# More with less: Simulation-based inference enables accurate diffusion-weighted MRI with minimal acquisition time

Maximilian F. Eggl[1,2,*] and Silvia De Santis[1,*]

[1]Institute of Neuroscience, CSIC-UMH, Alicante
[2]Institute of Experimental Epileptology and Cognition Research, University of Bonn Medical Center, Bonn
[*]Co-corresponding authors, meggl@umh.es, dsilvia@umh.es

## Abstract

Diffusion-weighted magnetic resonance imaging (dw-MRI) is an essential tool in neuroimaging, providing non-invasive insights into brain microstructure. However, obtaining reproducible and accurate maps requires lengthy acquisition due to the need to massively oversample the parameter space. This means that tensor-based dw-MRI accessibility is still relatively low in daily practice, and more advanced approaches with increased sensitivity and specificity to microstructure are seldom applied in research and clinical contexts. Motivated by recent advances in simulation-based inference (SBI) methods, this work uses neural networks to model the posterior distribution of key diffusion parameters when provided experimental data, allowing accurate estimation with fewer measurements and without the need to train on real data. We find that SBI outperforms standard non-linear least squares fitting under noisy and sparse data conditions in both diffusion tensor and kurtosis imaging, reducing imaging time by 90% while maintaining high accuracy and robustness. Demonstrated on simulated and real data in healthy and pathological brains, this approach can substantially impact radiology by: *i)* increasing dw-MRI access to more patients, including those unable to undergo long exams; *ii)* promoting advanced dw-MRI protocols for greater microstructure sensitivity; and *iii)* rescuing older data where noise hindered analysis. Combining SBI with dw-MRI could greatly improve clinical MRI workflows by reducing patient discomfort, enhancing scan efficiency, and enabling advanced imaging approaches in a data and privacy friendly way.

## Introduction

Diffusion-weighted magnetic resonance imaging (dw-MRI) is a pivotal neuroimaging tool for non-invasive, *in vivo* access to brain microstructure. By sensitizing MRI signals to water's random motion, dw-MRI provides unique insights into tissue microstructural organization, particularly in the brain[1]. Extending this framework, techniques like diffusion tensor imaging (DTI)[2] and diffusion kurtosis imaging (DKI)[3] have significantly impacted clinical applications[4]. However, noisy measurements substantially affect these methods, compromising the accuracy and reliability of fitted diffusion parameters[5,6]. While denoising techniques exist (e.g.,[7,8]), they often require oversampling beyond theoretical needs for DTI and DKI[9], increasing scan times and patient discomfort. On the one hand, this oversampling can exclude individuals unable to undergo long examinations and, on the other, limits the combination of dw-MRI with other sequences, hindering deeper insights through multi-modal imaging. Consequently, DTI and DKI remain underused in clinical practice, and advanced methods rarely reach clinical researchers.

Current parameter fitting methods like non-linear least squares (NLLS) and weighted least squares (WLS)[2,10,11] are additionally sensitive to initial parameter guesses and may not fully capture uncertainties in noisy, low signal-to-noise ratio conditions[5]. To address these limitations, we employ simulation-based inference (SBI) methods, powerful alternatives for statistical inference in complex models[12–14]. SBI uses simulations to approximate the posterior distribution of model parameters[15]. Neural posterior estimation (NPE)[16,17], a more sophisticated SBI method, uses deep neural networks to learn the posterior distribution directly from simulated data, enhancing robustness and accuracy. NPE has already been successfully applied in neuroscience[18,19] and imaging neuroscience[20–23]. Nonetheless, in the context of dw-MRI this framework has not been used previously to achieve accurate biomarker maps with the minimal amount of theoretical required information. Importantly, SBI relies on simulated data, contrasting with deep learning approaches that require real data for training[24,25], making it data- and privacy-friendly.

Given that DTI and DKI have straightforward simulation paradigms, SBI, and particularly NPE, is a natural fit to enhance fitting accuracy. This work demonstrates how applying SBI to DTI and DKI allows direct fitting of relevant tensors

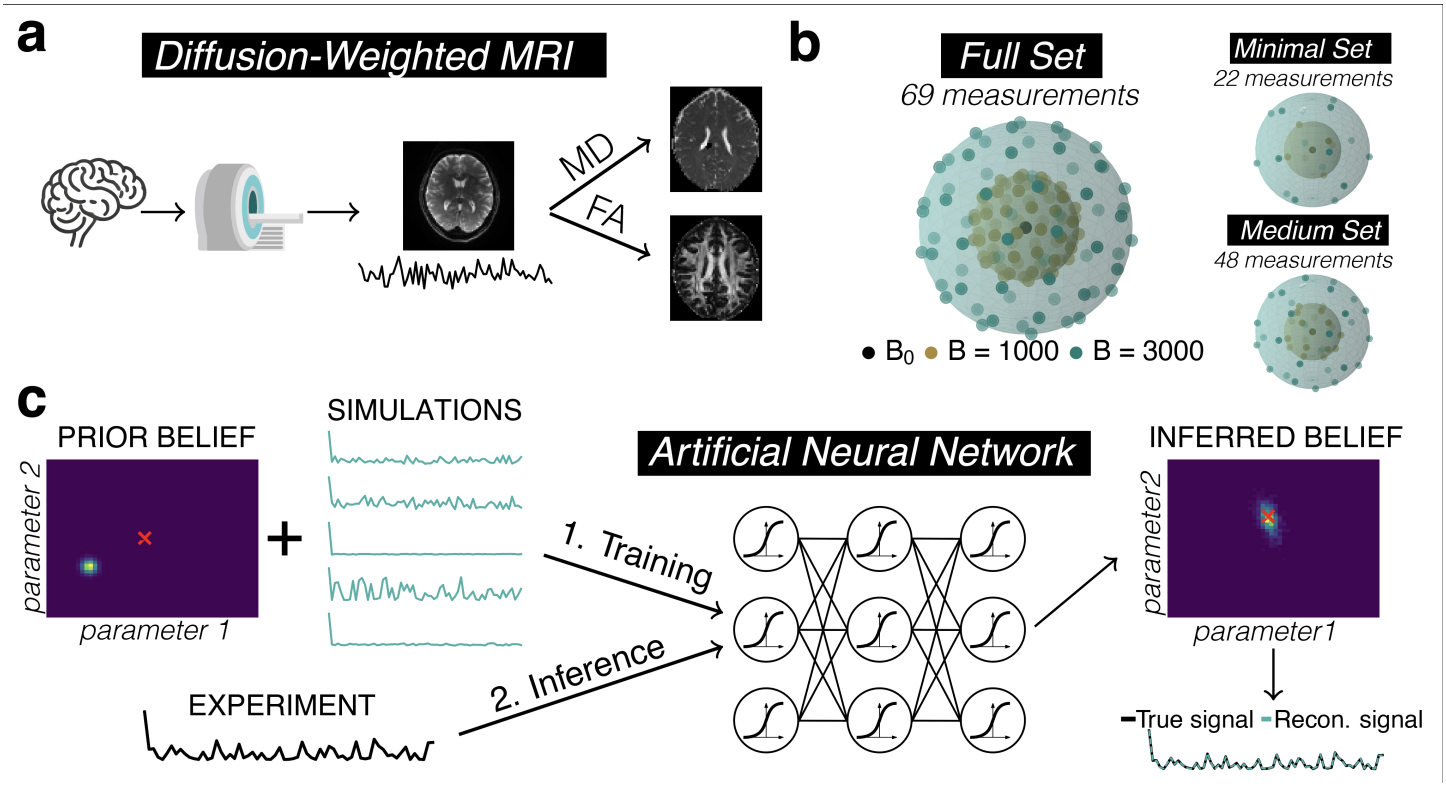


Figure 1: **Overview of the dw-MRI data processing and analysis pipeline. a)** Diffusion-weighted MRI data is processed by fitting a mathematical model to the signal, allowing for the extraction of key parameters such as mean diffusivity (MD) and fractional anisotropy (FA). These parameters provide insights into brain health and function. **b)** Both DTI and DKI frameworks rely on oversampling the diffusion tensor space using one or more b-values and many b-vectors. Left: full set of orientations for the $B = 1000$ (gold) and $B = 3000$ (teal) shells. Right: optimized b-vectors for a sub-sampled space (minimum set with 7 and 15 measurements for theoretical optimal values, medium set with 20 and 28 measurements for more realistic subspaces). **c)** In this work we utilize a simulation-based inference approach to fit DTI and DKI tensors. Here, an artificial neural network is trained with an initial prior belief of the possible model parameters, combined with a large set of simulations from that model, to provide the most likely parameters given an input. Once the neural network has converged, experimental signals are provided to the network to obtain the most likely tensor parameters, which can then be used to, e.g., reconstruct the signal or obtain MD and FA.

and extraction of accurate parameter maps with minimal measurements, reducing acquisition time. For example, SBI achieves accurate parameter estimates with a 90% reduction in DTI and an 85% reduction in DKI acquisition times over traditional methods, with an accuracy loss (measured by the structural similarity index[26]) of only 18% and 20% for DTI and DKI, respectively. We showcase the capabilities of SBI and dw-MRI using simulated and real-patient data from the HCP[27] and MS dataset[28]. Importantly, we demonstrate that the SBI framework, despite being trained on simulations that either made no assumptions on the data or were only informed by healthy tissue, was capable of accurately mapping pathological tissue supporting its potential as a radiological tool. We believe this approach can substantially impact radiology in research and clinical settings, providing a framework to extract the maximum amount of microstructural information with minimum acquisition time and substantially improve accessibility to both basic and advanced dw-MRI. Therefore, this framework can be applicable both in clinical practice as well as research in three fundamental ways: *i)* by shortening the dw-MRI procedure for DTI and DKI, more patients can access these imaging modalities, even those who are unable to undergo long examinations; *ii)* increasing sensitivity to microstructure across clinical studies by fostering the employment of non-tensor-based dw-MRI protocols; and *iii)* rescuing older data, such as longitudinal studies acquired with minimal protocols and previously fitted with NLLS, where noise prevented the analysis from highlighting patterns of interest.

# Methods

## DTI and DKI theoretical frameworks

In DTI, diffusion is modeled as a Gaussian process using the diffusion tensor $D$, a $3 \times 3$ symmetric positive-definite matrix representing diffusion coefficients in three-dimensional space. The signal attenuation is given by:

$$S(b, \mathbf{g}) = S_0 \exp\left(-b \cdot \mathbf{g}^\top D \mathbf{g}\right) \tag{1}$$

where $S(b, \mathbf{g})$ is the measured signal with diffusion weighting factor $b$, gradient orientation $\mathbf{g}$, and $S_0$ is the non-diffusion-weighted signal. This framework allows estimation of the diffusion tensor to derive metrics like mean diffusivity (MD) and fractional anisotropy (FA) (see Fig. 1a). However, DTI's Gaussian assumption limits sensitivity to complex diffusion processes.

To overcome this limitation, DKI extends the DTI model by incorporating non-Gaussian diffusion effects, capturing the kurtosis of water diffusion. The signal attenuation in DKI is modeled as:

$$S(b, \mathbf{g}) = S_0 \exp\left(-b \cdot \mathbf{g}^\top D \mathbf{g} + \frac{1}{6} b^2 \left(\mathbf{g}^\top D \mathbf{g}\right)^2 K(\mathbf{g})\right) \tag{2}$$

where $K(\mathbf{g})$ is the kurtosis coefficient in orientation $\mathbf{g}$[3]. This model describes more complex diffusion dynamics, offering metrics like mean kurtosis (MK) for deeper microstructural insights. Additional DKI metrics include axial kurtosis (AK), radial kurtosis, mean kurtosis tensor (MKT), and kurtosis fractional anisotropy (KFA). The latter two, though not standard, are defined solely by the kurtosis tensors, helping to determine if we are accurately estimating the kurtosis tensor in addition to the diffusion tensor[29].

## Simulation-based inference fitting framework

In this work, we employed Sequential Neural Posterior Estimation (SNPE)[17,30], an approach in the family of SBI-methods to fit the DTI and DKI tensors. A simplified pipeline of this approach is depicted in Fig. 1c. SNPE uses neural networks to approximate the posterior distribution over model parameters. During training, the network is provided simulated data $\mathbf{x}$ and corresponding parameters $\boldsymbol{\theta}$ used to generate it. Specifically, parameters are sampled from a prior distribution, also known as the proposal distribution, representing our prior beliefs and passed through a simulator to produce the data. The neural network then minimizes the negative log-likelihood loss between the provided parameters $\boldsymbol{\theta}$ and the estimated posterior conditioned on the respective simulation $\mathbf{x}$. By iteratively refining the proposal distribution based on the learned posterior, SNPE efficiently explores the parameter space. Once trained, the network can infer the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{x}_{\text{obs}})$ for experimental data $\mathbf{x}_{\text{obs}}$, facilitating robust parameter estimation in models where traditional likelihood-based methods are impractical. By taking, for example, the mean of the posterior distribution or its mode, we can then infer a single value for the provided data. Additionally, this parameter estimation provides a measure of uncertainty (for example, the coefficient of variation or relative inter-quartile range) as the posterior represents a distribution of the parameters for the given experimental data[21].

## Human dw-MRI data

In this retrospective study, to verify the performance of the SBI framework on real-world data, we selected five individuals (three females and two males, ages 20-55) from the Human Connectome Project (HCP) database. Initial pre-processing was performed by the HCP Minimal Preprocessing Pipeline[27], while the study-specific pre-processing consisted of downsampling the images from 1.5 mm/voxel to 2.5 mm/voxel. The datasets consisted of two sets of b-values, $b = 1000$ and $b = 3000$, with 69 different gradient acquisitions for each b-shell.

To validate the framework on non-healthy individuals, we additionally took 5 individuals (4 females and 1 male, ages 26-55) with a diagnosis of multiple sclerosis (MS)[28]. This data, available through the public repository DIGITAL.CSIC and pre-processed with the same HCP pipeline, consisted of two sets of b-values, $b = 2000$ and $b = 4000$, with 30 and 59 gradient acquisitions, respectively. Images were also downsampled to 2.5mm/voxel.

## In silico data

Since SBI relies on simulated data to train neural networks and infer experimental parameters, the quality and realism of these simulations are crucial for achieving accurate results[31]. Therefore, for both DTI and DKI, the priors used to sample diffusion and kurtosis tensors were designed to mimic features of real human data.

For DTI simulations, parameters were sampled from uniform distributions and adjusted to ensure the diffusion tensor was positive definite and symmetric. It was found that the uniform distribution for the diagonal entries needed to be roughly twice as wide as that for the off-diagonal entries. With this approach, we were able to effectively replicate the HCP data features without prior knowledge (see Supplementary Fig. S1).

DKI simulations required a more sophisticated approach due to the non-linear relationship between DTI and DKI terms and challenges in defining suitable priors. Generating stable signals was crucial because DKI includes a positive $b^2$ term, which can lead to unstable simulations if $D$ and $K$ are not carefully chosen. To address this, prior distributions were obtained from an HCP subset to generate diffusion and kurtosis tensors replicating real-world non-linear relationships while remaining biologically feasible. More precisely, where appropriate, we fit log-normal or normal distributions to each of the entries of the diffusion and kurtosis tensors (see Supplementary Figs. S2b-d). Tensors producing unstable signals (approximately 15% of cases) were resampled.

We generated simulation data for both DTI and DKI using a Monte Carlo approach, randomly drawing biologically realistic tensors from the defined priors. True signals were corrupted by adding Rician noise with varying signal-to-noise ratios (SNR = [20, 10, 5, 2]). Simulations were split into training and testing sets; neural networks were trained on the training set to estimate tensor entries, and the testing set was used to compare fitted parameters of both SBI and NLLS against the available ground truth. All simulations and tensor-specific metric calculations were performed using `Dipy` functions[32]. For inference of the simulated parameters, 50,000 and 65,000 samples were generated for training for DTI and DKI, respectively. In the case of the real-world data 200000 samples were generated. Unless stated otherwise, simulation-based results were generated using 500 simulated signals with four different SNR levels. DKI priors were fitted using `scipy.stats` functions and SNPE functions were performed using the `sbi-toolbox` with all the pre-defined hyperparameters, network sizes and training routines[14].

## Number of acquisitions used

To test the performance of both SBI and NLLS across a different number of acquisitions, we introduce three different acquisition sets: **the full set**, which uses the full information of all acquisitions (69 in the case of the HCP and 89 in case of the MS dataset); **the minimum set** which uses the theoretically minimum number of acquisitions (1 $B_0$ measurement plus 6 and 15 measurements for DTI and DKI, respectively) and the **medium set**, which according to Jones[9] represents an optimal and realistic set of 20 orientations that should be sufficient for the DTI framework. In the case of the DKI framework, we used an additional 28 of the higher b-value measurements for the kurtosis tensor (13 more than the theoretical minimum). The distributions of the gradients of these sets can be seen in Fig. 1b. We emphasize that the gradient orientations of the simulated data (both for the full set as well as the subsets) were chosen to maximize electrostatic repulsion, but note that this choice was primarily made to enhance the results of NLLS. SBI could perform at similar levels even if the gradient orientations were not chosen optimally.

## Metrics to evaluate SBI and NLLS performance

In our in-silico simulations, access to ground truth allows direct comparison between estimated and true metrics. Specifically, DTI errors are defined as:

- MD Error= $|\mathrm{MD}_{\mathrm{true}} - \mathrm{MD}_{\mathrm{guess}}|$
- FA Error= $|\mathrm{FA}_{\mathrm{true}} - \mathrm{FA}_{\mathrm{guess}}|$

with MK, RK, and AK errors similarly defined in the DKI setting.

For human data, lacking ground truth, we compared results from minimal and medium acquisition sets against the full set. Utilizing images rather than isolated signals, we used structural information to assess performance. As a proxy for accuracy, we used the Structural Similarity Index (SSIM)[26], which measures similarity based on luminance, contrast, and structure. SSIM values closer to 1 indicate higher similarity; we set a minimum desired performance threshold of 65%. Preliminary analyses showed that images with SSIM above 0.65 preserved essential structural features necessary for clinical interpretation, such as lesion boundaries and tissue interfaces, ensuring minimal perceptible differences for clinicians.

As a proxy for precision, we calculated the standard deviation (std) of the fitted metric in white matter (for healthy individuals) or lesions (for pathological subjects). We aimed for the std of minimal acquisition fits to be within 10% of the full acquisition results.

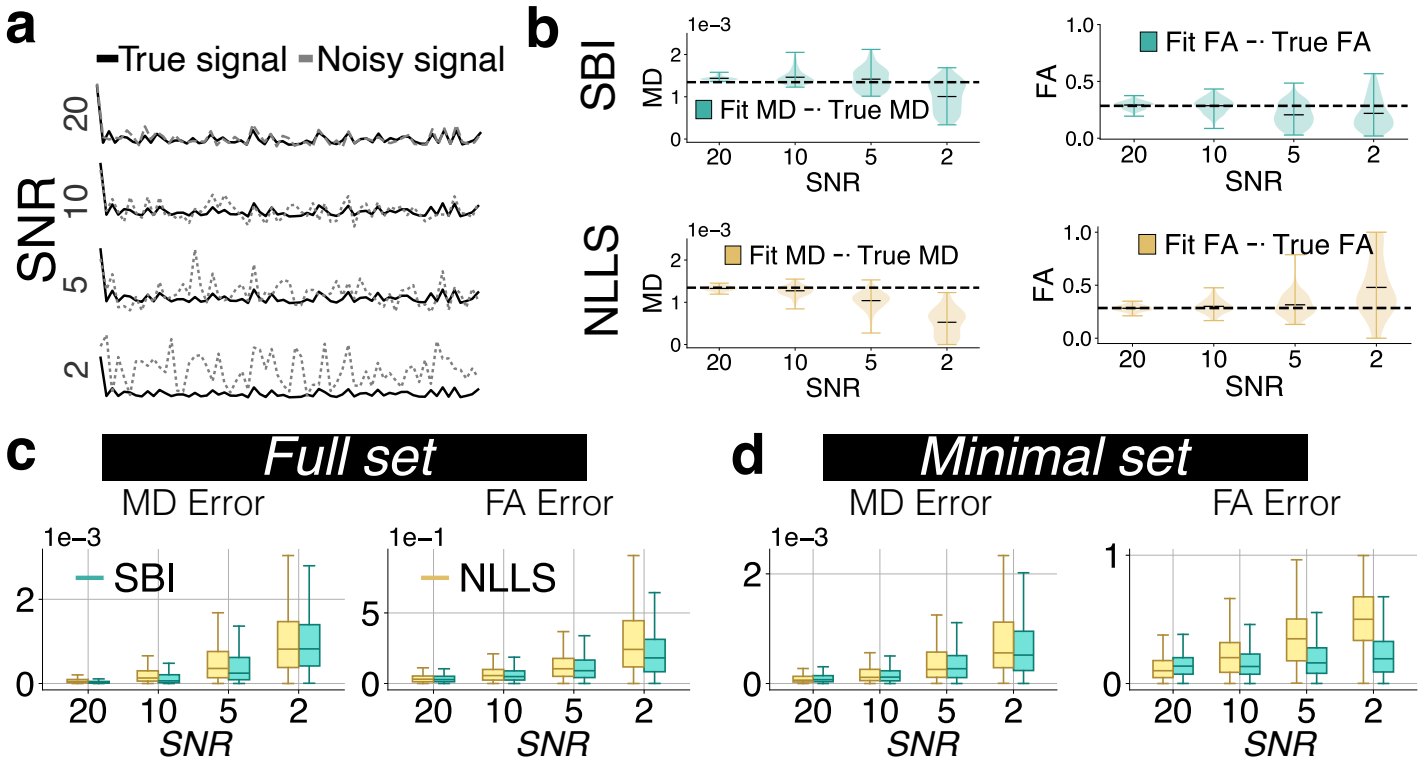


Figure 2: **SBI applied to simulated data shows more accurate performance than standard NLLS for DTI. a)** dw-MRI data is affected by varying levels of noise, which impacts signal clarity. The effect of different signal-to-noise ratios (SNRs, ranging from 2 to 20; dashed gray) is shown in comparison to the true signal (black). **b)** The accuracy of FA and MD estimates, obtained using the SBI (top) and NLLS (bottom) approach. The black dashed line indicates the true parameter values while the black solid line in violin-plot refers to mean. **c-d)** Using either the full set of measurements (c) or the minimum (d), both NLLS and SBI are fitted to 500 simulated signals and the errors in MD and FA are shown. Box plots represent median values, with boxes showing IQR and whiskers extending to 1.5 times the IQR. Gold denotes NLLS values; teal denotes SBI values.

# Results

## Fitting DTI and DKI to simulated data

We began by evaluating the performance of SBI compared to NLLS in both DTI and DKI frameworks using simulated data. For DTI, we generated a single signal and created 200 noisy versions with increasing noise levels (Fig. 2a). Estimating MD and FA using both methods (Fig. 2b), we found that NLLS aligns with true values at SNRs above 10 but loses accuracy below this level. In contrast, SBI remains robust even at SNR = 2, exhibiting smaller variance.

Extending this analysis, we generated 500 simulated signals at four different SNR levels. We calculated the absolute differences between the estimated MD and FA from both methods against the ground truth. For the full measurement set (Fig. 2c), both methods performed similarly at high SNR, but at lower SNRs, SBI's errors were smaller than those of NLLS. With the minimal measurement set (Fig. 2d), MD remained relatively unaffected for both methods due to the signal magnitude being discernible with fewer measurements. However, FA estimates from NLLS significantly deteriorated compared to SBI, indicating difficulty in extracting key information from noise with limited data. Supplemental Fig. S4 further illustrates that at SNR = 10, SBI reconstructs the true signal more accurately than NLLS, which appears to fit the noise (higher corelation with the noisy signal than SBI).

Building on the DTI results, we applied SBI to the DKI model, which introduces non-linearities through the kurtosis tensor—making neural network approaches particularly advantageous. Using simulated DKI data replicating key characteristics of the HCP dataset (Supplementary Fig. S5), we first compared SBI and NLLS using the full set of acquisitions. With a single signal at SNR = 20, both methods reconstructed the signal effectively (Fig. 3a). However, analyzing 200 noisy versions revealed that SBI estimates were closer to the true values across most metrics and SNR levels (Fig. 3b). As noise increased, NLLS estimates diverged more from the ground truth, indicating greater sensitivity to noise.

We then generated 500 new signals to evaluate kurtosis metrics at various SNR levels (Fig. 3c). While accuracy for both methods declined with decreasing SNR, SBI consistently exhibited greater robustness to noise, outperforming NLLS across most metrics.

Next, we used the minimal set of acquisitions for the DKI inference. In reconstructing the signal (Fig. 3d), SBI achieved more accurate results despite limited data, especially at higher b-values where NLLS deviated noticeably from the true signal. Parameter estimates from this minimal acquisition scheme (Fig. 3e) showed decreased performance for both methods compared to the full set, particularly in KFA estimation, which NLLS misidentified across all noise levels. SBI, however, continued to outperform NLLS, demonstrating superior accuracy in parameters like MK, AK, and MKT, even with limited data and high noise.

When comparing results across the 500 signals (Fig. 3f), we observed that at high SNR, both methods performed worse with minimal data than with full measurements due to reduced information. Interestingly, as SNR decreased, NLLS fits from the minimal acquisition sometimes performed better than with full data, aligning with DKI's known sensitivity to noise where additional noisy measurements can exacerbate errors. In contrast, SBI maintained consistent performance regardless of the number of measurements, effectively extracting sufficient information even with minimal data.

## Fitting the Human Connectome Data

Building on SBI's effectiveness with simulated data, we applied it to real-world data from the HCP for both DTI and DKI models. Unlike the simulations where $S_0$ was set at 200, the network had to infer $S_0$ from the real data, which, however did not affect SBI's performance. Additionally, while simulations allowed optimization of gradient orientations, the real-world data had pre-set orientations. For the minimal and medium subsets, we selected orientations from the full set that were maximally different (using Euclidean norms) to compare NLLS and SBI, primarily benefiting NLLS as SBI performed well even without optimal orientation choices.

To visually evaluate SBI's fitting results and mimic clinical application, we first examined a medial axial slice from one HCP individual, comparing tensor metrics MD and FA derived from SBI and NLLS for DTI (Figs. 4a,c). With the full measurement set, both methods yielded nearly identical structures. However, discrepancies appeared in the MD map, particularly in the lateral ventricles where NLLS produced slightly lower values. Based on trends from in-silico data—where NLLS tended to underestimate MD — we argue that SBI likely provides a more accurate reflection of the true tensor metrics. For FA, both methods yielded similar results, with differences mainly due to noise.

Examining the minimal measurement results in Figs. 4b,d, the advantages of SBI become evident. Despite limited data, SBI reconstructs a MD map which is indistinguishable from its full counterpart, while NLLS results are significantly noisier and less resolved (SSIM values of 0.97 for SBI and 0.59 for NLLS when comparing full and minimal images). Even the more complex FA metric is effectively reproduced by SBI, whereas NLLS loses structural information.

Next, we trained a generalist neural network to fit DTI to five HCP individuals, simultaneously. Te results of the absolute error, SSIM, and white matter std of the metrics (when compared against the respective full fit) are shown in Fig. 4e. For SBI, the accuracy of both MD and FA remains well above the 66% minimum accuracy we set (with the minimum SBI MD

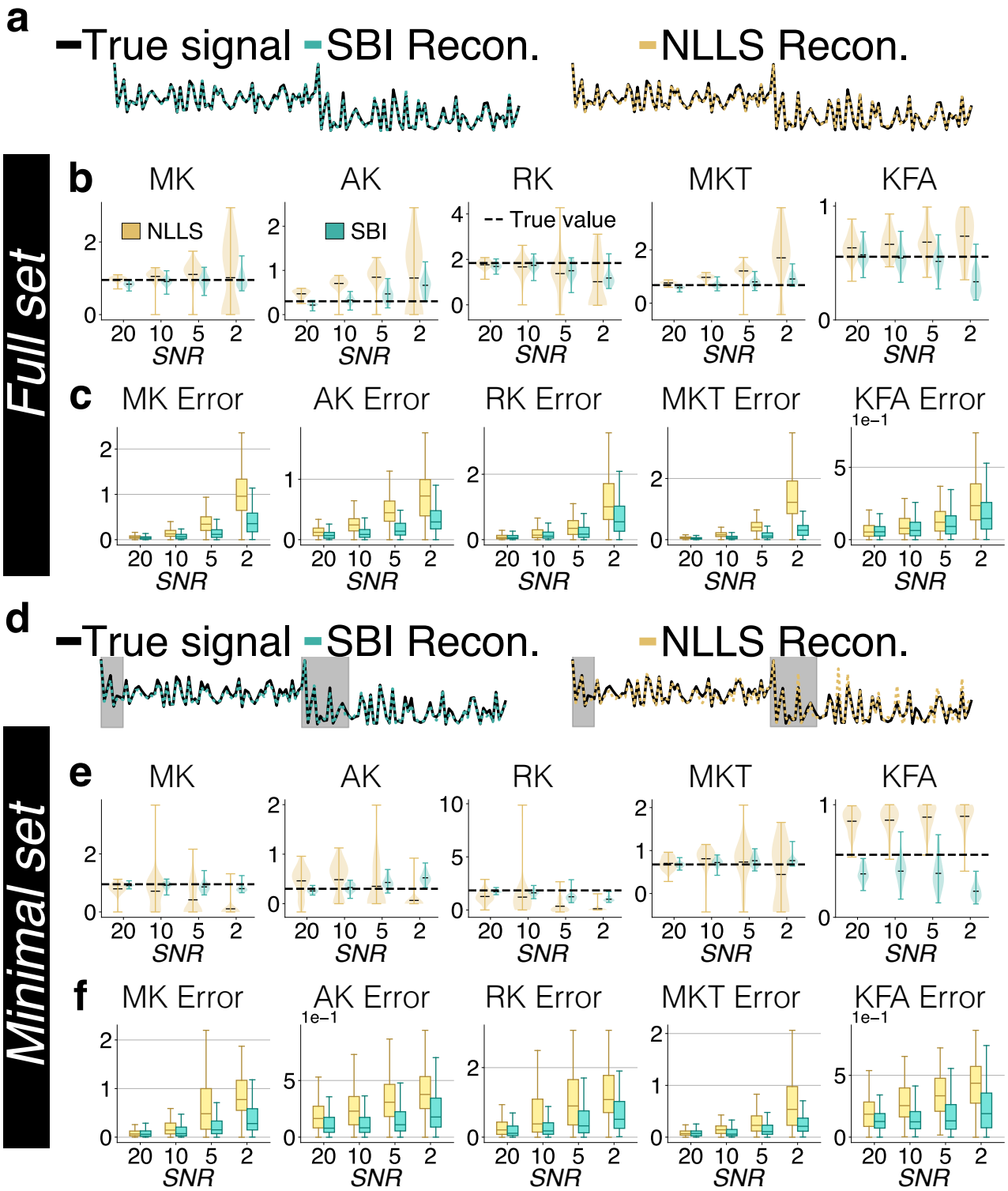


Figure 3: **SBI outperforms NLLS in the DKI framework.** **a)** SBI and NLLS are used to infer parameters and reconstruct an example signal using the full set of acquisitions (SNR of provided signal = 20). Both methods yield similar results with little difference. **b)** Applying varying noise levels (SNR=20, 10, 5, 2) to the same signal assesses the fitting routines' accuracy in estimating key parameters. SBI (teal) outperforms NLLS (gold) across different noise and parameter settings. The black dashed line indicates the true parameter values. Black solid line in violin-plot refers to mean. **c)** To evaluate performance across a range of simulated signals, 500 signals were generated and subjected to four noise levels. The resulting parameter fits were compared to the true parameters. SBI outperforms NLLS for all modalities (except KFA at SNR=20), and its errors increase more slowly with noise. Box plots represent median values, with boxes showing the IQR and whiskers extending to 1.5 times the IQR. Gold denotes NLLS values; teal denotes SBI values. **d-f)** Repeating the analysis in panels a–c with the minimum set of acquisitions, SBI continues to outperform NLLS even with reduced data.

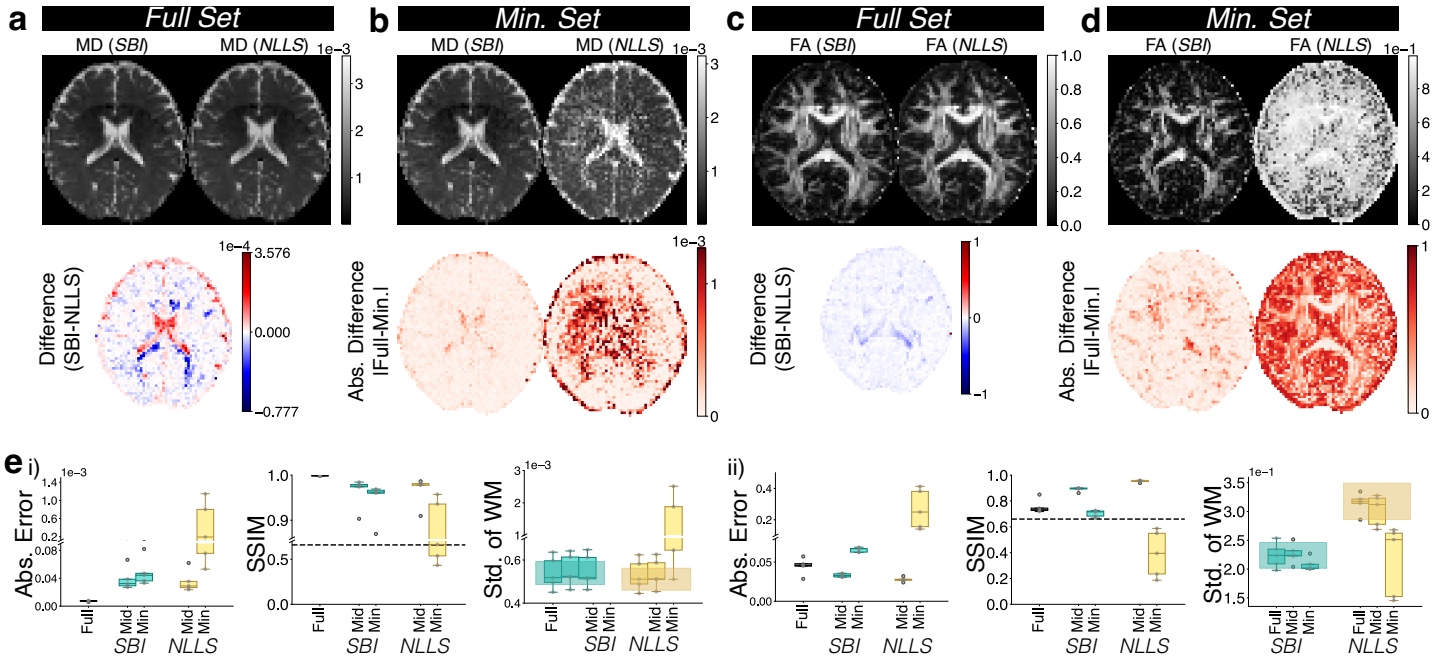


Figure 4: **Even with minimal information SBI is able to get meaningful information on real-world data where NLLS fails. a)** Comparison of MD values obtained with SBI and NLLS using the full set of acquisitions. Below we see a map of the differences between the two fits. **b)** Comparison of MD values obtained with SBI and NLLS using the minimal set. Below we see two maps that show the difference between the full and minimal fits for the SBI and NLLS implementations. **c-d)** Similar to a,b but instead of MD we show FA. **e)** Results of fitting MD and FA with SBI (one single network, teal) and NLLS (gold) using the full, medium and minimal set of acquisitions on the medial axial slice of 5 different individuals of the HCP dataset. For an objective comparison we present the absolute error of the MD and FA, the SSIM and the standard deviation of the respective metric in the white matter. Given a lack of an objective ground truth, the full NLLS and full SBI results are used for their respective subsets. A comparison between the full SBI and full NLLS is shown in the dark gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. The dashed black line in the SSIM plots refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished the standard deviation of the minimum fitting results to fall within.

| Modality | SSIM (mean, 90% CI) | SBI (HCP) | NLLS (HCP) | SBI (MS) | NLLS (MS) |
|---|---|---|---|---|---|
| DTI | MD | **0.95**, 0.91, 0.97 | **0.71**, 0.56, 0.86 | **0.96**, 0.92, 0.98 | **0.94**, 0.88, 0.98 |
| | FA | **0.7**, 0.68, 0.72 | **0.39**, 0.27, 0.5 | **0.64**, 0.5, 0.71* | **0.53**, 0.33, 0.68 |
| DKI | MK | **0.87**, 0.82, 0.92 | **0.01**, -0.05, 0.05 | **0.9**, 0.85,0.93 | **0.1**, 0.06, 0.15 |
| | AK | **0.71**, 0.67, 0.78 | **0.51**, 0.44, 0.57 | **0.64**, 0.51,0.71* | **0.59**, 0.51, 0.7 |
| | RK | **0.89**, 0.81, 0.96 | **0.13**, 0.02, 0.33 | **0.91**, 0.90,0.92 | **0.07**, 0.04,0.09 |

Table 1: **SSIM values for the boxplots the minimum set fitting results of the DTI and DKI results seen in Figs. 4, 5 and 6.** The bold number represents the mean SSIM while the two following numbers are the bootstrapped 90% confidence interval. *Low value is primarily driven by one outlier.

only losing approximately 5.6% SSIM accuracy with less than 10% of the available information) and far outperforms the minimum NLLS results (70% average SSIM MD accuracy loss).

Without ground truth, we compared each subset against its own full fit, plotting the . For both DTI and DKI, SBI maintained high accuracy even with less than 10% of the data, outperforming NLLS. In DKI, although medium NLLS results had smaller absolute errors than SBI, this did not necessarily imply better image reconstruction, as seen in the SSIM scores where minimal SBI fits had much higher SSIM values for MK and RK than NLLS. Assessing precision via the standard deviation of the fit metrics in white matter regions, SBI exhibited remarkable stability (within 10% of the full acquisition results), whereas NLLS values varied significantly with the number of acquisitions. Notably, NLLS's standard deviation for KFA with minimal acquisitions was significantly lower, but this was due to NLLS failing to fit KFA properly, assigning uniform values across the region. Therefore, considering all metrics together is essential for proper evaluation.

Next, we repeated this analysis for the DKI model. Importantly, the priors used for SBI training were derived from fits obtained from a different individual, demonstrating the method's generalizability across subjects. Using the full set of measurements, both NLLS and SBI produced similar qualitative results, highlighting consistent structural features across each kurtosis metric (Fig. 5a). Additionally, the kurtosis-specific metrics, such as MKT and KFA, were well resolved by both methods (see Fig. S7a), confirming accurate capture of both diffusion and kurtosis tensors. However, SBI demonstrated greater stability; NLLS resulted in gaps due to unrealistic or significantly negative values, visible as black dots within the brain region.

With the minimal set of acquisitions in Figs. 5b and S7b, SBI's effectiveness becomes even more apparent. Despite the reduced data comprising the theoretical minimum required, SBI achieved impressive results across most metrics. In contrast, NLLS failed under these conditions, breaking down and yielding unrealistically negative values. Comparing the differences between the full and minimal results, NLLS showed more structural discrepancies than SBI, indicating NLLS's inability to recreate accurate tensor metrics with limited data.

Again we we generated a generalist neural network capable of fitting DKI to five HCP individuals simultaneously, with the resulting absolute error, SSIM, and std of the metrics in Fig. 5d. The absolute error increase between the medium and minimum results of the SBI result are much smaller than those of the NLLS results, implying that SBI is more robust to less information (in line with results we saw in the DTI case). Additionally, SBI achieves higher SSIM scores than its NLLS counterpart. Finally, the precision of SBI is much more stable with less information than the results of NLLS. Notably, NLLS's standard deviation for KFA (Fig. S7d) with minimal acquisitions was significantly lower, but this was due to NLLS failing to fit KFA properly, assigning uniform values across the region. Therefore, considering all metrics together is essential for proper evaluation.

## SBI applied to MS individuals

As a final validation of SBI's performance with minimal acquisitions, we applied the DTI and DKI frameworks to data consisting of individuals with MS and brain lesions[28]. This allowed us to test whether our previously designed priors could generalize beyond the HCP dataset to pathological individuals.

An example axial slice is shown in Fig. 6a, with lesions highlighted. Figures 6b show the MD and FA values obtained by fitting the DTI framework using SBI and NLLS with the minimal number of acquisitions. Similar to the HCP results, when compared to fits from the full acquisitions (Supplementary Fig. S9a-b), SBI accurately reproduces MD and FA, while NLLS only reproduces MD with significant noise.

Next we trained a network to fit the DTI framework to five individuals simultaneously, selecting axial slices with the most lesions. For MD and FA, we calculated the absolute difference from full acquisition results (both across the whole image and within the lesions), the SSIM of the fitted images, and the std across the lesions, as shown in Fig. 6c. SBI outperforms NLLS in both MD and FA, achieving average SSIMs of 96% and 64%, respectively, compared to 9% and 53% for NLLS. While both methods perform well for FA in most cases, one individual showed a significantly lower SSIM for the minimal fit in both

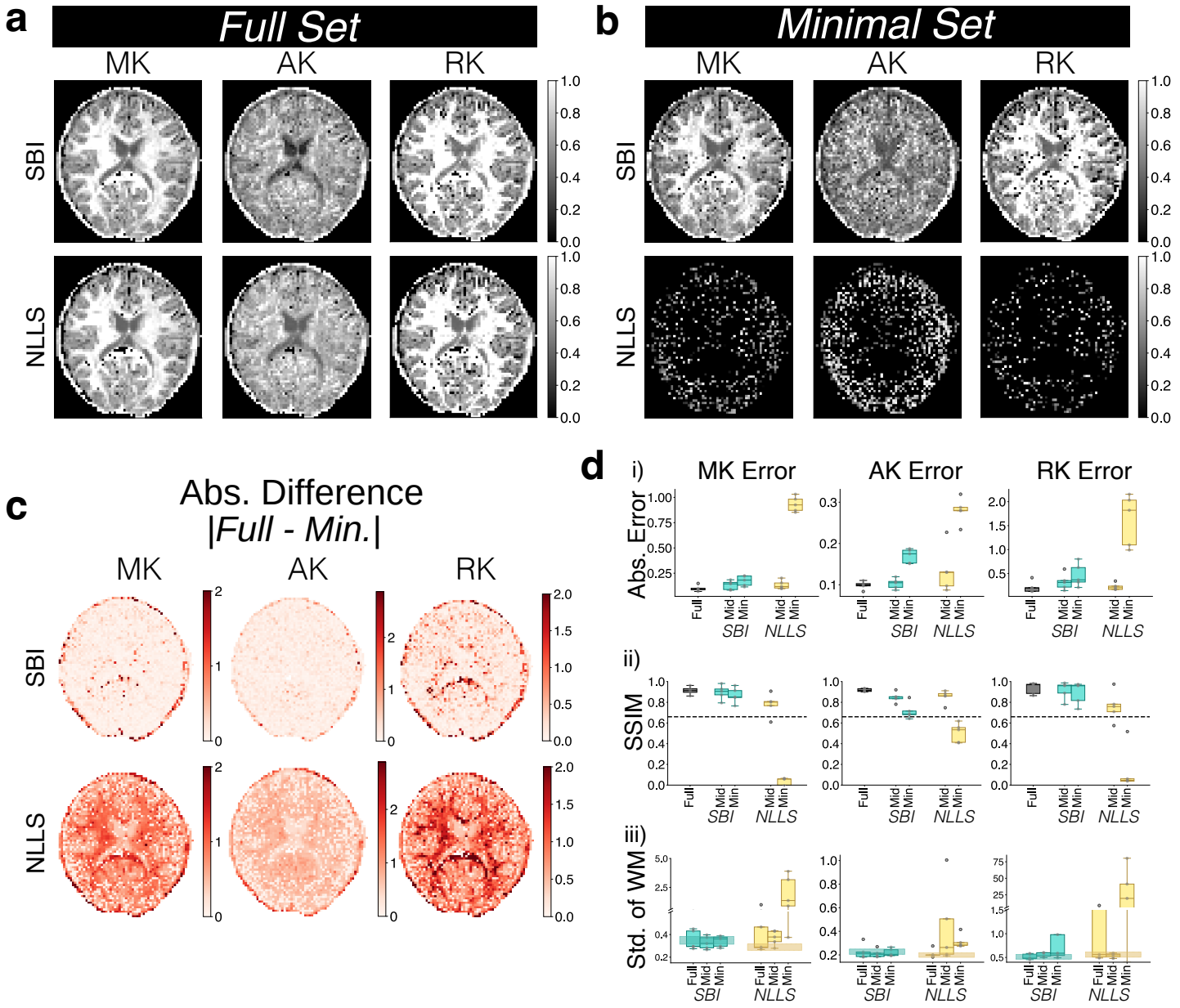


Figure 5: **SBI resolves kurtosis metrics even with minimal acquisitions. a)** Applying NLLS and SBI to the axial slice of an individual (different from the one used for the prior generation) and using the full set of acquisitions provides qualitative comparison of the fitted kurtosis tensor metrics. Note that black dots in the images indicate either large negative values or invalid numbers. **b)** Similar to panels a, but using only the minimal set of acquisitions. **c)** For each of the metrics we compute the absolute difference between the full and minimum set of SBI and NLLS. Darker red hues refer to larger differences. **d)** By fitting both SBI (teal) and NLLS (gold) using different amounts of orientations (Full set, medium and minimal) to the medial axial slice of 5 different individuals of the HCP dataset (using one single network), provides insight into the performance of the two methods. Similar to the DTI calculation, the absolute error (i), the SSIM of the axial slices (ii) as well as the standard deviation of the respective metric in the white matter (iii). Given a lack of an objective ground truth, the full NLLS (gold) and full SBI results (teal boxplots) are used as ground truths for their respective subsets. A comparison between the full SBI and full NLLS for the absolute error and SSIM is shown in the gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. Gold refers to the fit results the NLLS and teal to the SBI values. The dashed black line in (ii) refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished our minimum fitting results to fall within.

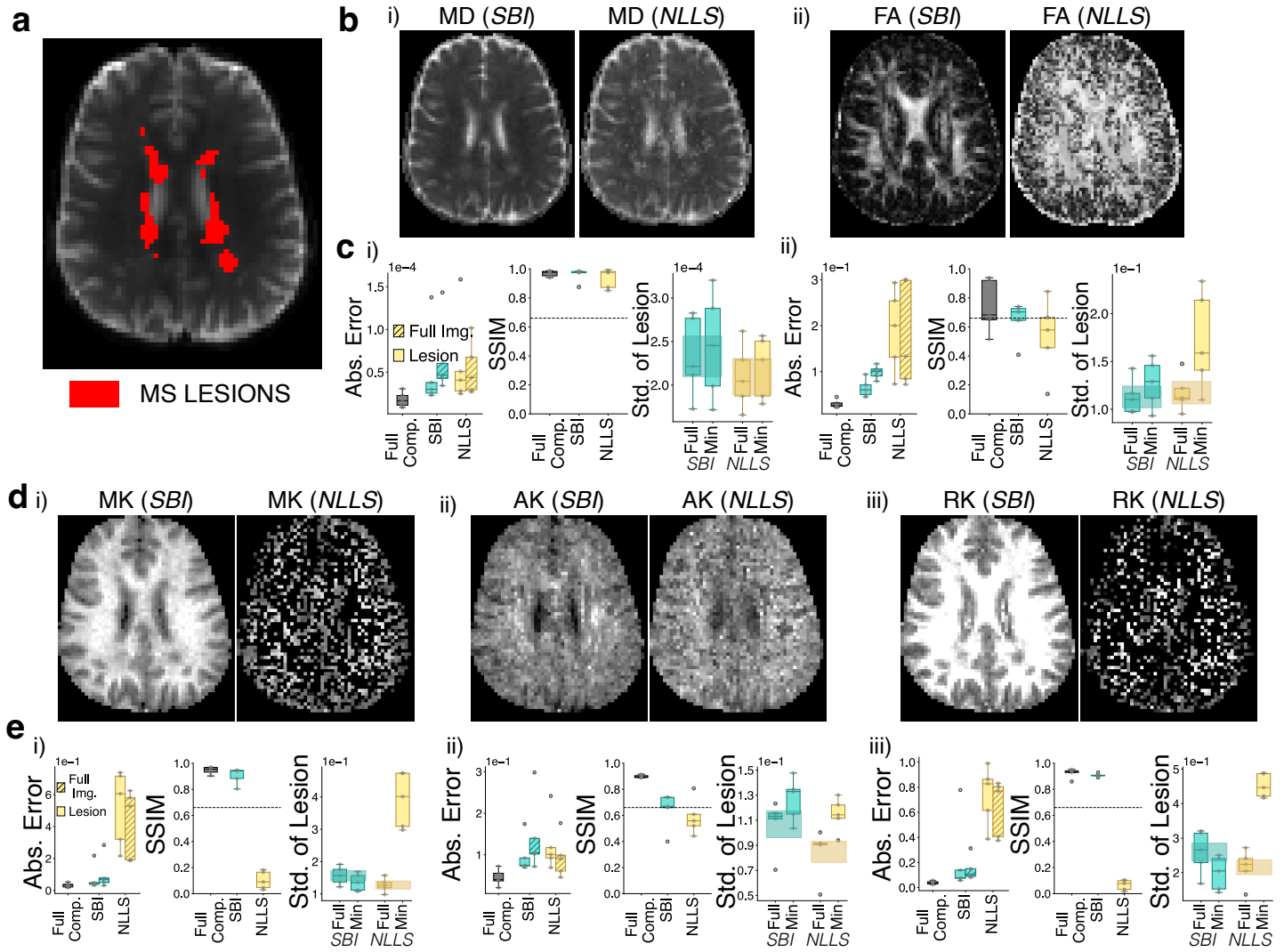


Figure 6: **The priors used to generate the accurate fittings in the HCP are also suitable to fit DTI and DKI using SBI in real-world data with MS lesions. a)** Example axial slice from the dataset of Cerdá et al.[28]. The red pixels represent locations of identified lesions. **b)** Comparison of MD and FA values obtained with SBI and NLLS using the minimal set of acquisitions. **c)** By fitting both SBI and NLLS using either the full or minimum set of acquisitions for 5 axial slices of different individuals of the MS dataset (using one single network), provides insight into the performance of SBI difference of MD (i) and FA (ii) across multiple patients. To study this performance, we calculated the absolute error of the MD and FA (both for the whole image and only the lesions - hatched), the SSIM and the standard deviation of the respective metric in the lesions. Given a lack of an objective ground truth, the full NLLS and full SBI results are used as ground truths for their respective subsets in the absolute error and SSIM metrics. A comparison between the full SBI and full NLLS is shown in the dark gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. Gold refers to the fit results the NLLS and teal to the SBI values. The dashed black line in the SSIM plots refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished our minimum fitting results to fall within. **d)** To evaluate the performance of the DKI fitting, we show the results of SBI and NLLS for MK, AK and RK using the minimum set of acquisitions (22). **e)** Similar to c, but now evaluating the performance of MK, AK and RK across 5 axial slices from 5 different individuals.

methods. Nonetheless, the structural integrity was sufficient for clinical lesion identification.

We then applied the same analysis to the DKI framework. As before, NLLS fails to fit the DKI model effectively with minimal data (see Fig. 6d and Supplementary Fig. S9c–e), whereas SBI successfully reconstructs metrics like MK, AK, and RK, allowing important insights into brain structure. Using a DKI generalist network on the same slices as mentioned above, SBI maintains high accuracy not achieved by NLLS (see Fig. 6e and Table 1). On average, SBI exhibits smaller absolute errors within lesions and across the whole image compared to NLLS. The standard deviation across lesions remains stable within our 10% target for SBI, but increases significantly in the NLLS setting.

# Discussion

In this work, we investigated how SBI, specifically SNPE performs in DTI and DKI frameworks under significant noise and minimal acquisitions. By generating extensive simulated datasets based on the mathematical formulations of these models, we trained neural networks for parameter estimation without relying on real data, thus avoiding issues related to data availability and ethical concerns[21,22,31]. This approach enhances parameter estimation accuracy while minimizing acquisition time, offering a robust and flexible tool for diffusion MRI analysis. Our results facilitate the use of advanced dw-MRI approaches like DKI in clinical applications, providing deeper insights than basic protocols typically used in clinical settings. Additionally, this framework allows us to rescue older data previously hindered by noise. Moreover, this method is versatile and can be extended to any diffusion model by training the neural network with an appropriate simulator, as demonstrated with SandiX[21].

Neural network approaches have been crucial in recent dw-MRI studies, including combining Bayesian frameworks, fiber tracking, and enhancing resolution[24,33–35]. Some works[25,36] considered using fewer acquisitions but relied on real data for training, which can introduce issues with data availability, bias, and generalizability.

Our work was motivated by studies combining dw-MRI and simulation-based methods. For example, Jallais and Palombo[21] demonstrated that SNPE is effective for parameter estimation in dw-MRI, providing quantifiable measures of model degeneracy not available with traditional methods. Other studies using SNPE include Hashemi et al.[37] for epilepsy models and Vasylechko et al.[38] for intravoxel incoherent motion (IVIM) models.

Our approach distinguishes itself by combining SNPE with dw-MRI to study DKI beyond the linear DTI framework, assessing different noise levels and evaluating minimal acquisition protocols without requiring real data for training. This offers a flexible and efficient framework for accurate parameter estimation when acquisition time is limited. Unlike previous SNPE approaches that use summary statistics via multi-layer perceptrons (MLPs)[20,21], we directly input the diffusion signal to the network, avoiding extra network training and reducing computational complexity. Two caveats arise: *i)* our neural networks are specific to the acquisition scheme, and *ii)* the input dimension can be large with full acquisitions. We addressed *i)* by providing an identifier for the acquisition scheme, and *ii)* computational resources remained manageable even with full acquisitions. Future work will aim to incorporate summary statistics to retain simplicity and interpretability while gaining generalization over acquisition numbers and directions.

SBI has limitations requiring future investigation. For example, it needs prior distributions to generate simulations[31]. For DTI, we used simple uniform distributions needing no pre-tuning while for DKI, priors initially required knowledge of parameter distributions from one individual. However, we showed these priors generalized across other individuals and pathological subjects[28], even with different b -values and protocols, suggesting minimal prior knowledge may suffice. Exploring priors for more complex models remains for future research. While DKI is a mathematical representation, biophysical models used to interpret dw-MRI signals are based on biological quantities, easing the definition of acceptable priors. Another limitation is inference time compared to NLLS; sampling the posterior 500 times, SBI inference took about 1 minute for the slice seen in Fig. 4, while NLLS operates in seconds. However, with parallelization and optimization, computation time can be substantially reduced.

In summary, this work demonstrates how data-friendly SBI methods can minimize acquisition times in both research and clinical settings. By reducing acquisition times by up to 90% in DTI and 85% in DKI while maintaining high accuracy, SBI addresses the key limitation of long scan durations in MRI. SBI's robustness to noise and effectiveness with fewer data points make it especially valuable where high-quality data acquisition is challenging. Demonstrating efficacy in both DTI and DKI, a more complex model sensitive to noise, opens avenues for impactful clinical research and practice in diffusion-weighted MRI and beyond.

# Data and code availability

All data and code is available in the github repository: *https://github.com/TIB-Lab/SBIDTI*. Within that repository we also provide a link to the trained networks that were used in this work.

The MS Data used in the preparation of this work were obtained at the Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH) for the NIH 1R21NS123419 Project (Principal Investigator: Caterina Mainero, M.D., Ph.D.). This data can be found at *https://digital.csic.es/handle/10261/340684*.

# References

[1] Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 13(4):534–546, 2001.

[2] Peter J Basser, James Mattiello, and Denis LeBihan. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[3] Jens H Jensen, Joseph A Helpern, Anita Ramani, Hanzhang Lu, and Kyle Kaczynski. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(6):1432–1440, 2005.

[4] Dow-Mu Koh and David J Collins. Diffusion-weighted mri in the body: applications and challenges in oncology. *American Journal of Roentgenology*, 188(6):1622–1635, 2007.

[5] Derek K Jones and Peter J Basser. "squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted mr data. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(5):979–993, 2004.

[6] Jonathan AD Farrell, Bennett A Landman, Craig K Jones, Seth A Smith, Jerry L Prince, Peter CM Van Zijl, and Susumu Mori. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 t. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 26(3):756–767, 2007.

[7] Jelle Veraart, Els Fieremans, and Dmitry S Novikov. Diffusion mri noise mapping using random matrix theory. *Magnetic resonance in medicine*, 76(5):1582–1593, 2016.

[8] Shreyas Fadnavis, Joshua Batson, and Eleftherios Garyfallidis. Patch2self: Denoising diffusion mri with self-supervised learning. *Advances in Neural Information Processing Systems*, 33:16293–16303, 2020.

[9] Derek K Jones. The effect of gradient sampling schemes on measures derived from diffusion tensor mri: a monte carlo study. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 51(4):807–815, 2004.

[10] Cheng Guan Koay and Peter J Basser. Analytically exact correction scheme for signal extraction from noisy magnitude mr signals. *Journal of magnetic resonance*, 179(2):317–322, 2006.

[11] Jelle Veraart, Jan Sijbers, Stefan Sunaert, Alexander Leemans, and Ben Jeurissen. Weighted linear least squares estimation of diffusion mri parameters: strengths, limitations, and pitfalls. *Neuroimage*, 81:335–346, 2013.

[12] Johann Brehmer. Simulation-based inference in particle physics. *Nature Reviews Physics*, 3(5):305–305, 2021.

[13] Jan Boelts, Jan-Matthis Lueckmann, Richard Gao, and Jakob H Macke. Flexible and efficient simulation-based inference for models of decision-making. *Elife*, 11:e77220, 2022.

[14] Alvaro Tejero-Cantero, Jan Boelts, Michael Deistler, Jan-Matthis Lueckmann, Conor Durkan, Pedro J Gonçalves, David S Greenberg, and Jakob H Macke. Sbi–a toolkit for simulation-based inference. *arXiv preprint arXiv:2007.09114*, 2020.

[15] Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010.

[16] George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *The 22nd international conference on artificial intelligence and statistics*, pages 837–848. PMLR, 2019.

[17] Jan-Matthis Lueckmann, Pedro J Goncalves, Giacomo Bassetto, Kaan Öcal, Marcel Nonnenmacher, and Jakob H Macke. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, 30, 2017.

[18] David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019.

[19] Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife*, 9:e56261, 2020.

[20] Maëliss Jallais, Pedro Luiz Coelho Rodrigues, Alexandre Gramfort, and Demian Wassermann. Inverting brain grey matter models with likelihood-free inference: a tool for trustable cytoarchitecture measurements. *Journal of Machine Learning for Biomedical Imaging*, pages 1–27, 2022.

[21] Maëliss Jallais and Marco Palombo. $\mu$guide: a framework for quantitative imaging via generalized uncertainty-driven inference using deep learning. *eLife*, 13, 2024.

[22] João P de Almeida Martins, Markus Nilsson, Björn Lampinen, Marco Palombo, Peter T While, Carl-Fredrik Westin, and Filip Szczepankiewicz. Neural networks for parameter estimation in microstructural mri: application to a diffusion-relaxation model of white matter. *NeuroImage*, 244:118601, 2021.

[23] William Consagra, Lipeng Ning, and Yogesh Rathi. A deep learning approach to multi-fiber parameter estimation and uncertainty quantification in diffusion mri. *arXiv preprint arXiv:2405.13655*, 2024.

[24] Vladimir Golkov, Alexey Dosovitskiy, Jonathan I Sperl, Marion I Menzel, Michael Czisch, Philipp Sämann, Thomas Brox, and Daniel Cremers. Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE transactions on medical imaging*, 35(5):1344–1351, 2016.

[25] Abhishek Tiwari, Rajeev Kumar Singh, and Saurabh J Shigwan. Swindti: swin transformer-based generalized fast estimation of diffusion tensor parameters from sparse data. *Neural Computing and Applications*, 36(6):3179–3196, 2024.

[26] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[27] David C Van Essen, Kamil Ugurbil, Edward Auerbach, Deanna Barch, Timothy EJ Behrens, Richard Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[28] Antonio Cerdán Cerdá, Nicola Toschi, Constantina A Treaba, Valeria Barletta, Elena Herranz, Ambica Mehndiratta, Jose A Gomez-Sanchez, Caterina Mainero, and Silvia De Santis. A translational mri approach to validate acute axonal damage detection as an early event in multiple sclerosis. *Elife*, 13:e79169, 2024.

[29] Brian Hansen and Sune Nørhøj Jespersen. Data for evaluation of fast kurtosis strategies, b-value optimization and exploration of diffusion mri contrast. *Scientific data*, 3(1):1–5, 2016.

[30] Jan-Matthis Lueckmann, Jan Boelts, David Greenberg, Pedro Goncalves, and Jakob Macke. Benchmarking simulation-based inference. In *International conference on artificial intelligence and statistics*, pages 343–351. PMLR, 2021.

[31] Noemi G Gyori, Marco Palombo, Christopher A Clark, Hui Zhang, and Daniel C Alexander. Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. *Magnetic resonance in medicine*, 87 (2):932–947, 2022.

[32] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, Ian Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

[33] Marco Reisert, Elias Kellner, Bibek Dhital, Juergen Hennig, and Valerij G Kiselev. Disentangling micro from mesostructure by diffusion mri: a bayesian approach. *Neuroimage*, 147:964–975, 2017.

[34] David Abramian, Anders Eklund, and Evren Özarslan. Super-resolution mapping of anisotropic tissue structure with diffusion mri and deep learning. *bioRxiv*, pages 2023–04, 2023.

[35] Daniele Ravi, Nooshin Ghavami, Daniel C Alexander, and Andrada Ianus. Current applications and future promises of machine learning in diffusion mri. In *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain, September 2018 22*, pages 105–121. Springer, 2019.

[36] Eric Aliotta, Hamidreza Nourzadeh, and Sohil H Patel. Extracting diffusion tensor fractional anisotropy and mean diffusivity from 3-direction dwi scans using deep learning. *Magnetic Resonance in Medicine*, 85(2):845–854, 2021.

[37] Meysam Hashemi, Anirudh N Vattikonda, Jayant Jha, Viktor Sip, Marmaduke M Woodman, Fabrice Bartolomei, and Viktor K Jirsa. Amortized bayesian inference on generative dynamical network models of epilepsy using deep neural density estimators. *Neural Networks*, 163:178–194, 2023.

[38] Serge Didenko Vasylechko, Simon K Warfield, Onur Afacan, and Sila Kurugol. Self-supervised ivim dwi parameter estimation with a physics based forward model. *Magnetic resonance in medicine*, 87(2):904–914, 2022.

## Acknowledgments

We acknowledge the support of the Institute of Neuroscience, CSIC-UMH. We thank Pedro Gonçalves and Caterina Mainero for fruitful discussions and guidance. This research was funded by "la Caixa" Foundation with Project-ID LCF/BQ/PI23/11970039 (MFE). SDS was supported by the the Spanish Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación (PID2021-128909NA-I00 and CNS2023-14488), by the Programs for Centres of Excellence in R&D Severo Ochoa (CEX2021-001165-S), and by the Generalitat Valenciana through a Subvencion para la contratación de investigadoras e investigadores doctores de excelencia 2021 (CIDEGENT/2021/015).

## Competing Interests

The authors declare no competing interests

## Supplemental material

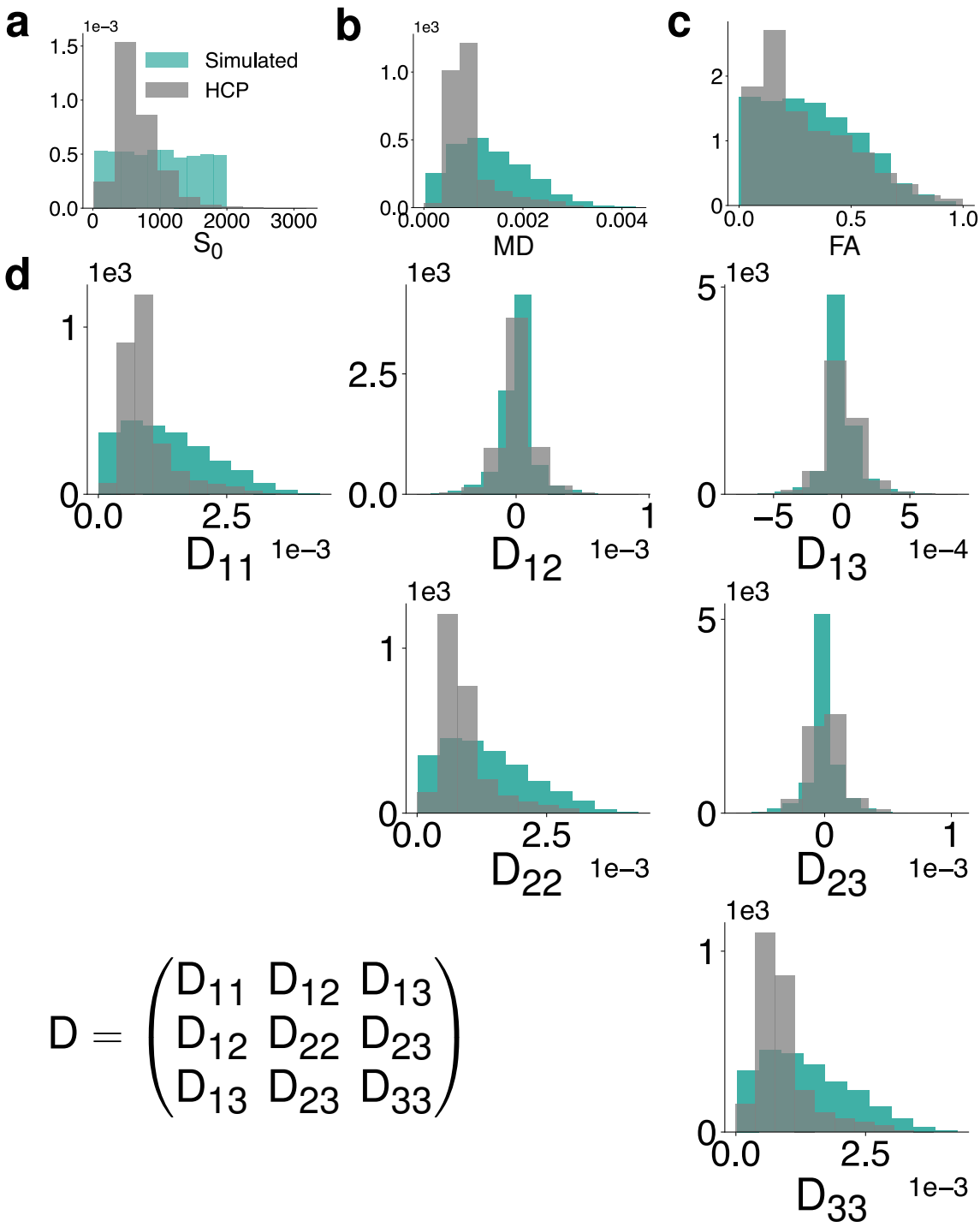


Figure S1: **The distributions of key features in the simulated DTI signals (teal) closely mimic those of the real dataset (gray). a)** Distribution of $S_0$. **b)** Distribution of the mean diffusivity. **c)** Distribution of the fractional anisotropy. **d)** The entries of the simulated diffusion tensor matrix (shown in the bottom left corner), generated as described in the Methods section, span the space of the DTI entries of the HCP data obtained with a NLLS fit.

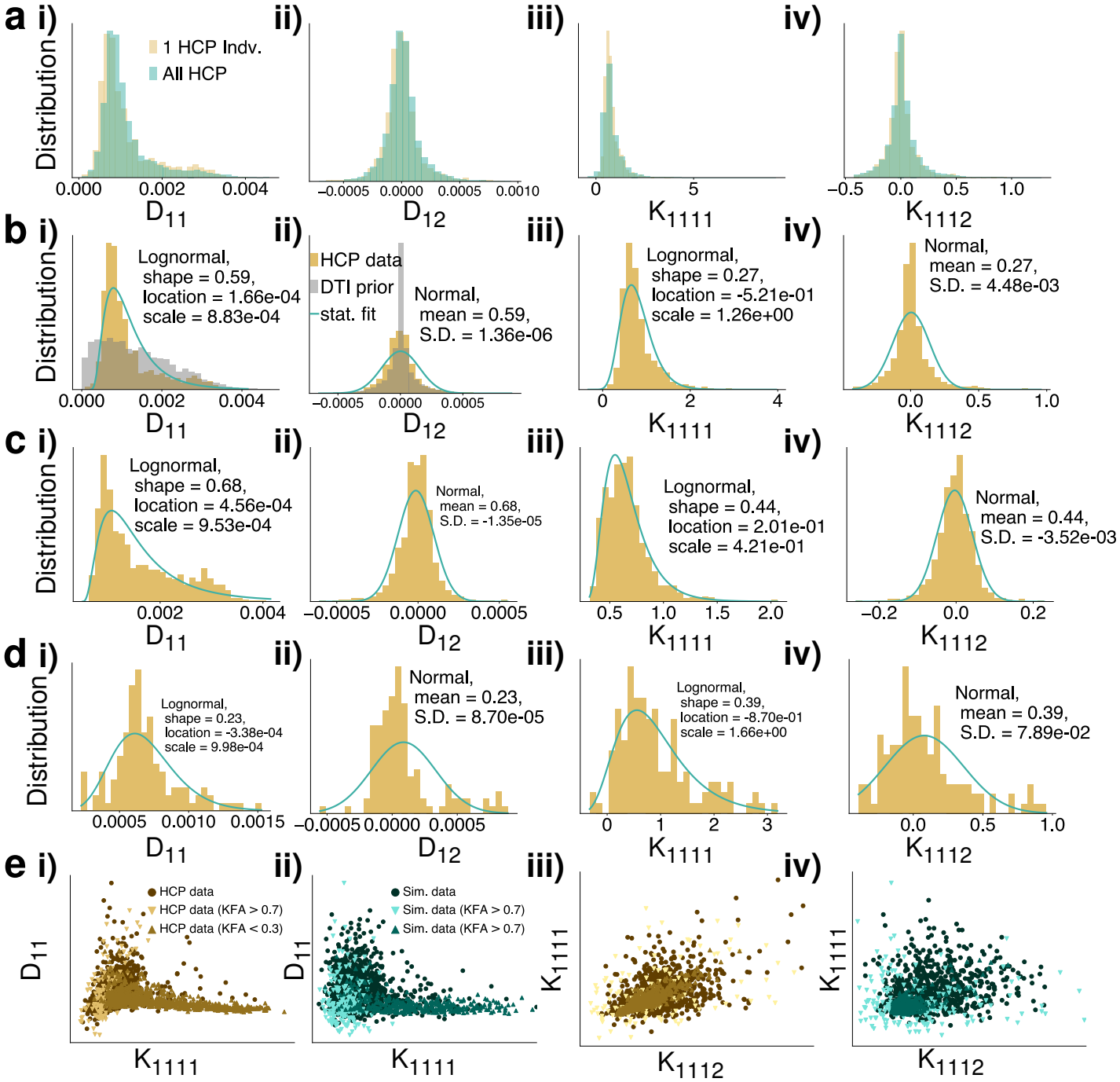


Figure S2: **The DTI and DKI metrics are distributed either normally or lognormally and demonstrate complex non-linear relationships.** **a)** Distributions of $D_{11}, D_{12}, K_{1111}$ and $K_{1112}$ terms of the a single individual (gold) and all individuals (teal) combined from the HCP dataset. **b)** Distribution of the $D_{11}, D_{12}, K_{1111}$ and $K_{1112}$ terms of the single HCP individual (fit with NLLS) in gold. Gray in i) and ii) depicts the results of the simulated DTI values if we sampled from a uniform distribution. Teal line is the best log-normal or normal fit. **c)** Same as b, except that tensors that have KFA >0.3 are filtered out. **d)** Same as b, except that tensors that have KFA <0.7 are filtered out. **e)** Plotting the true $D_{11}$ and the $K_{1111}$ (i), simulated $D_{11}$ and the $K_{1111}$ (ii), true $K_{1111}$ and $K_{1122}$ (iii) and simulated $K_{1111}$ and $K_{1122}$ (iv) entries of the HCP tensors against each reveals a complicated highly non-linear structure, with distinct regions for different KFA values.

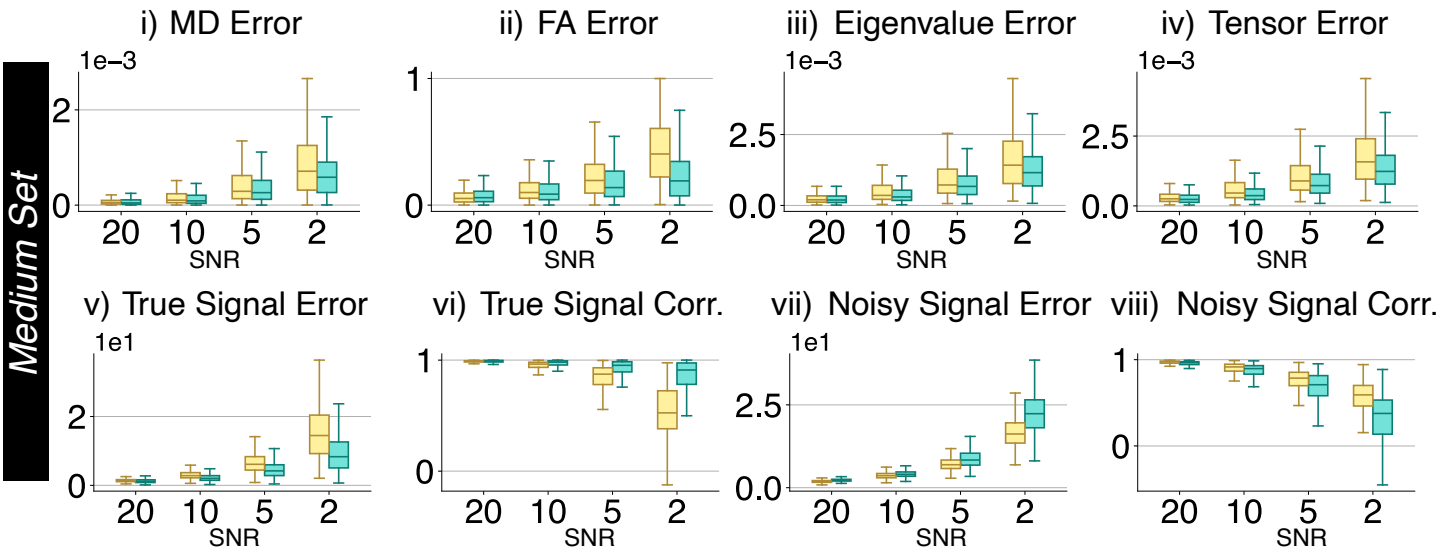


Figure S3: **Similarly to Figure 2, we can perform NLLS and SBI fits using only the medium set of acquisitions. Once again, SBI achieves a more accurate performance.** Both NLLS and SBI are fitted to 500 simulated signals while only being provided the medium subset of acquisitions. The obtained DTI tensors are then compared across eight different metrics: *i)* Difference in MD, *ii)* Difference in FA, *iii)* Difference in the sorted eigenvalues, *iv)* Difference in the Frobenius norm of the tensors, *v)* Integrated difference between the true (noiseless) signal and the reconstructed signal, *vi)* Pearson correlation between the reconstructed signal and the true (noiseless) signal, *vii)* Integrated difference between the provided noisy signal and the reconstructed signal, and *viii)* Pearson correlation between the reconstructed signal and the provided noisy signal. The box plots represent the median values, with the boxes showing the IQR and the whiskers extending to 1.5 times the IQR. Gold refers to the NLLS values, and teal to the SBI values.

## Additional DTI errors:

Given a true estimate of the diffusion tensor $D_{\text{true}}$ and a guess for the diffusion tensor $D_{\text{guess}}$, additional tensor specific errors in the DTI setting were calculated as follows:

- Eigenvalue Error= $\sqrt{\sum_{i=1}^{3}(\lambda_{\text{true},i} - \lambda_{\text{guess},i})^2}$, where the eigenvalues of the true and guess tensor were sorted according to their magnitude.
- Tensor error = $\sqrt{\sum_{i=1}^{3}\sum_{j=1}^{3}(D_{\text{true}} - D_{\text{guess}})_{i,j}^2}$

The signal specific errors are, after reconstructing the guess signal $S_{\text{guess}}$ to recreate the true signal $S_{\text{true}}$ given the noisy signal $S_{\text{noisy}}$, defined as:

- $X$ signal Error = $\sqrt{(S_{\text{guess}} - S_{\text{X}})^2}/N$, where $X$ refers to the noisy or true signal and $N$ is the number of measurements for the fitting.
- $X$ signal corr. is the Pearson correlation coefficient between the guess signal and the noisy or true signal.

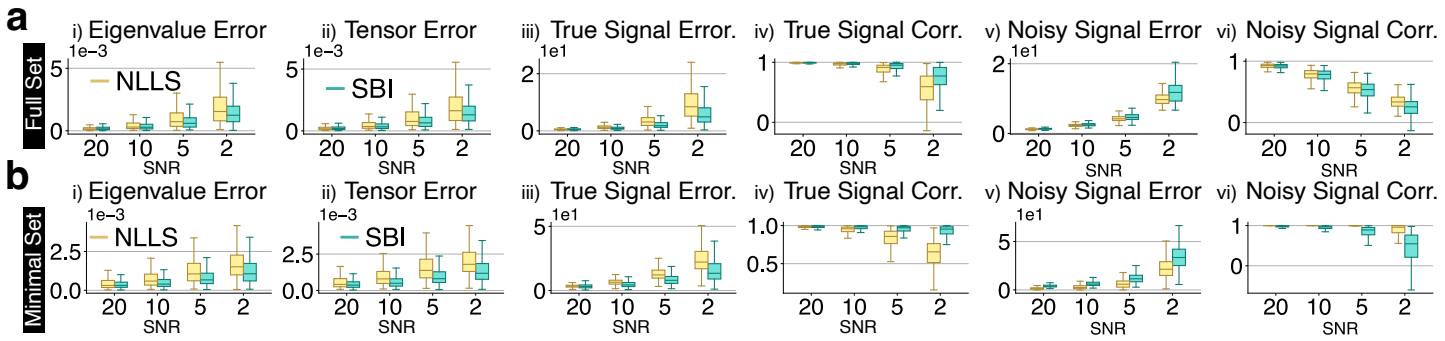


Figure S4: **Additional metrics provide insights into the decrease in performance of NLLS versus SBI. a,b)** Using either the full set of measurements (a) or the minimum (b), both NLLS and SBI are fitted to 500 simulated signals. The resulting DTI tensors are compared across six metrics: *i)* difference in sorted eigenvalues, *ii)* difference in Frobenius norm of tensors, *iii)* integrated difference between true (noiseless) and reconstructed signals, *iv)* Pearson correlation between reconstructed and true signals, *v)* integrated difference between noisy and reconstructed signals, and *vi)* Pearson correlation between reconstructed and noisy signals. Box plots represent median values, with boxes showing IQR and whiskers extending to 1.5 times the IQR. Gold denotes NLLS values; teal denotes SBI values.

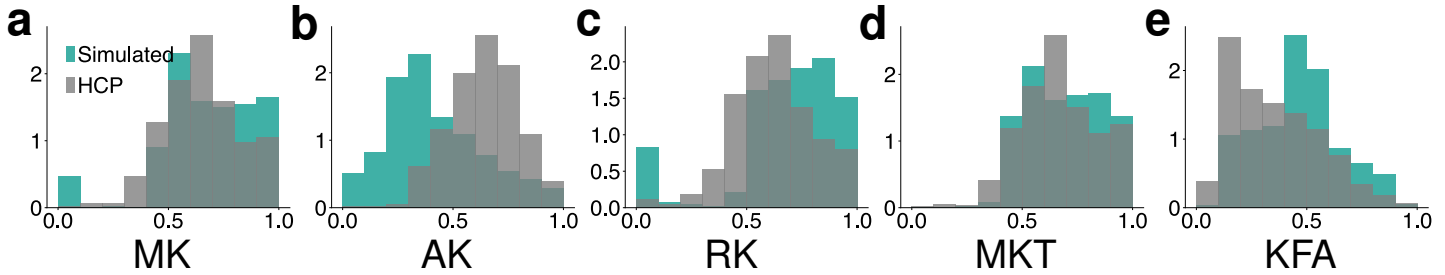


Figure S5: **The distributions of key features in the simulated DKI (teal) closely mimic those of the real dataset (gray). a)** Distribution of $S_0$. **b)** Distribution of the mean diffusivity. **c)** Distribution of the fractional anisotropy. **d)** The entries of the simulated diffusion tensor (DT) matrix (shown in the bottom left corner), generated as described in the Methods section, span the space of the DTI entries of the HCP data obtained with a NLLS fit.

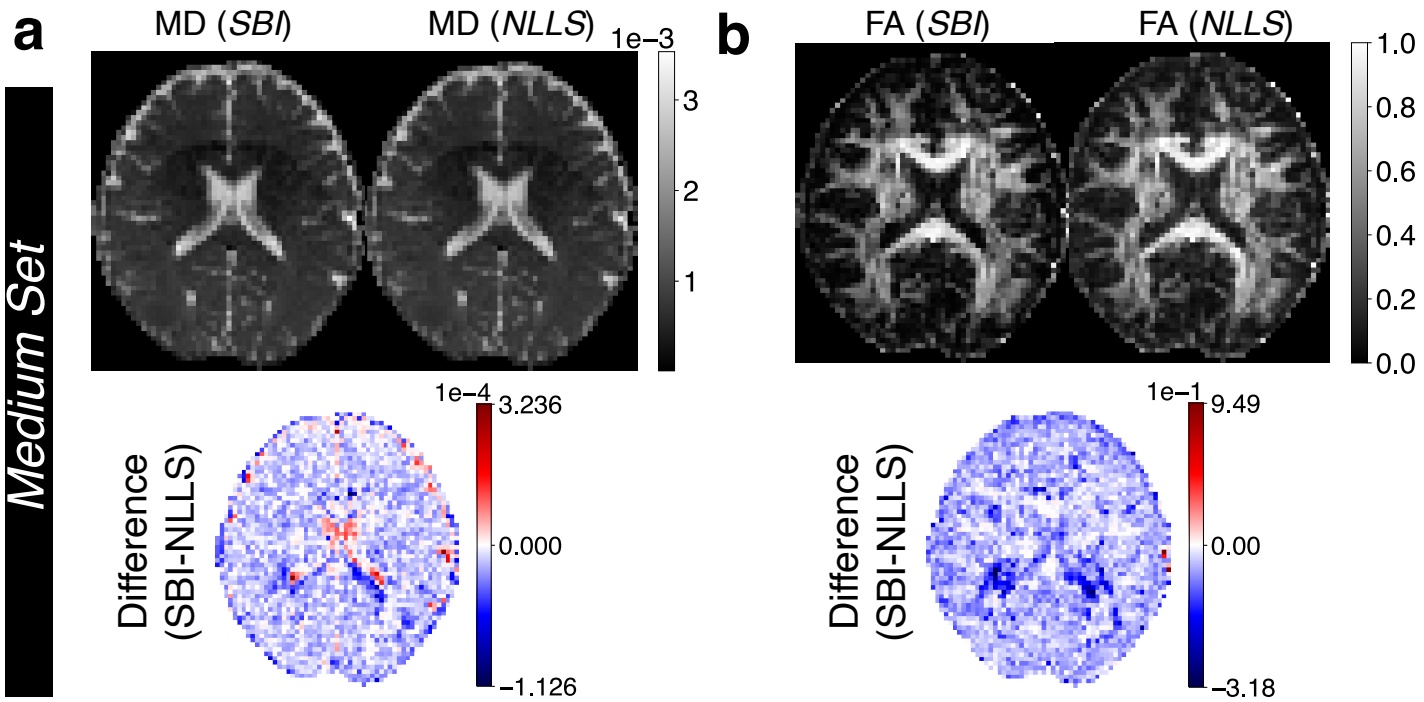


Figure S6: **Application of the SBI and NLLS fitting using the medium set. a)** Mean diffusivity obtained with SBI and NLLS, respectively. **c)** Fractional anisotropy.

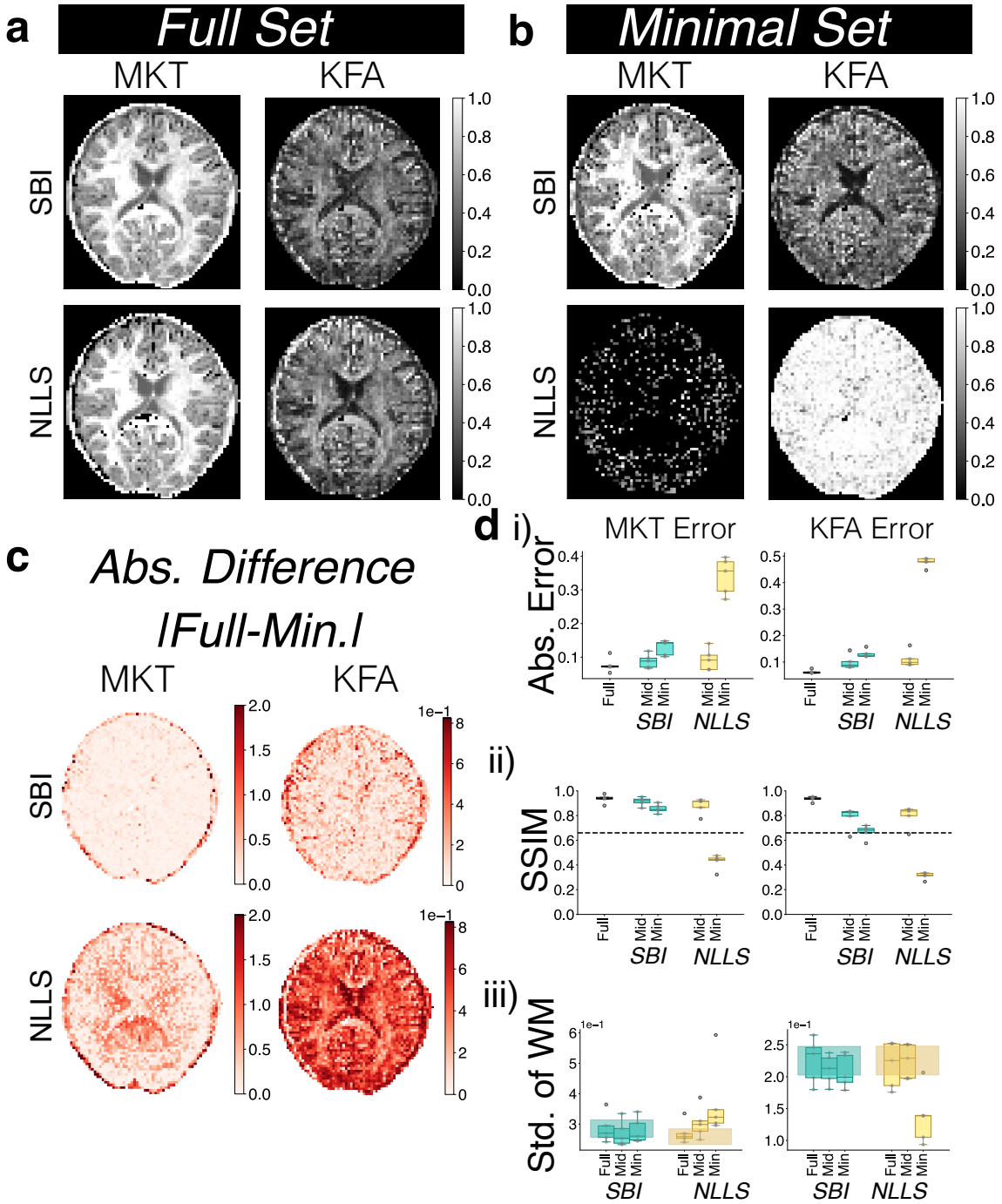


Figure S7: **SBI is able to resolve kurtosis tensor specific metrics even with minimal acquisitions. a)** Applying NLLS and SBI to the axial slice of an individual (different from the one used for the prior generation) and using the full set of acquisitions provides qualitative comparison of the mean of the kurtosis tensor and the kurtosis fractional anisotropy. **b)** Similar to panels a), but using only the minimal set of acquisitions. **c)** For each of the metrics we compute the absolute difference between the full and minimum set of SBI and NLLS. Darker red hues refer to larger differences. **d)** By fitting both SBI and NLLS using different amounts of orientations (Full set, medium and minimal) the medial axial slice of 5 different individuals of the HCP dataset (using one single network), provides insight into the performance of the two methods. Similar to the DTI calculation, the absolute error (i), the SSIM of the axial slices (ii) as well as the standard deviation of the respective metric in the white matter (iii). Given a lack of an objective ground truth, the full NLLS and full SBI results are used as ground truths for their respective subsets. A comparison between the full SBI and full NLLS for the absolute error and SSIM is shown in the gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. Gold refers to the fit results the NLLS and teal to the SBI values. The dashed black line in (ii) refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished our minimum fitting results to fall within.

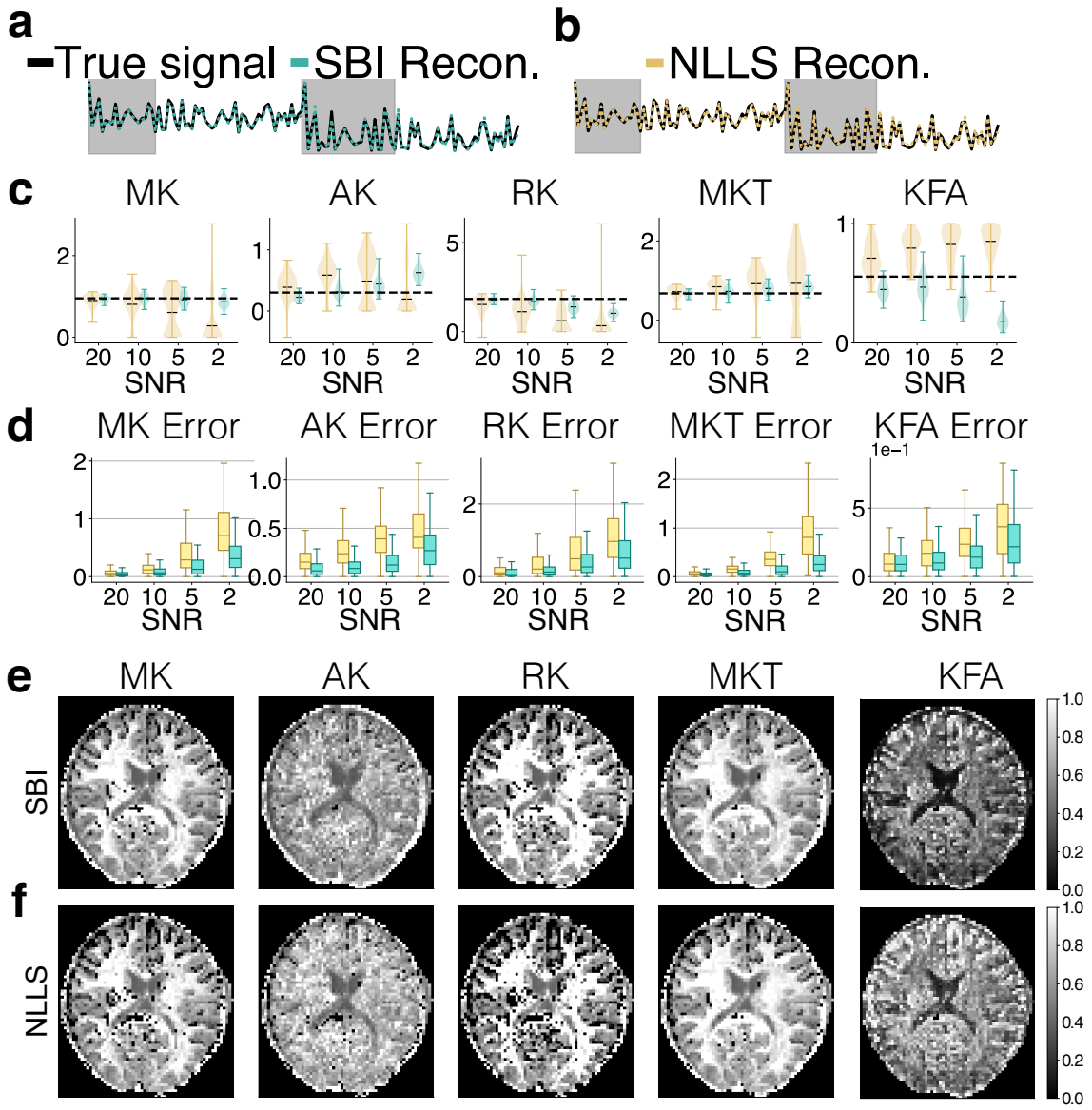


Figure S8: **Application of the SBI and NLLS DKI fitting using only the medium set of acquisitions. a)** SBI and NLLS are used to infer the parameters of an example signal and then reconstruct the signal. When the medium set of acquisitions is used there is little difference in the results (SNR=20). **b)** Using the same signal and adding a different amount of noise (SNR = 20,10,5,2) we can get an insight into how well the fitting routines are estimating the key parameters. We note that SBI (teal) outperforms NLLS (gold) across a variety of noise and parameter modalities. The black dashed line indicates the true value of the respective parameters while the solid line in the violinplot refers to the mean. **c)** To verify the performance of SBI and NLLS across a range of simulated signals, we generated 500 signals and then applied the 4 different noise strengths to each of them. The resulting parameter fits were then compared against the true parameters. We note that SBI outperforms NLLS for all modalities (apart from KFA at SNR=20). Additionally, the errors grow much more slowly with the noise than with the NLLS errors. The box plots represent the median values, with the boxes showing the IQR and the whiskers extending to 1.5 times the IQR. Gold refers to the NLLS values, and teal to the SBI values. **d)** Applying NLLS and SBI to an axial slice of a HCP individual provides a qualitative result we can use to compare the two methods after fitting. Here we have plotted the relevant kurtosis metrics.

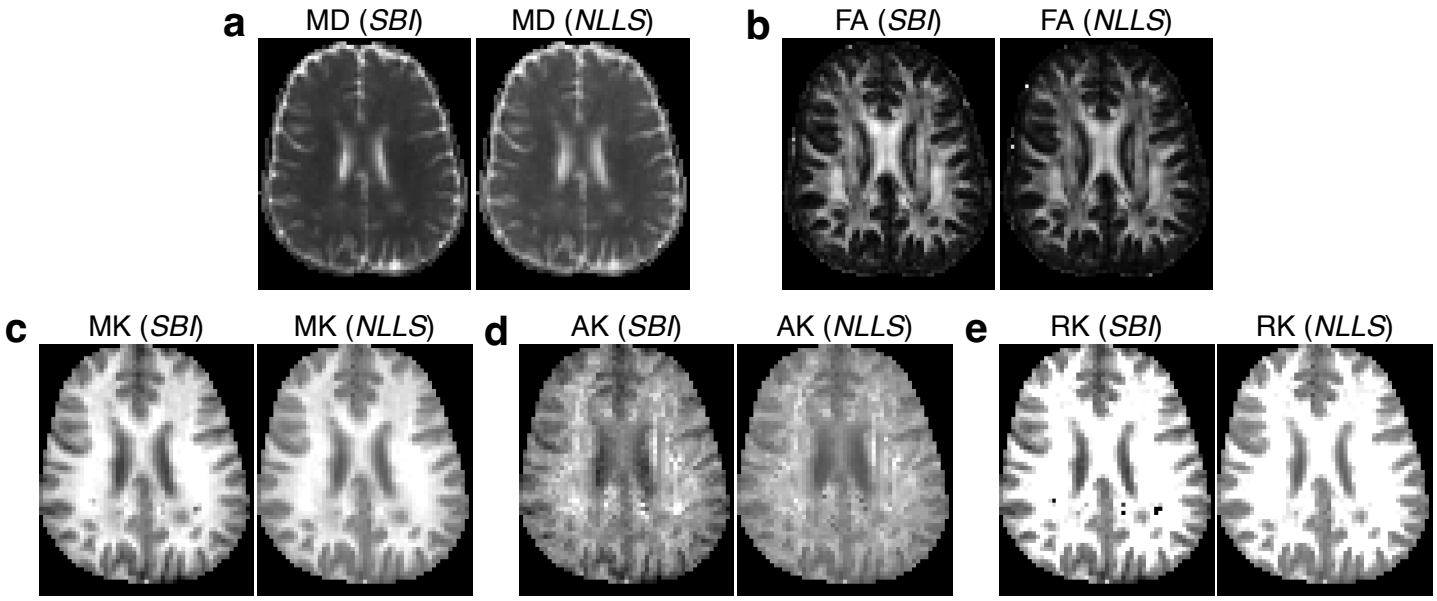


Figure S9: **Accurate DTI AND DKI metrics are achieved by both SBI and NLLS in the MS dataset[28] when the full set of acquisitions is used.** **(a–b)** Comparing the DTI metrics for both SBI and NLLS. **(c-e)** Example of the DKI metrics.

# More with less: Simulation-based inference enables accurate diffusion-weighted MRI with minimal acquisition time
## Figures

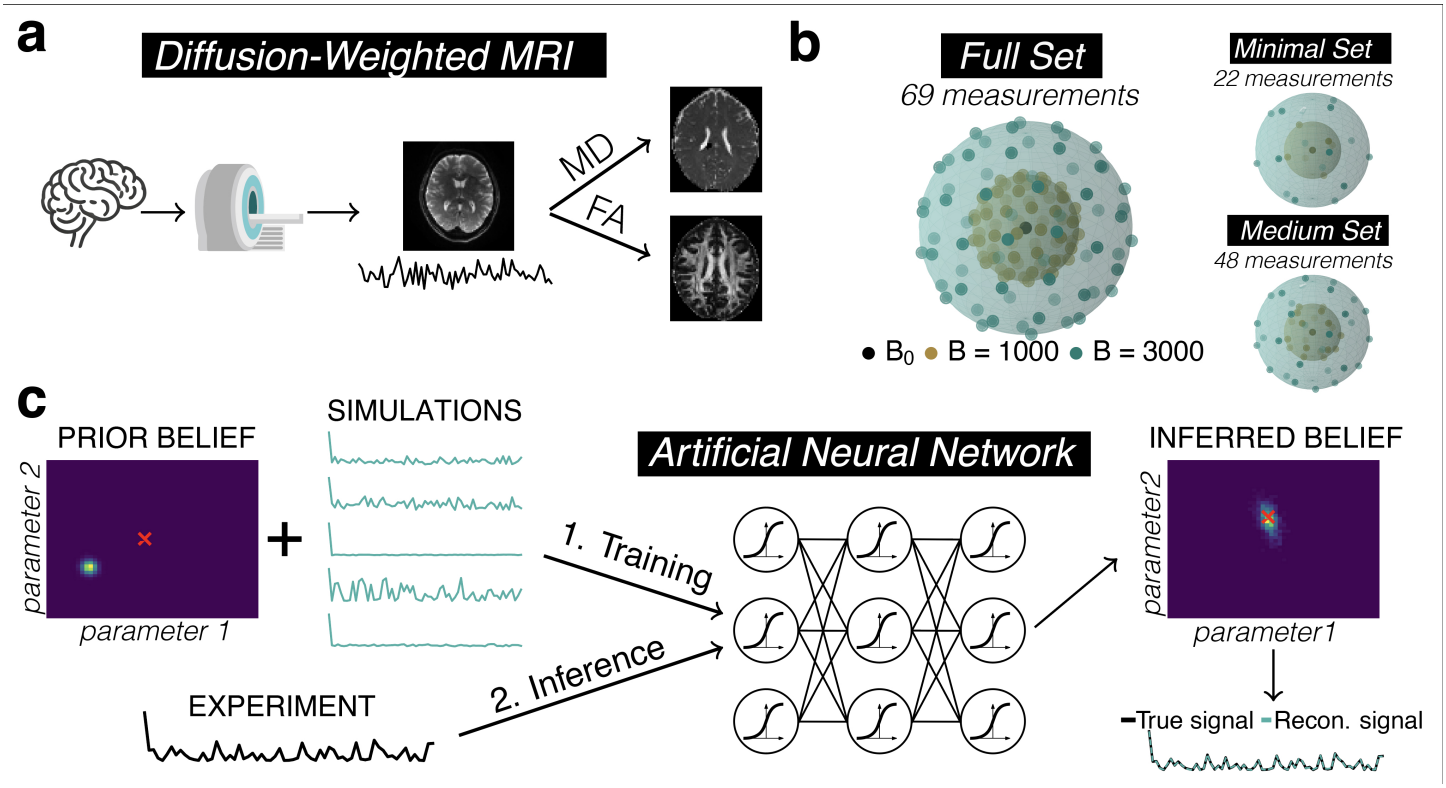


Figure 1: **Overview of the dw-MRI data processing and analysis pipeline. a)** Diffusion-weighted MRI data is processed by fitting a mathematical model to the signal, allowing for the extraction of key parameters such as mean diffusivity (MD) and fractional anisotropy (FA). These parameters provide insights into brain health and function. **b)** Both DTI and DKI frameworks rely on oversampling the diffusion tensor space using one or more b-values and many b-vectors. Left: full set of orientations for the $B = 1000$ (gold) and $B = 3000$ (teal) shells. Right: optimized b-vectors for a sub-sampled space (minimum set with 7 and 15 measurements for theoretical optimal values, medium set with 20 and 28 measurements for more realistic subspaces). **c)** In this work we utilize a simulation-based inference approach to fit DTI and DKI tensors. Here, an artificial neural network is trained with an initial prior belief of the possible model parameters, combined with a large set of simulations from that model, to provide the most likely parameters given an input. Once the neural network has converged, experimental signals are provided to the network to obtain the most likely tensor parameters, which can then be used to, e.g., reconstruct the signal or obtain MD and FA.

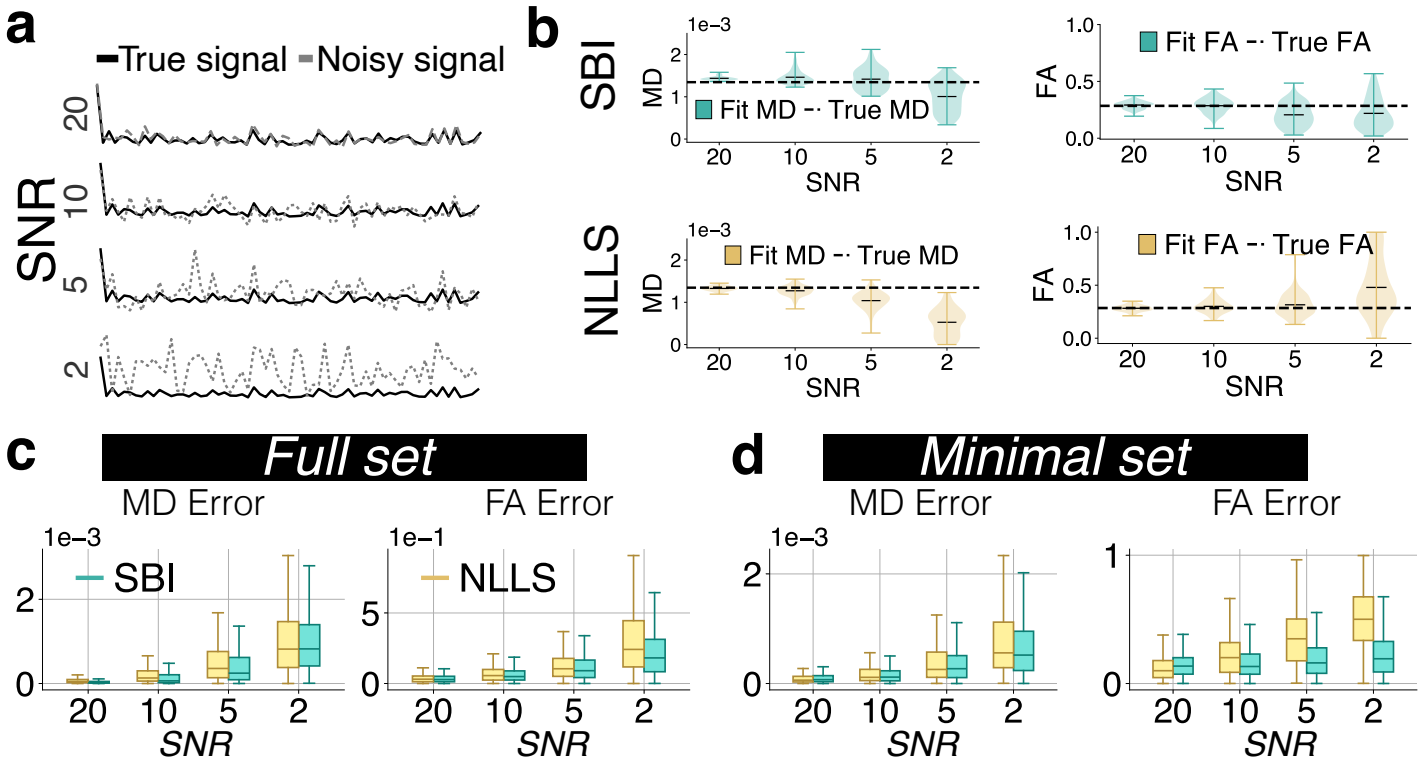


Figure 2: **SBI applied to simulated data shows more accurate performance than standard NLLS for DTI. a)** dw-MRI data is affected by varying levels of noise, which impacts signal clarity. The effect of different signal-to-noise ratios (SNRs, ranging from 2 to 20; dashed gray) is shown in comparison to the true signal (black). **b)** The accuracy of FA and MD estimates, obtained using the SBI (top) and NLLS (bottom) approach. The black dashed line indicates the true parameter values while the black solid line in violin-plot refers to mean. **c-d)** Using either the full set of measurements (c) or the minimum (d), both NLLS and SBI are fitted to 500 simulated signals and the errors in MD and FA are shown. Box plots represent median values, with boxes showing IQR and whiskers extending to 1.5 times the IQR. Gold denotes NLLS values; teal denotes SBI values.

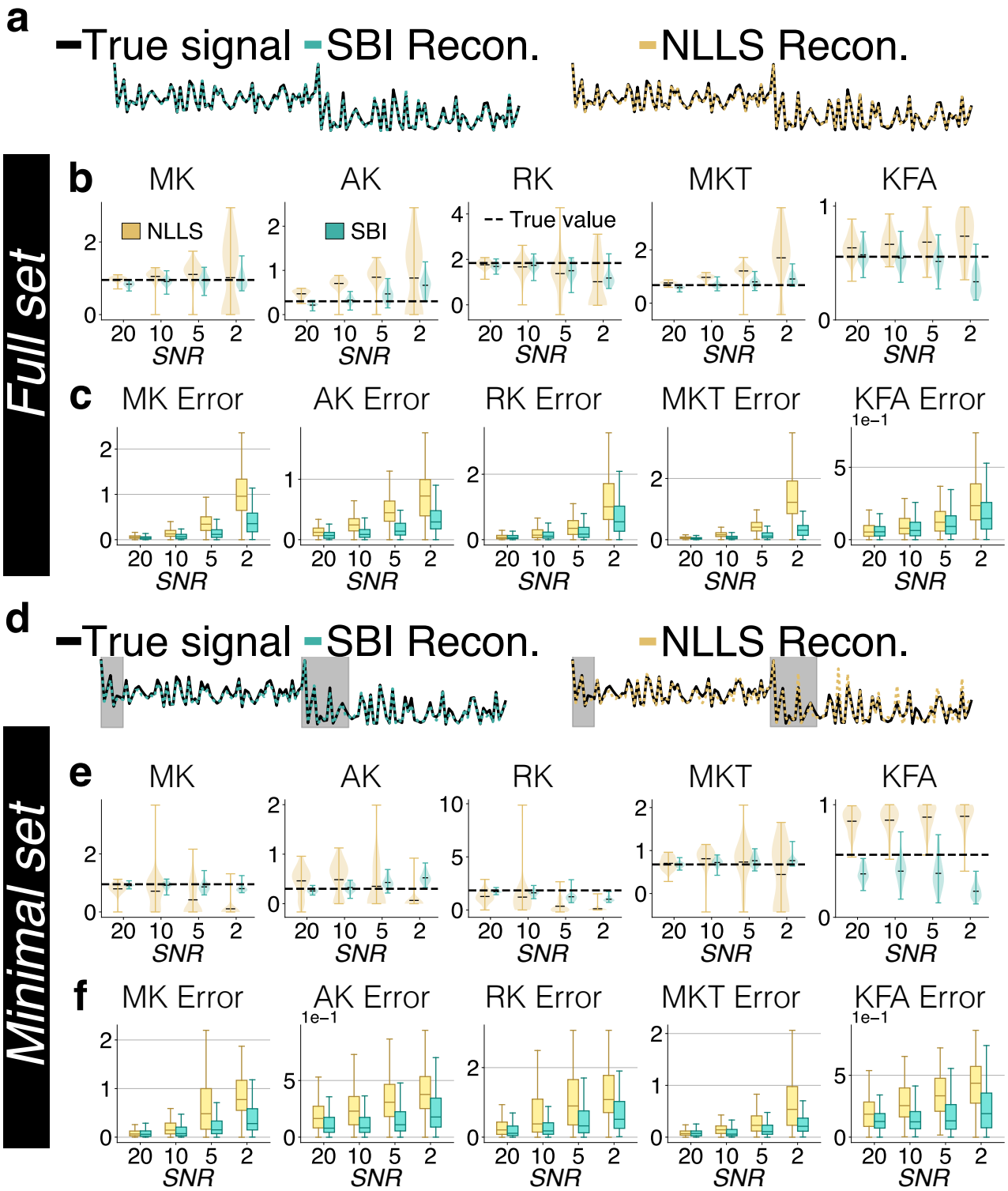


Figure 3: **SBI outperforms NLLS in the DKI framework.** **a)** SBI and NLLS are used to infer parameters and reconstruct an example signal using the full set of acquisitions (SNR of provided signal = 20). Both methods yield similar results with little difference. **b)** Applying varying noise levels (SNR=20, 10, 5, 2) to the same signal assesses the fitting routines' accuracy in estimating key parameters. SBI (teal) outperforms NLLS (gold) across different noise and parameter settings. The black dashed line indicates the true parameter values. Black solid line in violin-plot refers to mean. **c)** To evaluate performance across a range of simulated signals, 500 signals were generated and subjected to four noise levels. The resulting parameter fits were compared to the true parameters. SBI outperforms NLLS for all modalities (except KFA at SNR=20), and its errors increase more slowly with noise. Box plots represent median values, with boxes showing the IQR and whiskers extending to 1.5 times the IQR. Gold denotes NLLS values; teal denotes SBI values. **d-f)** Repeating the analysis in panels a–c with the minimum set of acquisitions, SBI continues to outperform NLLS even with reduced data.

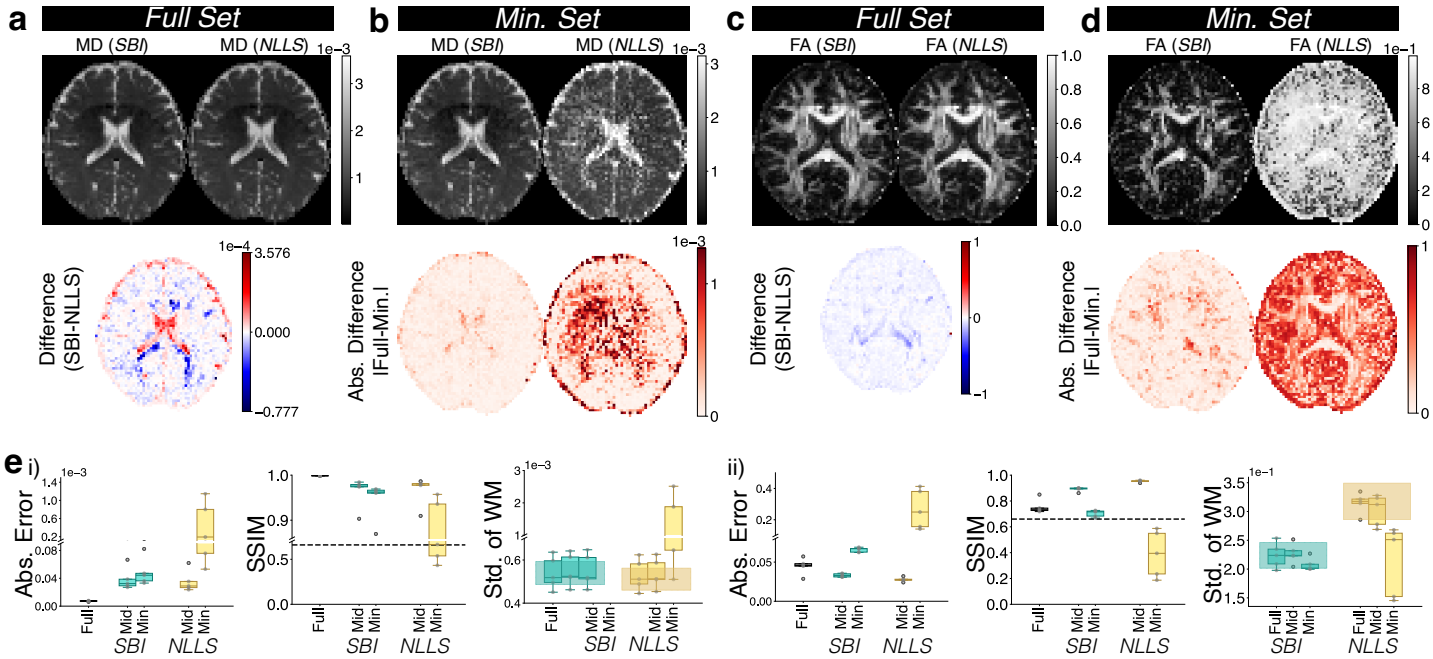


Figure 4: **Even with minimal information SBI is able to get meaningful information on real-world data where NLLS fails. a)** Comparison of MD values obtained with SBI and NLLS using the full set of acquisitions. Below we see a map of the differences between the two fits. **b)** Comparison of MD values obtained with SBI and NLLS using the minimal set. Below we see two maps that show the difference between the full and minimal fits for the SBI and NLLS implementations. **c-d)** Similar to a,b but instead of MD we show FA. **e)** Results of fitting MD and FA with SBI (one single network, teal) and NLLS (gold) using the full, medium and minimal set of acquisitions on the medial axial slice of 5 different individuals of the HCP dataset. For an objective comparison we present the absolute error of the MD and FA, the SSIM and the standard deviation of the respective metric in the white matter. Given a lack of an objective ground truth, the full NLLS and full SBI results are used for their respective subsets. A comparison between the full SBI and full NLLS is shown in the dark gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. The dashed black line in the SSIM plots refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished the standard deviation of the minimum fitting results to fall within.

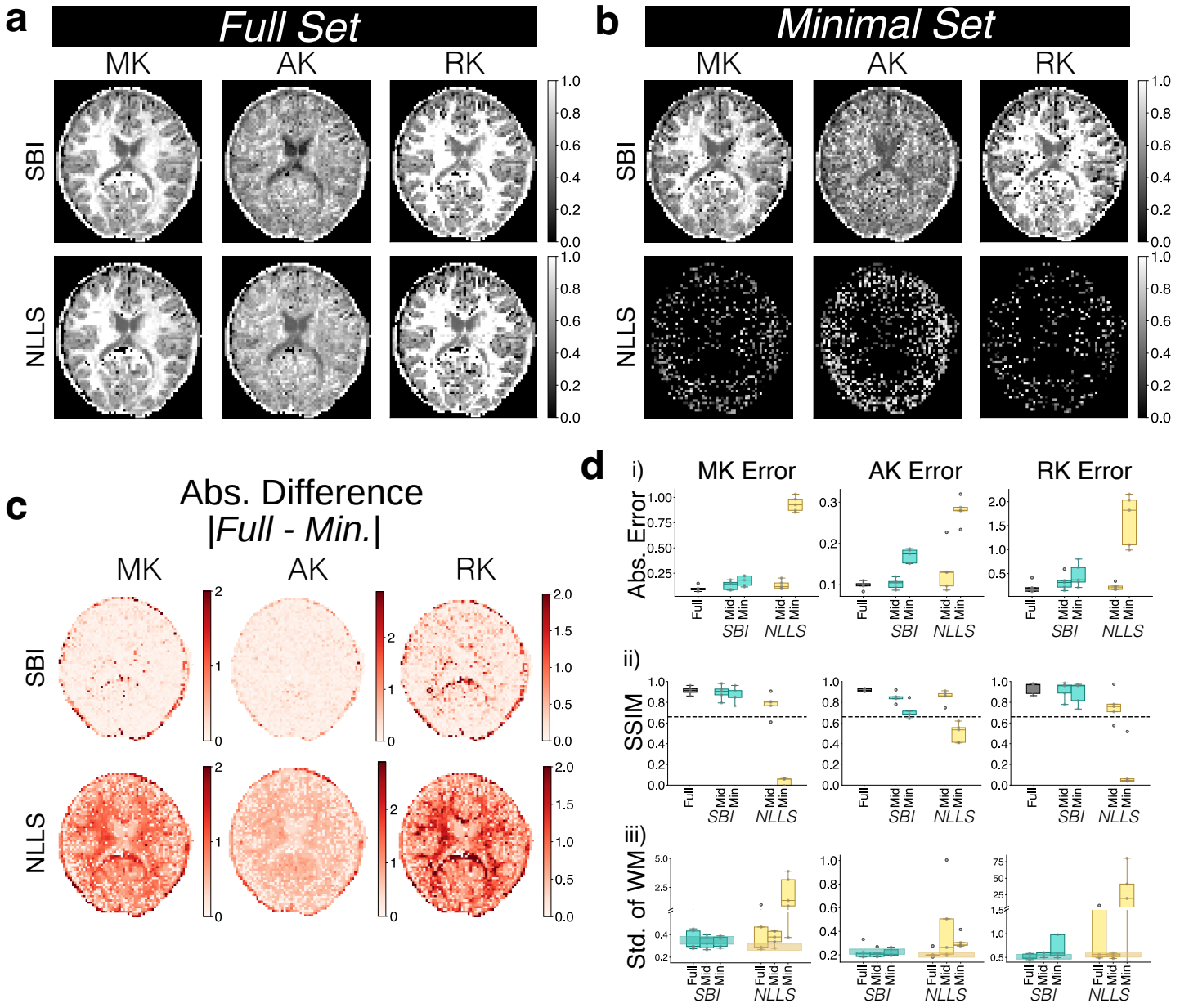


Figure 5: **SBI resolves kurtosis metrics even with minimal acquisitions. a)** Applying NLLS and SBI to the axial slice of an individual (different from the one used for the prior generation) and using the full set of acquisitions provides qualitative comparison of the fitted kurtosis tensor metrics. Note that black dots in the images indicate either large negative values or invalid numbers. **b)** Similar to panels a, but using only the minimal set of acquisitions. **c)** For each of the metrics we compute the absolute difference between the full and minimum set of SBI and NLLS. Darker red hues refer to larger differences. **d)** By fitting both SBI (teal) and NLLS (gold) using different amounts of orientations (Full set, medium and minimal) to the medial axial slice of 5 different individuals of the HCP dataset (using one single network), provides insight into the performance of the two methods. Similar to the DTI calculation, the absolute error (i), the SSIM of the axial slices (ii) as well as the standard deviation of the respective metric in the white matter (iii). Given a lack of an objective ground truth, the full NLLS (gold) and full SBI results (teal boxplots) are used as ground truths for their respective subsets. A comparison between the full SBI and full NLLS for the absolute error and SSIM is shown in the gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. Gold refers to the fit results the NLLS and teal to the SBI values. The dashed black line in (ii) refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished our minimum fitting results to fall within.

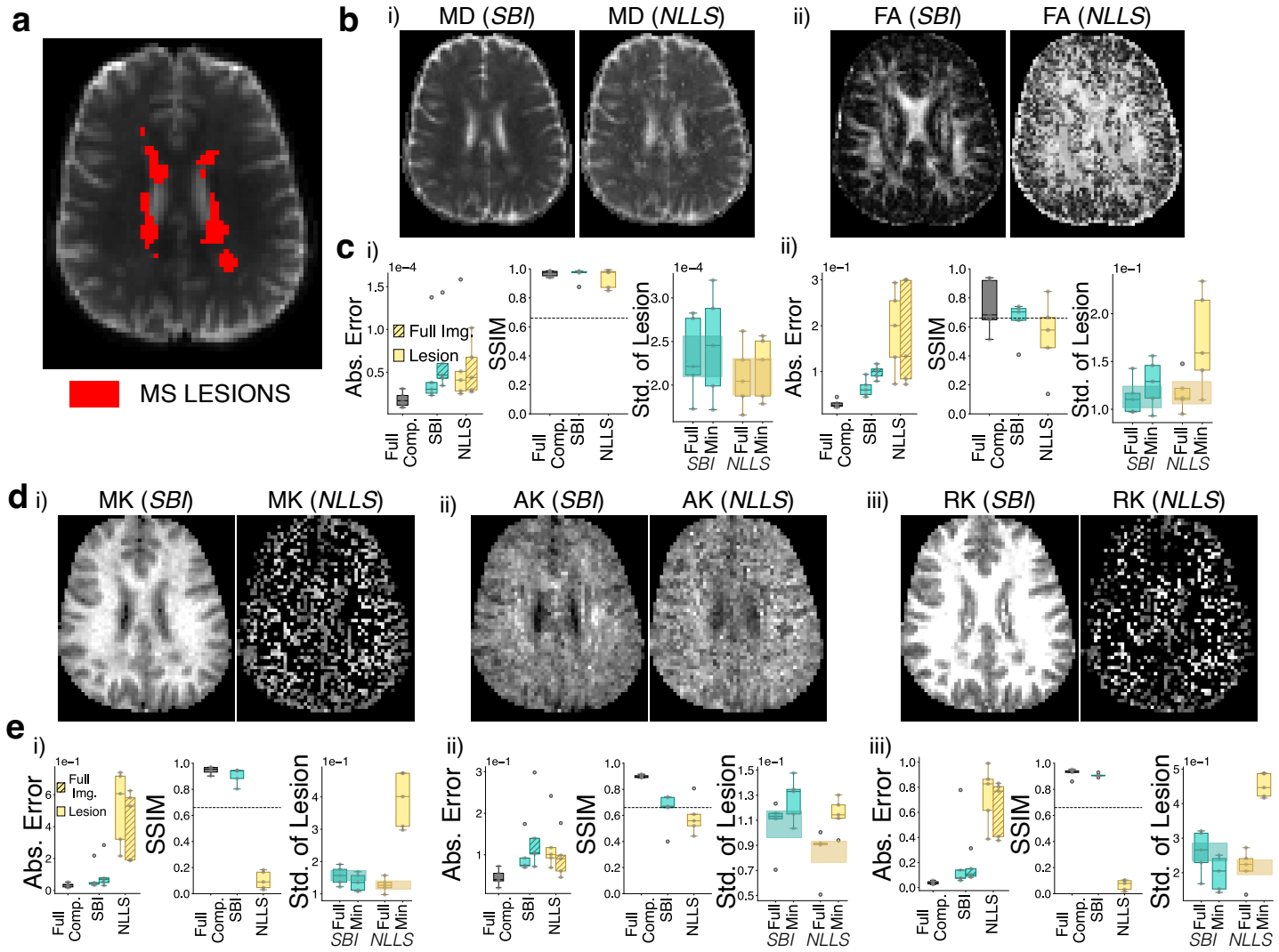


Figure 6: **The priors used to generate the accurate fittings in the HCP are also suitable to fit DTI and DKI using SBI in real-world data with MS lesions. a)** Example axial slice from the dataset of **?**. The red pixels represent locations of identified lesions. **b)** Comparison of MD and FA values obtained with SBI and NLLS using the minimal set of acquisitions. **c)** By fitting both SBI and NLLS using either the full or minimum set of acquisitions for 5 axial slices of different individuals of the MS dataset (using one single network), provides insight into the performance of SBI difference of MD (i) and FA (ii) across multiple patients. To study this performance, we calculated the absolute error of the MD and FA (both for the whole image and only the lesions - hatched), the SSIM and the standard deviation of the respective metric in the lesions. Given a lack of an objective ground truth, the full NLLS and full SBI results are used as ground truths for their respective subsets in the absolute error and SSIM metrics. A comparison between the full SBI and full NLLS is shown in the dark gray boxplot. Boxplots represent the median values while the boxes represent the IQR and the whiskers 1.5× the IQR. Gold refers to the fit results the NLLS and teal to the SBI values. The dashed black line in the SSIM plots refers to our acceptable minimum accuracy. The shaded areas in the std plots refer to the ±10% range we wished our minimum fitting results to fall within. **d)** To evaluate the performance of the DKI fitting, we show the results of SBI and NLLS for MK, AK and RK using the minimum set of acquisitions (22). **e)** Similar to c, but now evaluating the performance of MK, AK and RK across 5 axial slices from 5 different individuals.