predicted timeseries for each voxel. We then determined how well the predicted timeseries matched the true timeseries using the coefficient of determination, R^2. We first considered the K=1 case, in which one run was used to predict the timeseries of all remaining runs. The R^2 (and standard error over permutations) of the FIR cross-validation scheme was calculated for voxels ranging from 5% of variance explained in the conventional model to the max R^2 for that dataset.

For K>=2, we summarized error across permutations using a R^2>=15% mask from the conventional model. This process was repeated to generate the following three comparisons: Standard predicting Standard, NORDIC predicting NORDIC, and, most importantly, NORDIC predicting Standard. This final comparison is significant, as it represents the NORDIC data predicting the non-denoised data, which has undergone minimal processing, and retains all of the signals of interest, albeit mixed with noise.

Acknowledgements and Funding Sources

The authors would like to thank Dr. Kendrick Kay for his input. This work was supported by NIH grants: U01EB025144 (K.U.), P41 EB027061 (K.U.), P30 NS076408 (K.U.) and RF1 MH116978 (E.Y.), and RF1 MH117015 (G.G.)

References

1. K. Ugurbil, Magnetic resonance imaging at ultrahigh fields. *IEEE Trans Biomed Eng* **61**, 1364-1379 (2014).
2. K. Ugurbil, Imaging at ultrahigh magnetic fields: History, challenges, and solutions. *Neuroimage* **168**, 7-32 (2018).
3. P. J. Koopmans, M. Barth, D. G. Norris, Layer-specific BOLD activation in human V1. *Human Brain Mapping* **31**, 1297-1304 (2010).
4. C. A. Olman *et al.*, Layer-Specific fMRI Reflects Different Neuronal Computations at Different Depths in Human V1. *PLOS ONE* **7** (2012).
5. F. De Martino *et al.*, Cortical Depth Dependent Functional Responses in Humans at 7T: Improved Specificity with 3D GRASE. *PLOS ONE* **8**, e60514 (2013).
6. T. Q. Duong, D. S. Kim, K. Ugurbil, S. G. Kim, Localized cerebral blood flow response at submillimeter columnar resolution. *Proc Natl Acad Sci U S A* **98**, 10904-10909 (2001).
7. K. Ugurbil, What is feasible with imaging human brain function and connectivity using functional magnetic resonance imaging. *Philos Trans R Soc Lond B Biol Sci* **371** (2016).
8. S. O. Dumoulin, A. Fracasso, W. van der Zwaag, J. C. W. Siero, N. Petridou, Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function. *Neuroimage* **168**, 345-357 (2018).
9. F. De Martino *et al.*, The impact of ultra-high field MRI on cognitive and computational neuroimaging. *Neuroimage* **168**, 366-382 (2018).
10. E. A. Stringer, L. M. Chen, R. M. Friedman, C. Gatenby, J. C. Gore, Differentiation of somatosensory cortices by high-resolution fMRI at 7 T. *Neuroimage* **54**, 1012-1020 (2011).
11. L. Huber *et al.*, Sub-millimeter fMRI reveals multiple topographical digit representations that form action maps in human motor cortex. *NeuroImage* **208**, 116463 (2020).
12. E. Yacoub, N. Harel, K. Ugurbil, High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci U S A* **105**, 10607-10612 (2008).
13. E. Yacoub, N. Harel, K. Uğurbil, High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences* **105**, 10607-10612 (2008).
14. C. Triantafyllou, J. R. Polimeni, L. L. Wald, Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. *NeuroImage* **55**, 597-606 (2011).
15. C. Triantafyllou *et al.*, Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage* **26**, 243-250 (2005).
16. S. Moeller *et al.*, Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* **63**, 1144-1153 (2010).
17. K. Ugurbil *et al.*, Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80-104 (2013).
18. S. M. Smith *et al.*, Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144-168 (2013).
19. K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger, SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* **42**, 952-962 (1999).
20. N. Todd *et al.*, Functional Sensitivity of 2D Simultaneous Multi-Slice Echo-Planar Imaging: Effects of Acceleration on g-factor and Physiological Noise. *Frontiers in Neuroscience* **11** (2017).
21. J. Gonzalez-Castillo *et al.*, Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5487-5492 (2012).

22. D. A. Handwerker, J. M. Ollinger, M. D'Esposito, Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage* **21**, 1639-1651 (2004).
23. A. J. Taylor, J. H. Kim, D. Ress, Characterization of the hemodynamic response function across the majority of human cerebral cortex. *NeuroImage* **173**, 322-331 (2018).
24. S. G. Warren, E. Yacoub, G. M. Ghose, Featural and temporal attention selectively enhance task-appropriate representations in human primary visual cortex. *Nature Communications* **5**, 1-12 (2014).
25. X. Hu, S. G. Kim, Reduction of signal fluctuation in functional MRI using navigator echoes. *Magn Reson Med* **31**, 495-503 (1994).
26. K. N. Kay, A. Rokem, J. Winawer, R. F. Dougherty, B. A. Wandell, GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front Neurosci* **7**, 247 (2013).
27. G. H. Glover, T. Q. Li, D. Ress, Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* **44**, 162-167 (2000).
28. K. Shmueli *et al.*, Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. *Neuroimage* **38**, 306-320 (2007).
29. M. Bianciardi, P. van Gelderen, J. H. Duyn, M. Fukunaga, J. A. de Zwart, Making the most of fMRI at 7 T by suppressing spontaneous signal fluctuations. *Neuroimage* **44**, 448-454 (2009).
30. T. E. Lund, K. H. Madsen, K. Sidaros, W. L. Luo, T. E. Nichols, Non-white noise in fMRI: does modelling have an impact? *Neuroimage* **29**, 54-66 (2006).
31. C. Triantafyllou, R. D. Hoge, L. L. Wald, Effect of spatial smoothing on physiological noise in high-resolution fMRI. *Neuroimage* **32**, 551-557 (2006).
32. J. V. Haxby, A. C. Connolly, J. S. Guntupalli, Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience* **37**, 435-456 (2014).
33. K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, Identifying natural images from human brain activity. *Nature* **452**, 352-355 (2008).
34. T. Naselaris, K. N. Kay, S. Nishimoto, J. L. Gallant, Encoding and decoding in fMRI. *Neuroimage* **56**, 400-410 (2011).
35. V. Q. Vu *et al.*, Encoding and Decoding V1 Fmri Responses to Natural Images with Sparse Nonparametric Models. *Ann Appl Stat* **5**, 1159-1182 (2011).
36. L. Vizioli *et al.*, A Paradigm Change in Functional Brain Mapping: Suppressing the Thermal Noise in fMRI. *BioRxiv* 10.1101/2020.11.04.368357 (2020).
37. S. Moeller *et al.*, NOise reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing. *NeuroImage* **226**, 117539 (2021).
38. J. Veraart *et al.*, Denoising of diffusion MRI using random matrix theory. *NeuroImage* **142**, 394-406 (2016).
39. L. Cordero-Grande, D. Christiaens, J. Hutter, A. N. Price, J. V. Hajnal, Complex diffusion-weighted image estimation via matrix recovery under general noise models. *NeuroImage* **200**, 391-404 (2019).
40. J. D. Tournier *et al.*, MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137 (2019).
41. J. Gonzalez-Castillo *et al.*, Evaluation of multi-echo ICA denoising for task based fMRI studies: Block designs, rapid event-related designs, and cardiac-gated fMRI. *NeuroImage* **141**, 452-468 (2016).
42. P. Kundu *et al.*, Multi-Echo fMRI: A Review of Applications in fMRI Denoising and Analysis of BOLD Signals. *NeuroImage* 10.1016/j.neuroimage.2017.03.033 (2017).
43. R. H. Pruim, M. Mennes, J. K. Buitelaar, C. F. Beckmann, Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *NeuroImage* **112**, 278-287 (2015).
44. S. M. Smith *et al.*, Resting-state fMRI in the Human Connectome Project. *NeuroImage* **80**, 144-168 (2013).
45. G. Chen *et al.*, 10.1101/2020.05.19.102111 (2020).

46. J. R. Polimeni, V. Renvall, N. Zaretskaya, B. Fischl, Analysis strategies for high-resolution UHF-fMRI data. *NeuroImage* **168**, 296-320 (2018).
47. R. W. Cox, G. Chen, D. R. Glen, R. C. Reynolds, P. A. Taylor, FMRI Clustering in AFNI: False-Positive Rates Redux. *Brain Connect* **7**, 152-171 (2017).
48. A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, K. Ugurbil, Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage* **35**, 539-552 (2007).
49. K. Kay, A. Rokem, J. Winawer, R. Dougherty, B. Wandell, GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in Neuroscience* **7** (2013).
50. E. Yacoub *et al.*, Spin-echo fMRI in humans using high spatial resolutions and high magnetic fields. *Magnetic Resonance in Medicine* **49**, 655-664 (2003).
51. D. A. Roberts, J. A. Detre, L. Bolinger, E. K. Insko, J. S. Leigh, Jr., Quantitative magnetic resonance imaging of human brain perfusion at 1.5 T using steady-state inversion of arterial water. *Proc Natl Acad Sci U S A* **91**, 33-37 (1994).
52. L. Huber *et al.*, Slab-selective, BOLD-corrected VASO at 7 Tesla provides measures of cerebral blood volume reactivity with high signal-to-noise ratio. *Magn Reson Med* **72**, 137-148 (2014).
53. L. R. Huber *et al.*, LayNii: A software suite for layer-fMRI. *NeuroImage* 10.1016/j.neuroimage.2021.118091, 118091 (2021).
54. A. I. Blazejewska, B. Fischl, L. L. Wald, J. R. Polimeni, Intracortical smoothing of small-voxel fMRI data can provide increased detection power without spatial resolution losses compared to conventional large-voxel fMRI data. *NeuroImage* **189**, 601-614 (2019).
55. B. M. Adhikair *et al.*, Heritability estimates on resting state fMRI data using ENIGMA analysis pipeline. *Pac Symp Biocomput* **23**, 307-318 (2018).
56. B. M. Adhikari *et al.*, A resting state fMRI analysis pipeline for pooling inference across diverse cohorts: an ENIGMA rs-fMRI protocol. *Brain imaging and behavior* **13**, 1453-1467 (2019).
57. B. Ades-Aron *et al.*, Improved Task-based Functional MRI Language Mapping in Patients with Brain Tumors through Marchenko-Pastur Principal Component Analysis Denoising. *Radiology* **298**, 365-373 (2021).
58. Jr, M. J. P., Brookeman, Rapid three-dimensional T1-weighted MR imaging with the MP-RAGE sequence. *Journal of magnetic resonance imaging : JMRI* **1** (1991).
59. J. P. Marques *et al.*, MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage* **49**, 1271-1281 (2010).
60. R. W. Cox, AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research* **29**, 162-173 (1996).
61. W. Olszowy, J. Aston, C. Rua, G. B. Williams, Accurate autocorrelation modeling substantially improves fMRI reliability. *Nature Communications* **10**, 1-11 (2019).
62. J. Ashburner, K. J. Friston, Unified segmentation. *NeuroImage* **26**, 839-851 (2005).
63. Z. S. Saad *et al.*, A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *NeuroImage* **44**, 839-848 (2009).

Supplemental Figures

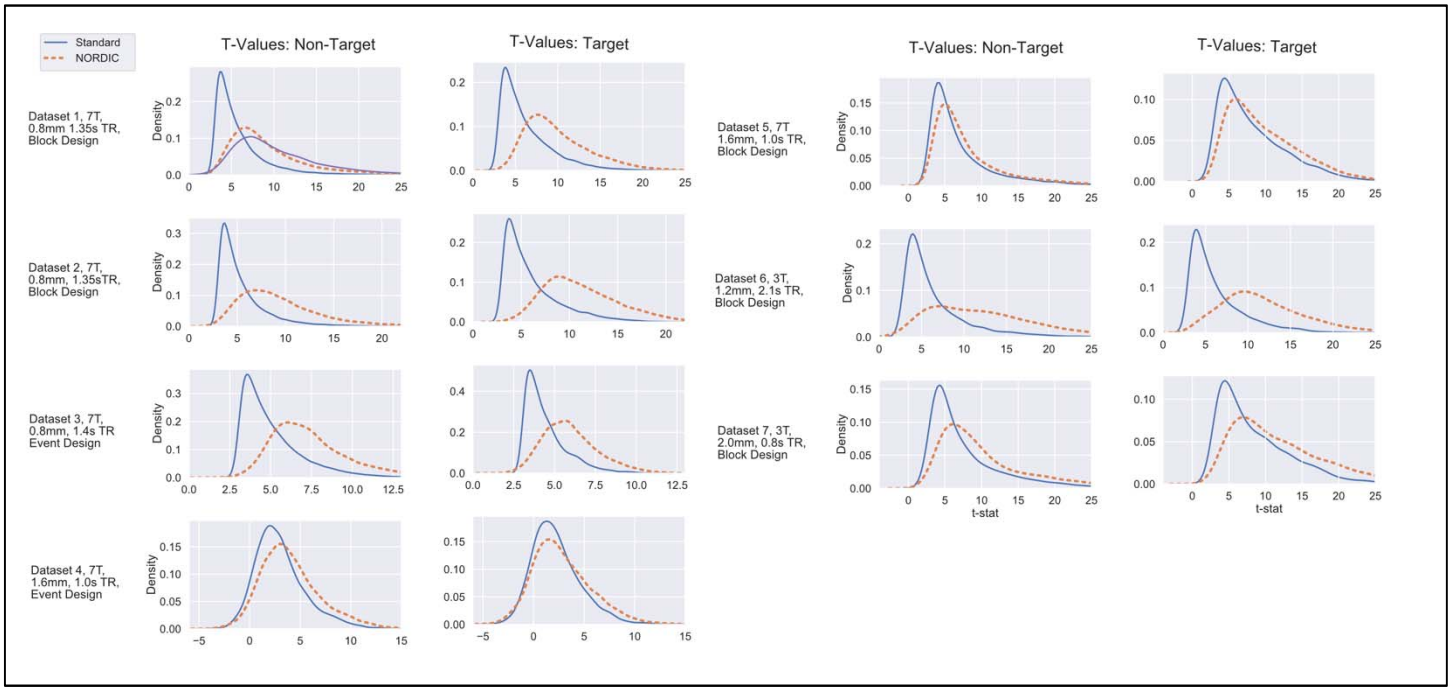


Figure S1. T-statistic histograms comparing Standard and NORDIC data within the Non-Target and Target ROIs.

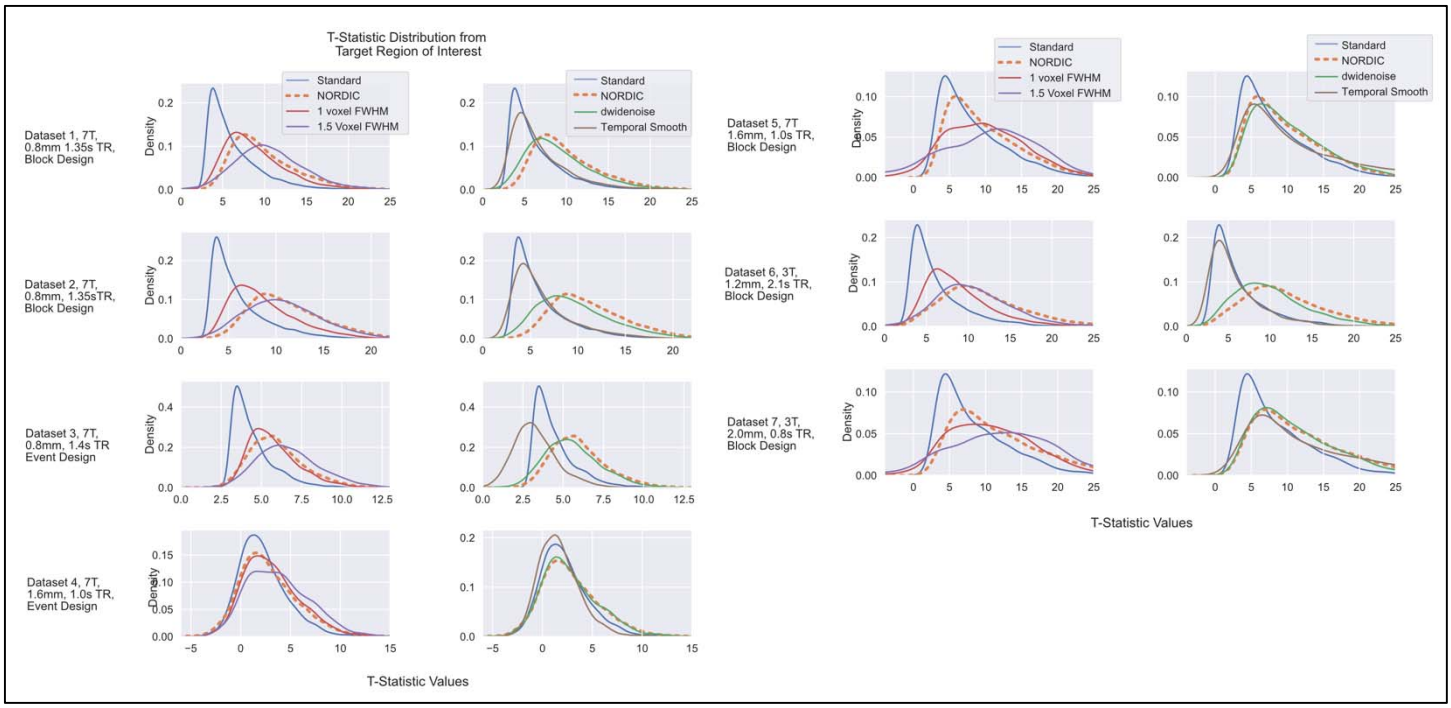


Figure S2. T-statistic values from Datasets 1 through 7 in the Target ROI, under different processing schemes.

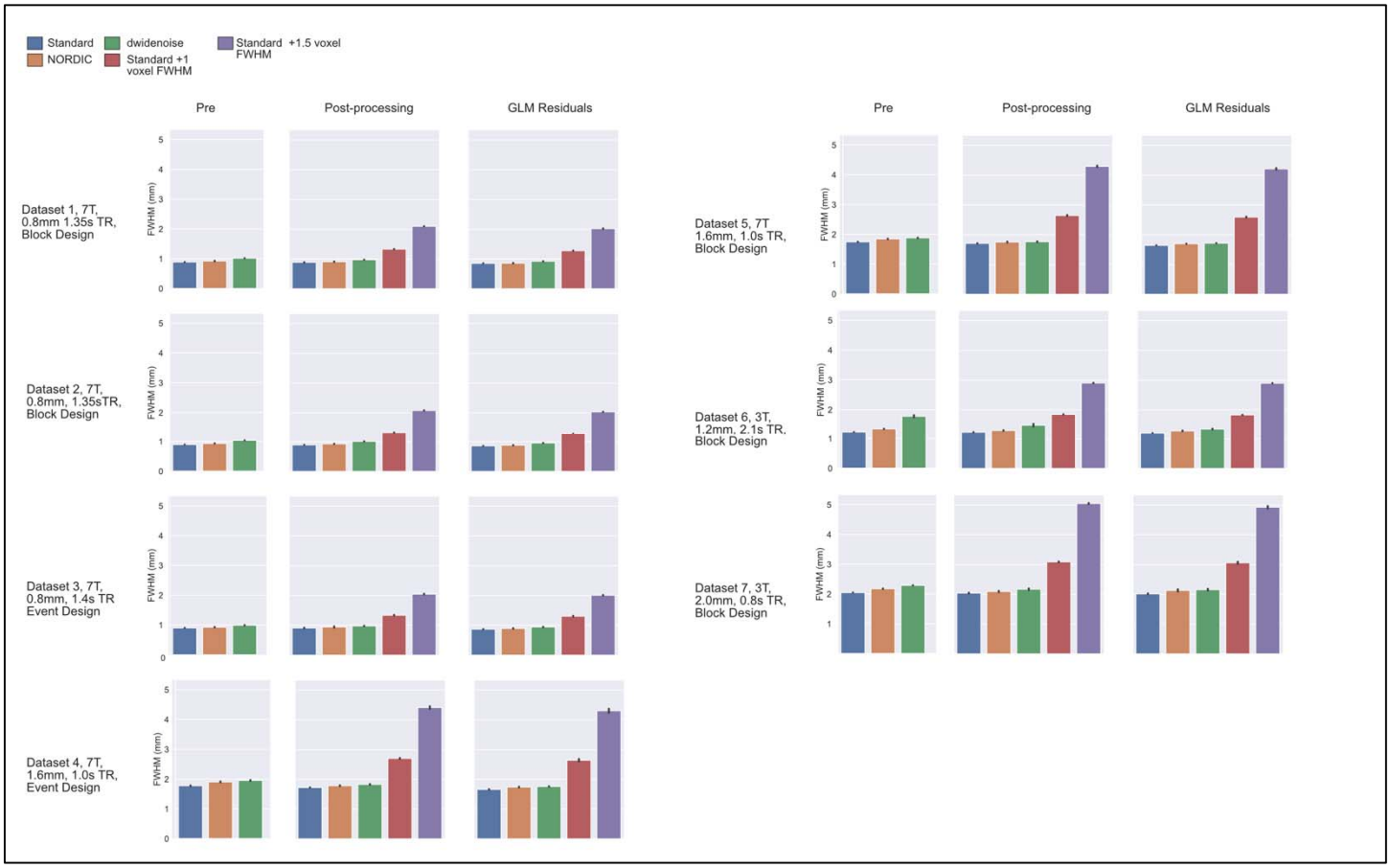
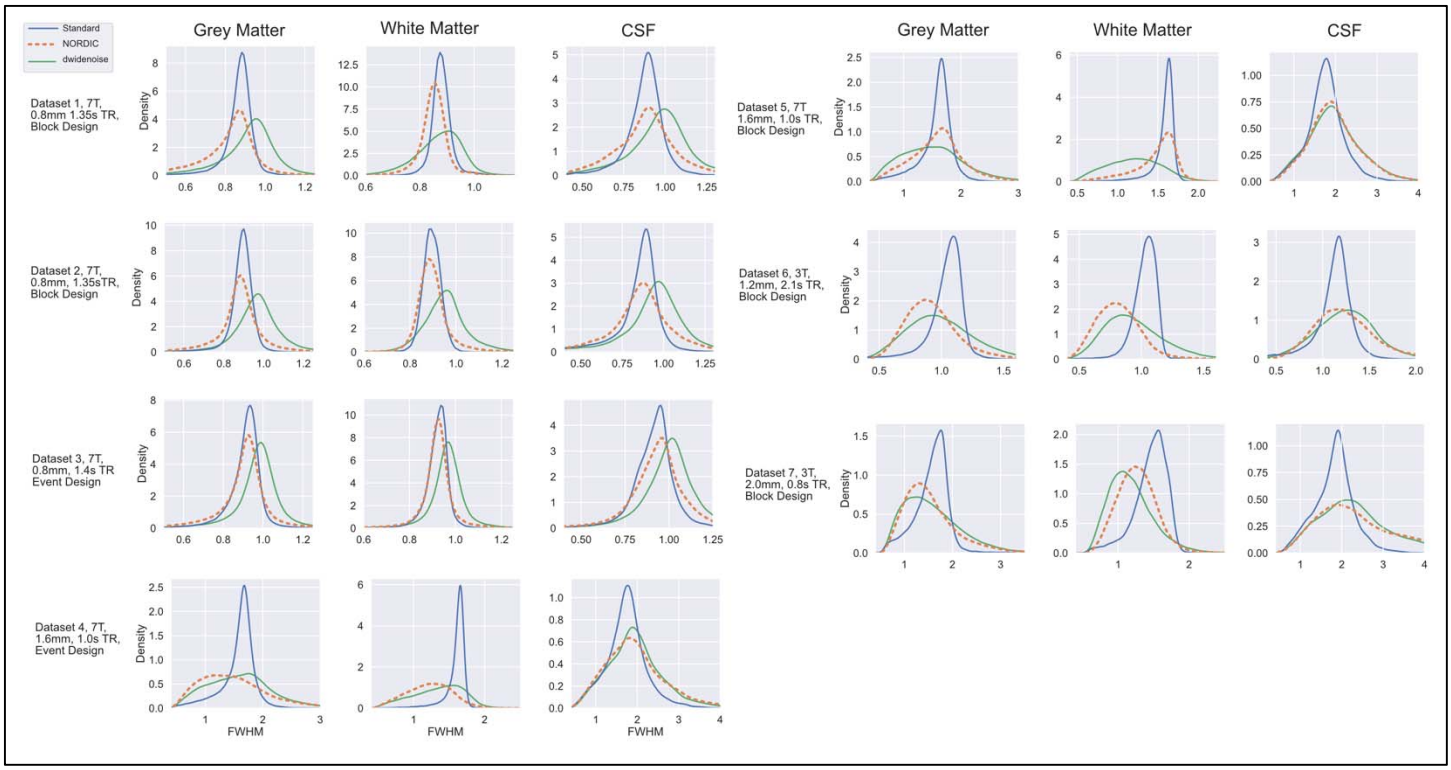
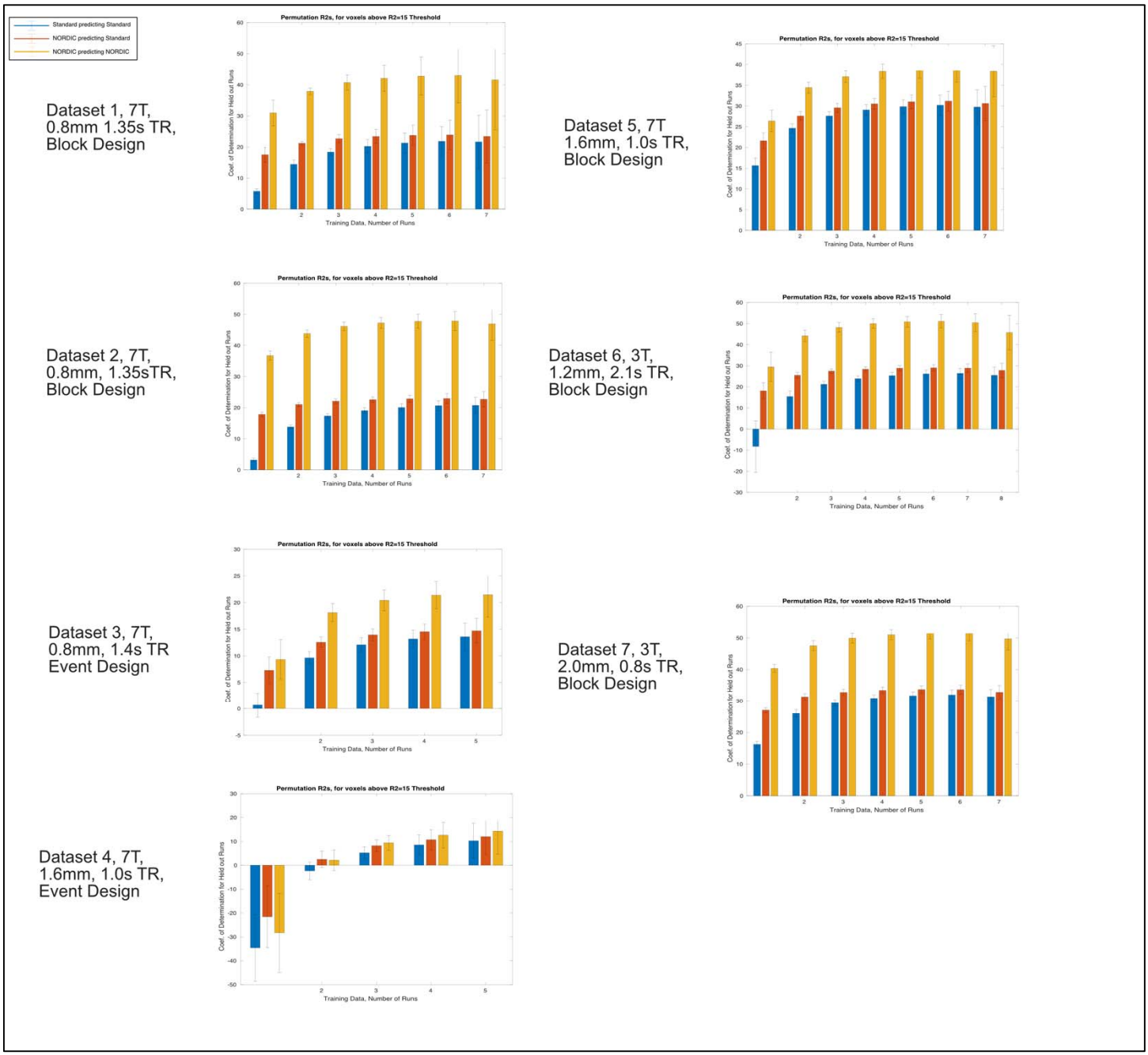


Figure S3. Global Smoothness Estimates for Datasets 1 – 7.



Supplemental Figure 4. Local Smoothness from Datasets 1 – 7.



Supplemental Figure 5. Cross Validated R^2 for Datasets 1 through 7.

		Gray Matter	White Matter	CSF
DS1	NORDIC	0.82±0.148	0.85±0.056	0.89±0.214
	Offline	0.87±0.084	0.88±0.039	0.88±0.119
	dwidenoise	0.91±0.153	0.87±0.094	0.98±0.207
DS2	NORDIC	0.88±0.133	0.90±0.082	0.88±0.204
	Offline	0.89±0.074	0.90±0.041	0.84±0.151
	dwidenoise	0.97±0.133	0.96±0.102	0.96±0.195
DS3	NORDIC	0.91±0.147	0.92±0.089	0.95±0.192
	Offline	0.91±0.10	0.92±0.062	0.92±0.138
	dwidenoise	0.98±0.15	0.97±0.102	1.02±0.19
DS4	NORDIC	1.53±0.657	1.22±0.32	1.96±0.749
	Offline	1.57±0.301	1.59±0.159	1.74±0.487
	dwidenoise	1.62±0.592	1.34±0.357	1.95±0.667
DS5	NORDIC	1.6±0.476	1.47±0.268	1.98±0.622
	Offline	1.58±0.287	1.58±0.157	1.76±0.449
	dwidenoise	1.53±0.569	1.21±0.336	1.97±0.649
DS6	NORDIC	0.93±0.238	0.81±0.191	1.24±0.36
	Offline	1.04±0.154	1.03±0.102	1.13±0.202
	dwidenoise	1.00±0.289	0.94±0.247	1.25±0.33
DS7	NORDIC	1.54±0.57	1.26±0.273	2.5±1.15
	Offline	1.57±0.344	1.46±0.236	1.82±0.473
	dwidenoise	1.58±0.6	1.18±0.311	2.36±0.929

Supplemental Table 1. Mean and Standard Deviation of Local Smoothness Estimates in mm FWHM for Datasets 1 through 7.