Quantitative evaluation of biophysical models of the diffusion with in vivo data by assessment of the generalization error

Benoit Scherrer[1], Maxime Taquet[1], Mustafa Sahin[2], Sanjay P Prabhu[1], and Simon K. Warfield[1]
[1]*Radiology, Harvard Medical School, Boston Children's Hospital, Boston, MA, United States,* [2]*Neurology, Harvard Medical School, Boston Children's Hospital, Boston, MA, United States*

PURPOSE – Biophysical models of the diffusion focus on describing the MR signal formation with a model whose parameters reflect the underlying biophysical mechanisms. They are of great interest to characterize and compare tissue properties. Their practical evaluation, however, remains challenging. We propose here a novel framework for comparing biophysical models *with in vivo data*.

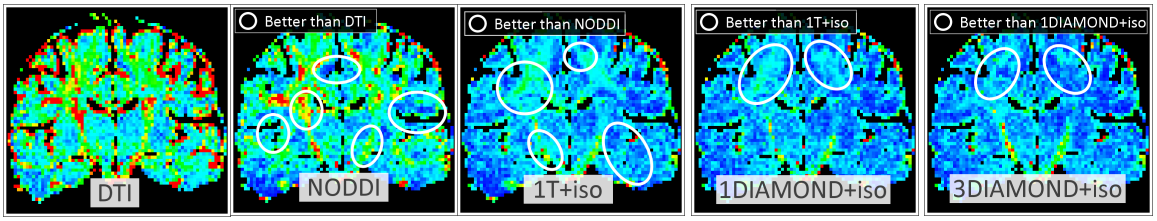
THEORY – Our hypothesis is that a biophysical model that well captures the underlying biophysical mechanisms ought to be able to accurately predict the signal for new gradient directions and strengths. This can be measured by assessing the *generalization error* of the model. Given $\boldsymbol{z} = \{z_1, ..., z_n\}$ with $z_i = (x_i, y_i)$ the set of n observed data points, in which x_i are inputs to the model (the b-values and gradient directions in diffusion images) and y_i are outputs (the signal attenuation in diffusion images), and given the model $r_{\boldsymbol{z}}(x)$ that predict the outputs y from an input x, the generalization error is defined by: $E_g = E_{z_i \overset{\text{iid}}{\sim} F}\left\{ E_{z_0 \sim F}\left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2 \middle| \boldsymbol{z} \right] \right\}$, [$z_0 = (x_0, y_0)$: new data point following the distribution F; $E_{z_0 \sim F}[\cdot]$: generalization error conditional on the observed data; $E_{z_i \overset{\text{iid}}{\sim} F}[\cdot]$: accounts for the variability of the observed data points]. Because the distribution F is unknown, the generalization error cannot be computed exactly. Leave-one-out cross validation provides an estimate of E_g with low bias but high variance. K-fold cross validation provides an estimate with lower variance but higher bias. Instead, Efron et al introduced the 632 bootstrap approach[1] which counterbalances the positive bias of the fitting error by the negative bias of the leave-one out cross-validation bootstrap estimate, by computing:

$$\hat{E}_g^{632} = (1-\tfrac{1}{n})^n \tfrac{1}{n} \sum_{i=1}^n |y_i - r_{\boldsymbol{z}}(x_i)|^2 + \left(1-(1-\tfrac{1}{n})^n\right) \tfrac{1}{n} \sum_{i=1}^n \left[\sum_{b=1}^B \delta(N_i^b) \left| y_i - r_{\boldsymbol{z}^*_{(-i)}}(x_i) \right|^2 \middle/ \sum_{b=1}^B \delta(N_i^b) \right]$$

where B is the number of boostrap iterations, N_i^b is the number of times sample *i* is used in the training set of the b[th] bootstrap replicate and $\delta(x)$ is the Dirac function. The factor $\delta(N_i^b)$ guarantees that sample *i* can be used as a testing sample in the bootstrap replicate *b*. The B632 estimate of the generalization error was shown to have low bias and low variance[1].

METHODS – We evaluated five biophysical models of the diffusion: (1) the single tensor model, (2) NODDI[2], (3) a model with one anisotropic tensor and one isotropic tensor to model free water diffusion (Diso=3x10^{-3} mm^2/s) (1T+iso), (4) a model with one DIAMOND compartment[3] and an isotropic compartment (1DIAMOND+iso) and (5) a model with up to three DIAMOND compartments and an isotropic compartment (3DIAMOND+iso). The number of fascicles was estimated at each voxel by minimizing the generalization error. We compared the models using a CUSP65 acquisition[4] (FOV=240mm, matrix-size=128x128, 68 slices, resolution=1.8x1.8x2mm^3, TE=78ms, TR=10.1s, ~12min acquisition time) which provides a large number of different b-values between 1000s/mm^2 and 3000s/mm^2 with low TE and high SNR. We computed the generalization error using B=300 bootstrap iterations.

RESULTS – We report a coronal view the generalization error maps computed for the five models. Fig.f reports the table of the corresponding average generalization errors.



	Mean Generalization Error
1T	57.72
NODDI	41.20
1T+iso	28.93
1DIAMOND+iso	26.54
3DIAMOND+iso	26.37

a b c d e f

DISCUSSION – Fig.a shows that DTI is the worst predictor of the diffusion signal, likely because it does not account for the non-monoexponential decay. NODDI (Fig. b) provides a lower generalization error in regions of crossing and close to the cortex because models the fascicle dispersion in each voxel, and accounts for freely diffusing water. However, Fig.c shows that 1T+iso better predicts the signal than NODDI. This is likely because a number of parameters are fixed in NODDI (fixed parallel diffusivity, no radial diffusivity) which is inconsistent with the imaging data. Fig.c shows that estimation of all the parameters of the 1T+iso model enables ultimately better prediction of the signal. Fig.d shows that accounting for the heterogeneity of each compartment by estimation of a DIAMOND model for each of them slightly improves the generalization error in regions of crossings. Finally, Fig.e shows that accounting for each fascicle in each voxel and accounting for the compartment heterogeneity provides the best generalization error in these regions.

CONCLUSION – We proposed a novel framework to achieve quantitative evaluation of biophysical models of the diffusion with in-vivo data. It enables identification of the model that best predicts the diffusion signal, and therefore identification of the model that best captures the underlying biophysical mechanisms. It is important to note that there is no absolute best model. This is strongly dependent on the acquisition. In future work we will investigate the optimal choice of diffusion model for various acquisition schemes and acquisition duration time.

REFERENCES - 1.Efron, B., *et al.*, Journal of the American Statistical Association, 1997. **92**(438): p. 548-560. 2.Zhang, H., *et al.*, Neuroimage, 2012. **61**(4): p. 1000-16. 3.Scherrer, B., *et al.*, Med Image Comput Comput Assist Interv., 2013. **16**: p. 518-526. 4.Scherrer, B., *et al.*, PLoS One, 2012. **7**(11).