Analysis of OCT decays

Pascal PERNOT

2018/12/03

Contents

1 Introduction

Analysis of OCT signals by a mono-exponential decay reveals two features which condition the proposed data analysis method: an heterogeneous random noise (Poisson-like; see Fig. @ref{fig:plotNoise1}), and medium-scale oscillations around the exponential decay (model inadequacy; see Fig. @ref{fig:plotMono1}). To be able to partition unambiguously the model residuals between these two components, we proceed in two steps:

1. estimation and modeling of the random noise component, to be injected in
2. the estimation of the parameters of the decay model:
    a. test of a simple monoexponential model
    b. if the latter fails, use of a modulated decay model

2 Methods

Considering a set of N measured data points $\boldsymbol{D} = \{z_i, y_i\}_{i=1}^{N}$, one considers a measurement model with heterogeneous additive noise

$$y_i = f(z_i) + \epsilon_i \tag{1}$$

where $f(.)$ is a model function to be defined, and

$$\epsilon_i \sim Norm(0, \sigma_i) \tag{2}$$

represents an heterogeneous measurement noise with a normal distribution of standard deviation σ_i.

2.1 Estimation and modeling of the random noise

A cubic smoothing spline function is used to estimate the random part of the signal. The residuals of the smoothing function are assigned to random noise ϵ.

Considering that the OCT decays result from photon counting experiments, one can expect that the noise obbeys a Poisson law. In consequence, the standard deviation of the noise is modeled by an exponential decay

$$\sigma_i = a_1 * \exp\left(-\frac{2 * z_i}{a_2}\right) \tag{3}$$

This shape enables also to account for cases of *in vivo* measurements with nearly uniform noise.

The parameters are obtained by Bayesian inference (Gelman et al. 2013) from Eqns. 2-3, with uniform priors in the range $]0, a_{max}]$, the upper value being chosen to accommodate a quasi-uniform noise model.

The posterior pdf is sampled by Markov Chain Monte Carlo in `stan` (Gelman, Lee, and Guo 2015) (see below for details of implementation), and the mean values are used to define the measurement uncertainty

$$u_{yi} = \bar{a_1} * \exp\left(-\frac{2 * z_i}{\bar{a_2}}\right) \tag{4}$$

to be used in the next steps, for which one has now a data set augmented with measurement uncertainties $\boldsymbol{u}_y$, *i.e.* $\boldsymbol{D} = \{z_i, y_i, u_{yi}\}_{i=1,N}$.

2.2 Calibration of a decay model

2.2.1 Mono-exponential decay

The mono-exponential decay curve with parameters $\boldsymbol{\vartheta} = \{a, b, l_0\}$

$$f(z; \boldsymbol{\vartheta}) = a + b * \exp\left(-\frac{2 * z}{l_0}\right) \tag{5}$$

is first fitted to the data by maximization of the posterior pdf (MAP). The likelihood is

$$\boldsymbol{y}|\boldsymbol{\vartheta} \sim \exp\left\{-\frac{1}{2}\ \chi^2_w(\boldsymbol{y};\boldsymbol{z},\boldsymbol{u}_y,\boldsymbol{\vartheta})\right\} \tag{6}$$

where the weighted chi-square function is defined as

$$\chi^2_w(\boldsymbol{y};\boldsymbol{z},\boldsymbol{u}_y,\boldsymbol{\vartheta}) = \sum_{i=1}^{N} \frac{[y_i - f(z_i;\boldsymbol{\vartheta})]^2}{u^2_{yi}} \tag{7}$$

The parameters have uniform priors on $[0,\infty[$.

2.2.1.1 Validation

The value of the reduced chi-square $\chi^2_r = \chi^2_w/(N-3)$ should be close to 1 ($\chi^2_r \in IQ_{95}$), based on the quantiles of the reduced chi-square distribution with $N-3$ degrees of freedom. Moreover, the residuals should not present serial correlation. If these conditions are not met, one has to use the more elaborate model, described below.

2.2.2 Modulated decay model

The mono-exponential decay model is improved with a z-dependent optical depth $l(z;l_0,\boldsymbol{\kappa})$

$$f(z;\boldsymbol{\vartheta},\boldsymbol{\kappa}) = a + b * \exp\left(-\frac{2*z}{l(z;l_0,\boldsymbol{\kappa})}\right) \tag{8}$$

where the shape of the optical depth is defined as

$$l(z;l_0,\boldsymbol{\kappa}) = l_0 * (1 + \delta l(z;\boldsymbol{\kappa})) \tag{9}$$

$\delta l(.)$, *the lodulation function*, is a Gaussian Process (GP) of mean 0,conditioned on M control values $\boldsymbol{\kappa} = \{\kappa_i\}_{i=1}^M$ at predefined locations $\hat{\boldsymbol{z}} = \{\hat{z}_i\}_{i=1}^M$. The mean value of the GP is used here as an interpolator between the control points, and we choose a Gaussian kernel for its smoothness properties

$$C(z,z') = \alpha^2 * \exp\left(-\frac{(z-z')^2}{\rho^2}\right) \tag{10}$$

The α and ρ parameters of the GP are fixed *a priori*.

Considering the set of M control values $\boldsymbol{\kappa}$ for the OD modulation at locations $\hat{\boldsymbol{z}}$, δl can be obtained at any depth as the mean value of the GP:

$$\delta l(z;\boldsymbol{\kappa}) = \boldsymbol{\Omega}^T * \boldsymbol{K}^{-1} * \boldsymbol{\kappa} \tag{11}$$

where $\boldsymbol{K}$ is a $M \times M$ covariance matrix with elements $K_{ij} = C(\hat{z}_i,\hat{z}_j)$ and $\boldsymbol{\Omega}$ is a M-vector with

elements $\Omega_i = C(\hat{z}_i, z)$.

The control points positions, $\hat{\boldsymbol{z}}$, are chosen *a priori* on a regular grid spanning the experimental depth range. As one does not expect short scale modulations, a small number of points is used, typically $M \simeq 10$. Similarly, for the smoothness of the interpolation, one picks the correlation length of the kernel *a priori*, at a value large enough to avoid undue oscillations between the control points and small enough to avoid excessive rigidity of the model. In the present configuration, a good compromise has been found to be $\rho = 1/M^{th}$ of the total depth range. In the same spirit, the variance parameter of the GP is taken as a small fraction of the standard deviation of the control values $\alpha = 0.1 * sd(\boldsymbol{\kappa})$. This choice of α and ρ has been found to provide a well behaved interpolator for test simulated signals. Besides, small changes around these values do not affect significantly the mean prediction of the GP (Eq. 8).

2.2.2.1 Likelihood function

To compensate for defaults in the estimation of the noise, a parameter σ is introduced as a multiplicative factor of $\boldsymbol{u}(y)$. The unknown parameters are therefore $\boldsymbol{\vartheta}$, $\boldsymbol{\kappa}$, and σ. The likelihood function is a multivariate Normal distribution

$$\boldsymbol{y}|\boldsymbol{\vartheta}, \boldsymbol{\kappa}, \sigma \sim \exp\left\{-\frac{1}{2}\chi^2_w(\boldsymbol{y}; \boldsymbol{z}, \boldsymbol{u}_y, \boldsymbol{\vartheta}, \boldsymbol{\kappa}, \sigma)\right\} \tag{12}$$

where

$$\chi^2_w(\boldsymbol{y}; \boldsymbol{z}, \boldsymbol{u}_y, \boldsymbol{\vartheta}, \boldsymbol{\kappa}, \sigma) = \sum_{i=1}^{N} \frac{[y_i - f(z_i; \boldsymbol{\vartheta}, \boldsymbol{\kappa})]^2}{[\sigma * u_{yi}]^2} \tag{13}$$

2.2.2.2 Prior pdfs

$\boldsymbol{\kappa}$. The definition of the modulation function is a source of indetermination between l_0 and $\boldsymbol{\kappa}$; for instance, setting all values of $\boldsymbol{\kappa}$ to 1 would be exactly compensated by halving l_0. One therefore constrains $\boldsymbol{\kappa}$ to be close to zero with a Bayesian Lasso-type prior (Park and Casella 2008), in a version based on a hierarchical prior adapted from Ref.(Mallick and Yi 2014):

$$\begin{aligned} \kappa_i|u_i &\sim Normal(0, s_i); i = 1, M \\ s_i|\lambda &\sim Gamma(2, \lambda); i = 1, M \\ \lambda &\sim Gamma(2, \lambda_r) \end{aligned} \tag{14}$$

whete λ_r is chosen *a priori*. I defines the scale of expected deviations from zero of $\boldsymbol{\kappa}$ (typically $\lambda_r = 0.1$).

$\boldsymbol{\vartheta}$. The prior on $\boldsymbol{\vartheta}$ is a multivariate normal distribution $\pi(\boldsymbol{\vartheta}) = N(\hat{\boldsymbol{\vartheta}}_1, \boldsymbol{\Sigma}_\vartheta)$ centered on the best estimate from the mono-exponential fit, $\hat{\boldsymbol{\vartheta}}_1$, with a covariance matrix $\boldsymbol{\Sigma}_\vartheta$. Because the present model is used when the mono-exponential decay is inadequate, one cannot rely directly on the covariance matrix extracted from its calibration, $\boldsymbol{\Sigma}_{\vartheta_1}$. Instead, a covariance matrix is estimated to cover the range ov variations of the mono-exponential residuals (Pernot 2017), following two approaches:

1. the covariance matrix is built from the correlation matrix $\boldsymbol{C}_{\vartheta_1}$ issued from the monoexponential fit, and a vector of standard deviations specified from relative uncertainties on the parameters:

$$\boldsymbol{\Sigma}_{\vartheta} = \boldsymbol{I}(\boldsymbol{u}_{\vartheta}) * \boldsymbol{C}_{\vartheta_1} * \boldsymbol{I}(\boldsymbol{u}_{\vartheta}) \tag{15}$$

where $\boldsymbol{I}(\boldsymbol{u}_{\vartheta})$ is a diagonal matrix with elements $\boldsymbol{u}_{\vartheta} = r * \hat{\boldsymbol{\vartheta}}_1$. The uncertainty factor r is typically chosen as a small percentage, *e.g.*, $r = 0.05$.

2. a *diagonal* covariance matrix $\boldsymbol{\Sigma}_{\vartheta} = \boldsymbol{I}(\boldsymbol{u}_{\vartheta}^2)$ is built by a moments-matching procedure. The standard deviations $\boldsymbol{u}_{\vartheta}$ are optimized to match two criteria:

   a. S_1, the 2-sigma prediction uncertainty of the mono-exponential model has to match Q_{95}, the 95^{th} quantile of the absolute errors of the mono-exponential model (all statistics are weighted by $\boldsymbol{u}_y$) (Pernot and Savin 2018);

   b. the standard deviation of the prediction uncertainty has to be as small as possible.

   The first criterion ensures that the mean prediction uncertainty of the mono-exponential model is in agreement with the amplitude of the model's residuals (Pernot 2017). This criterion can typically be matched by an infinity of solutions, and the second one selects those parameters which provides the 'flatest' prediction band. The optimization is done by sampling the posterior pdf of the parameters $\boldsymbol{u}_{\vartheta}$ with uniform priors on $[0, \infty[$, and a likelihood function

$$S_1, S_2 | \boldsymbol{u}_{\vartheta} \sim \exp\left\{ -\frac{1}{2} \frac{(S_1 - Q_{95})^2 + S_2^2}{\varepsilon^2} \right\} \tag{16}$$

   where $S_1 = 1.96 * \sqrt{< \boldsymbol{u}_p^2 / \boldsymbol{u}_y^2 >}$, $\boldsymbol{u}_p$ is the prediction uncertainty of the mono-exponential model estimated by linear uncertainty propagation (BIPM et al. 2008), Q_{95} is the the 95^{th} quantile of the absolute weighted residuals $|\{\boldsymbol{y} - f(\boldsymbol{z}; \boldsymbol{\vartheta})\} / \boldsymbol{u}_y|$, $S_2 = \mathrm{sd}(\boldsymbol{u}_p / \boldsymbol{u}_y)$ and ε is a predefined precision factor ($\varepsilon = 10^{-3}$).

σ. The prior on σ is a normal distribution, centered on 1, with standard deviation 0.1.

The parameters to be sampled are therefore $\boldsymbol{\vartheta}$, $\boldsymbol{\kappa}$, λ and σ.

2.2.2.3 Validation

The quality of the fit can be estimated by inspection of the residuals which should not present serial correlations and should conform with the random experimental noise.

The value of the reduced chi-square $\chi_r^2 = \chi_w^2 / (N - \nu)$, where ν is the number of effective free parameters, should be close to 1. One has $\nu = 5 + \hat{M}$, where $0 \leq \hat{M} \leq M$ is the number of control values significantly different from zero, *i.e.* those for which $0 \notin IQ_{90}(\kappa_i)$.

Posterior predictive samples are also generated and plotted with the reference data to confirm the quality of the fit.

3 Implementation

Algorithm

1. Noise estimation
    i. Perform a smoothing of data and extract residuals $\boldsymbol{R}$
    ii. Fit residuals by a normal distribution with exponentially decreasing amplitude
2. Mono-exponential fit
3. Modulated exponential fit

Estimation of the random errors is done with the `smooth.spline` from `R` (R Core Team 2017); A satisfying degree of smoothing for all examples considered here was obtained by setting the smoothing `df` parameter to 15.

The Bayesian models are implemented in `stan` (Gelman, Lee, and Guo 2015), using the `rstan` interface package (Stan Development Team 2016) for `R` (R Core Team 2017). `Stan` is a very flexible and efficient probabilistic programming language to implement Bayesian statistical models. The No-U-Turn sampler (Hoffman and Gelman 2014) was used for this study.

The main outputs of the stan codes are samples of the posterior pdf of the parameters, from which statistics and plots can be generated in R. Convergence of the sampling is assessed by examining the traces of parameters samples and the 'split Rhat' statistics provided by rstan. In the present application, the Markov chains converge rapidly, and all models are run with four parallel Markov Chains of 1500 iterations each, 1000 of which are used as warm-up for the No-U-Turn sampler and dispatched. The convergence criteria and parameters statistics are therefore estimated on a sample of 2000 points.

4 Example(s)

4.1 OCT signal

4.1.1 Estimate noise

A fit of the residuals of smoothing splines provides the parameters for a model of the random noise.

```
##
##  Noise fit parameters:
## a_1 :  8.53
## a_2 :  224
```

The smoothing curve, residuals and noise model can be seen in Fig. 1. The noise presents a notable exponential decay.

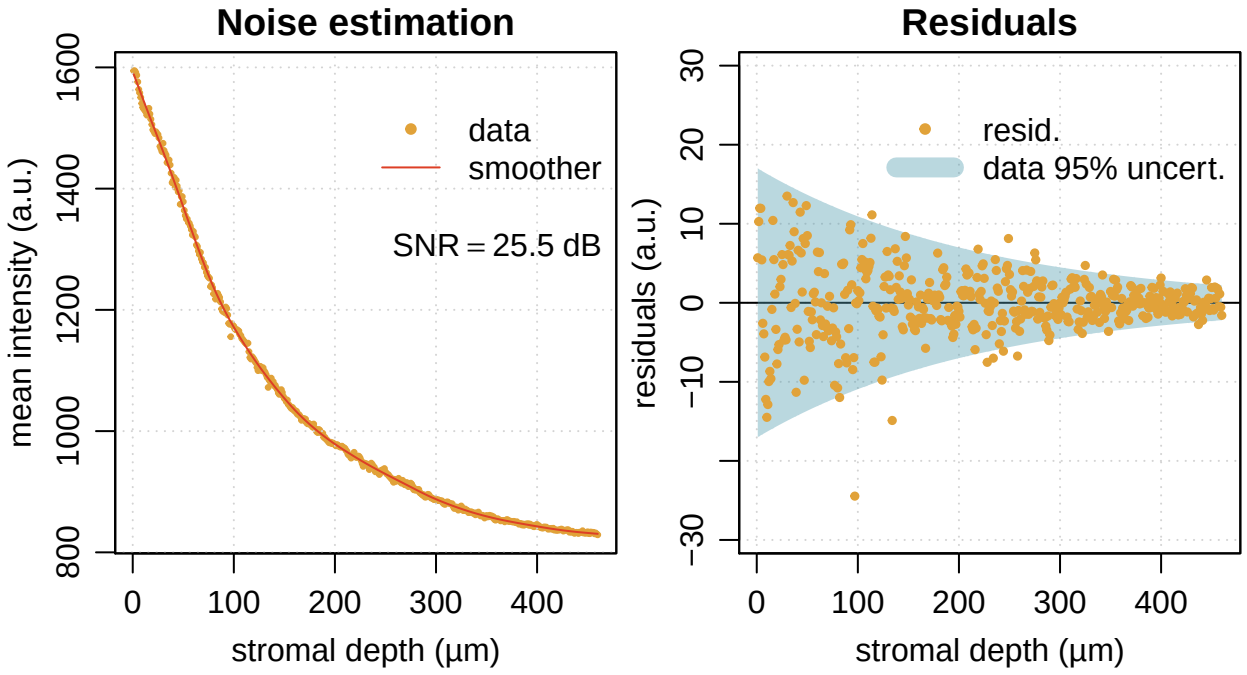

Figure 1: Splines smoothing and noise estimation

4.1.2 Mono-exponential fit

A mono-exponential fit is attempted using the noise model.

The Birge ratio analysis rejects this model

```
##
##  MonoExp decay parameters:
## b_1 :  807
## b_2 :  808
## b_3 :  259
##
##
## ndf         : 457
## br          : 2.5
## CI95(br)    : 0.87-1.1
## !!! WARNING: br out of interval !!!
```

It can be seen in Fig. 2 that the residuals present notable serial correlation.

4.1.3 Modulated-exponential fit

Considering that the mono-exponential is deemed inadequate, the model with mean-depts modulation is fit to the data. The default parameters are used, notably with a number of control points $M = 10$.

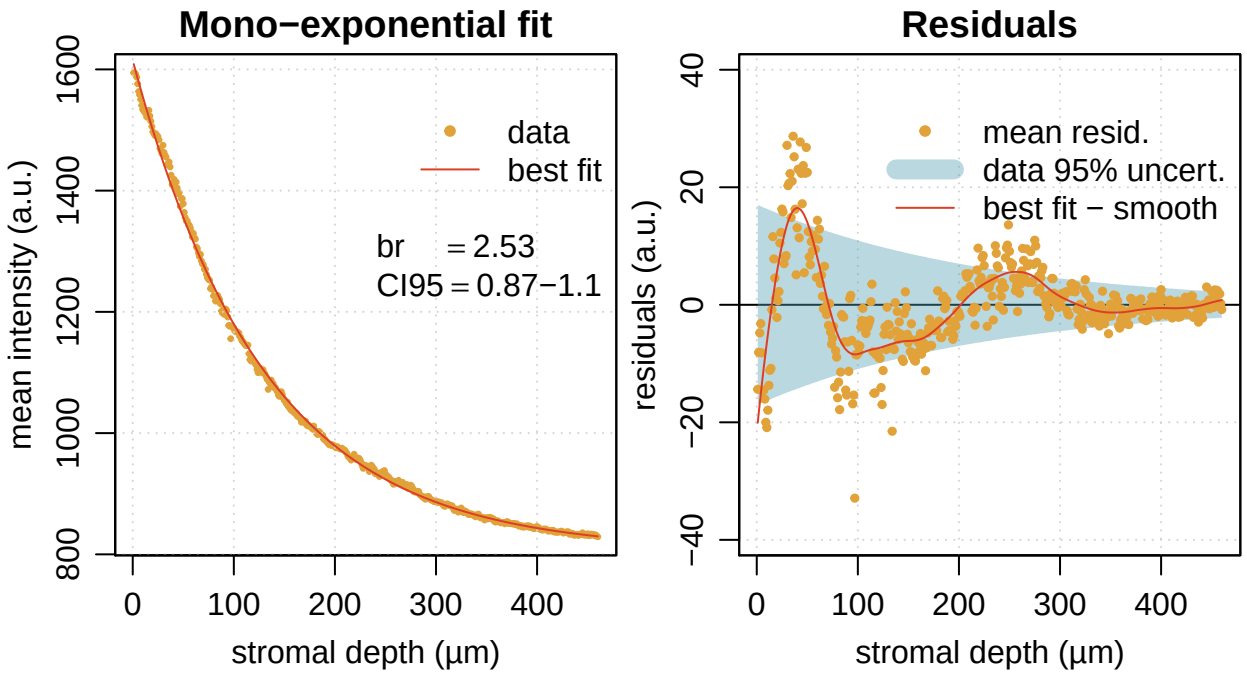

Figure 2: Mono-exponential fit and residuals.

4.1.3.1 Prior sampling

The moments-matching prior is used for the exponential parameters, and one can check the dispersion of the exponential curves generated from this prior (prior predictive sampling) in Fig. 3. Prior probability intervals for the control values are also shown in this figure (right panel).

4.1.3.2 Posterior sampling

The priors pdf seems correct, and one can now draw samples from the posterior pdf.

The Birge ratio analysis validates this model.

```
## Active pts.: 4 / 10
## ndf        : 451
## br         : 1
## CI95(br)   : 0.87-1.1
##
```

As can be seen in Fig. 4, the residuals have been notably improved from the mono-exponential model.

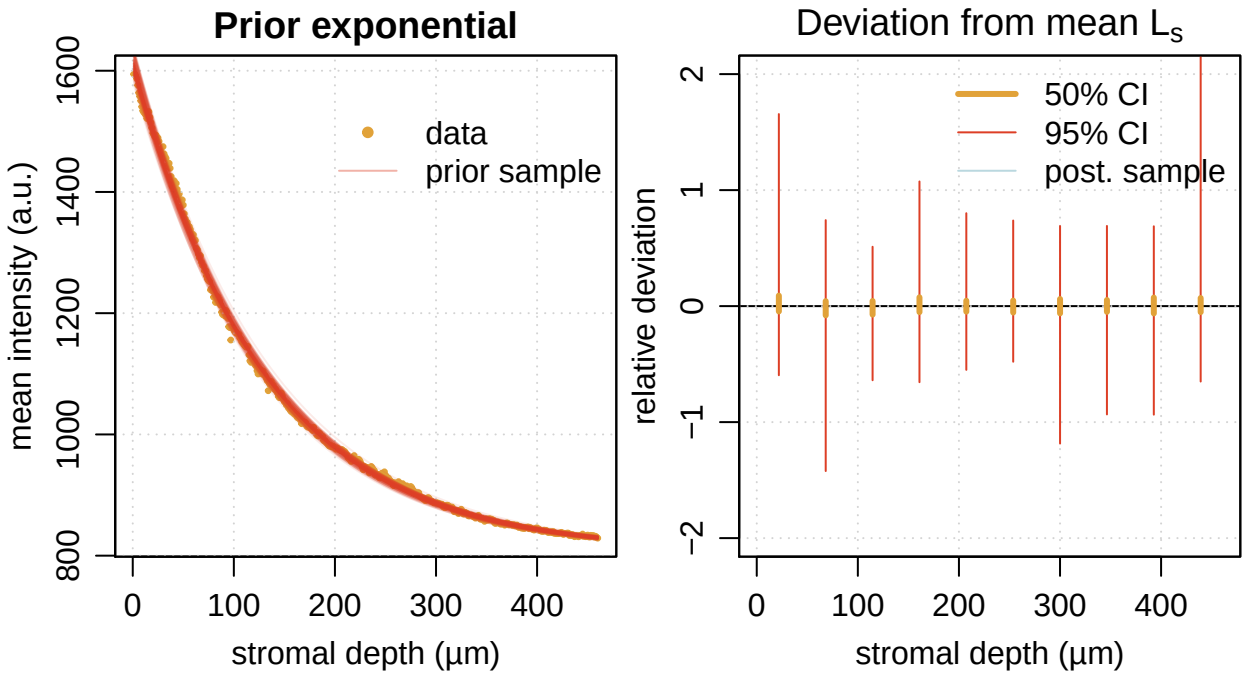

Figure 3: Samples from prior pdf.

4.2 In-vivo signal

4.2.1 Estimate noise

A fit of the residuals of smoothing splines provides the parameters for a model of the random noise.

```
##
##  Noise fit parameters:
## a_1 :  0.0092
## a_2 :  10000
```

The smoothing curve, residuals and noise model can be seen in Fig. 5. In thsi example, the noise is homogeneous, and the exponential model tends to a constant.

4.2.2 Mono-exponential fit

A mono-exponential fit is attempted using the noise model.

The Birge ratio analysis rejects this model

```
##
##  MonoExp decay parameters:
## b_1 :  0.434
## b_2 :  0.265
## b_3 :  279
```

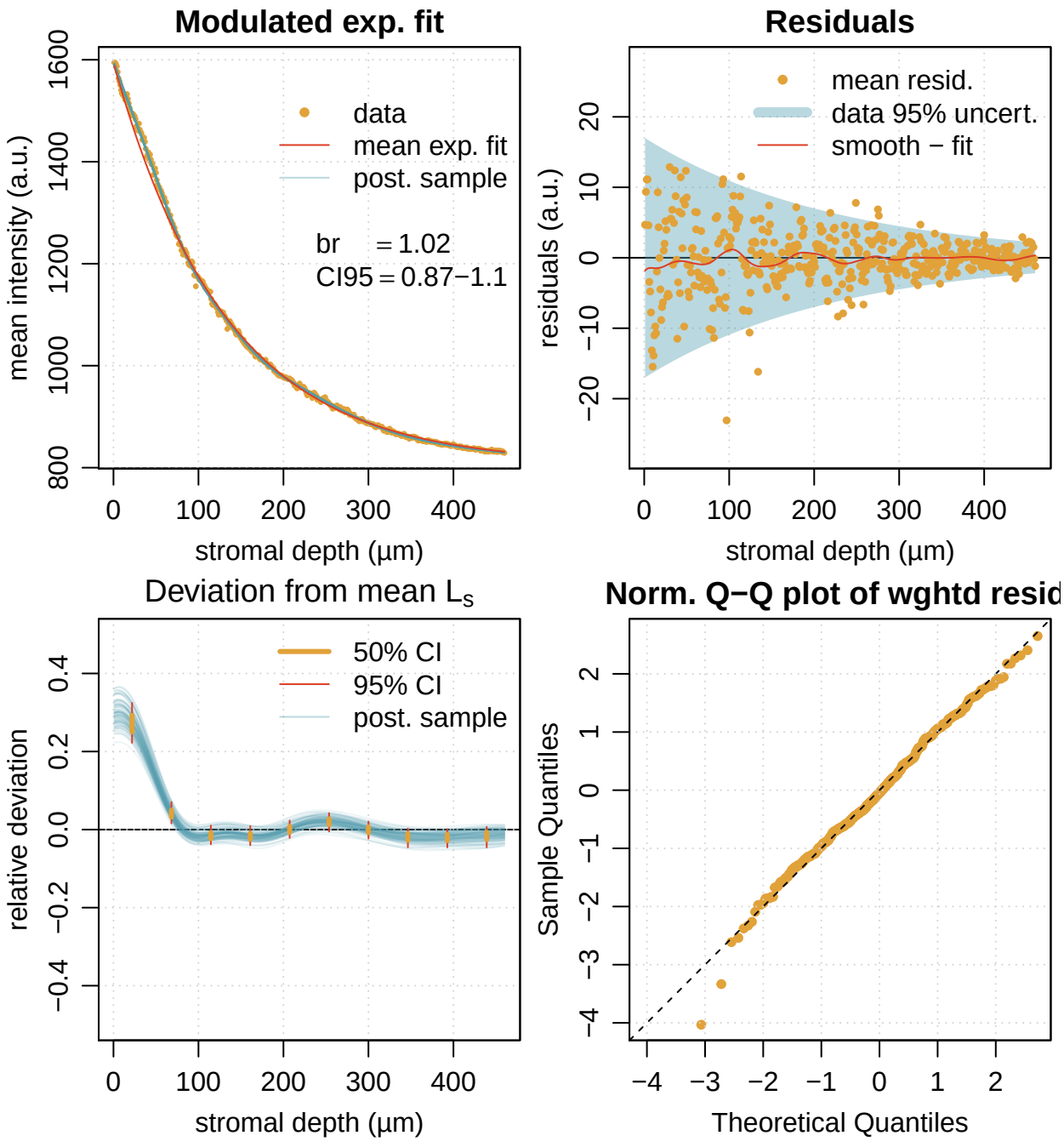

Figure 4: Modulated-exponential fit, residuals, Modulation function and Normal QQ-plot of residuals.

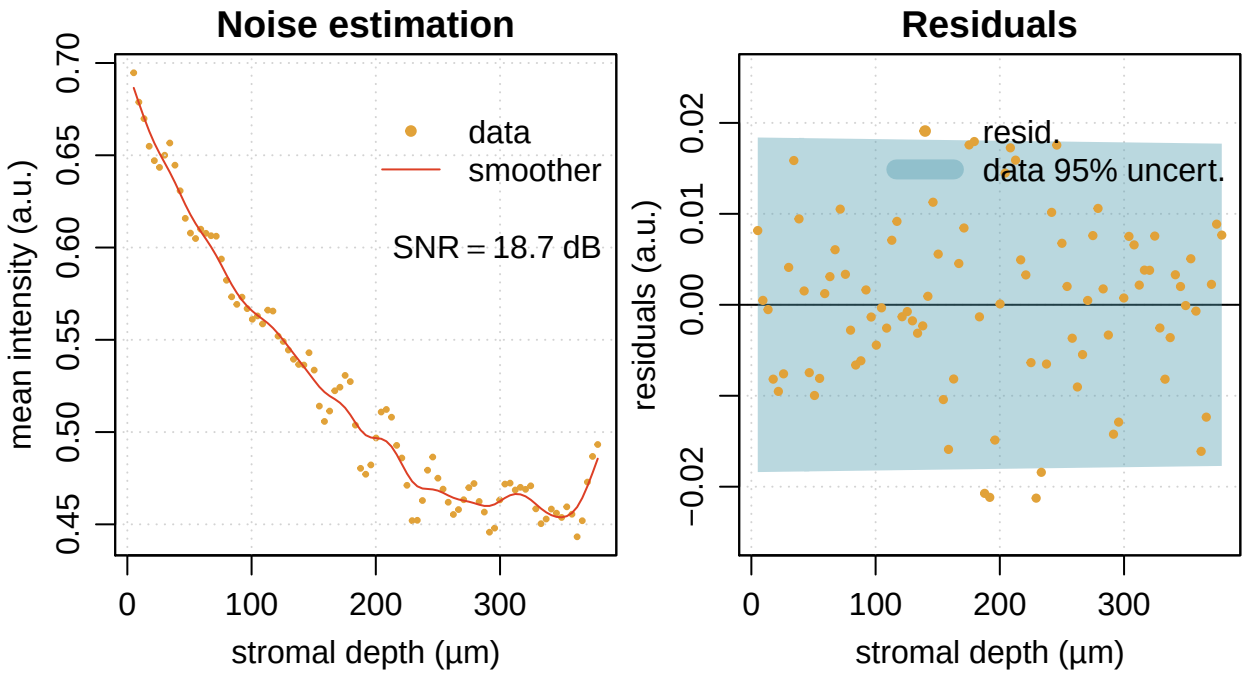

Figure 5: Splines smoothing and noise estimation

```
##
##
## ndf         : 88
## br          : 2.1
## CI95(br)    : 0.73-1.3
## !!! WARNING: br out of interval !!!
```

It can be seen in Fig. 6 that the residuals present notable serial correlation.

4.2.3 Modulated-exponential fit

Considering that the mono-exponential is deemed inadequate, the model with mean-depts modulation is fit to the data. The default parameters are used, notably with a number of control points $M = 10$.

4.2.3.1 Prior sampling

The moments-matching prior is used for the exponential parameters, and one can check the dispersion of the exponential curves generated from this prior (prior predictive sampling) in Fig. 7. Prior probability intervals for the control values are also shown in this figure (right panel).

4.2.3.2 Posterior sampling

The priors pdf seems correct, and one can now draw samples from the posterior pdf.

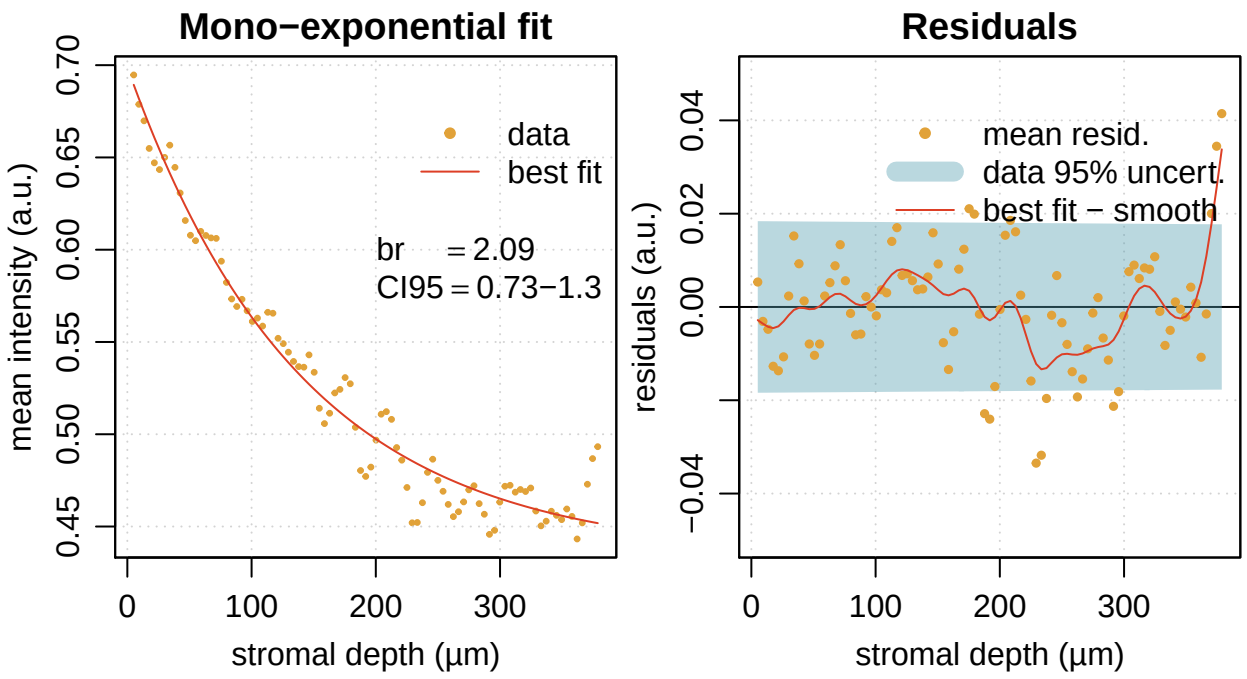

Figure 6: Mono-exponential fit and residuals.

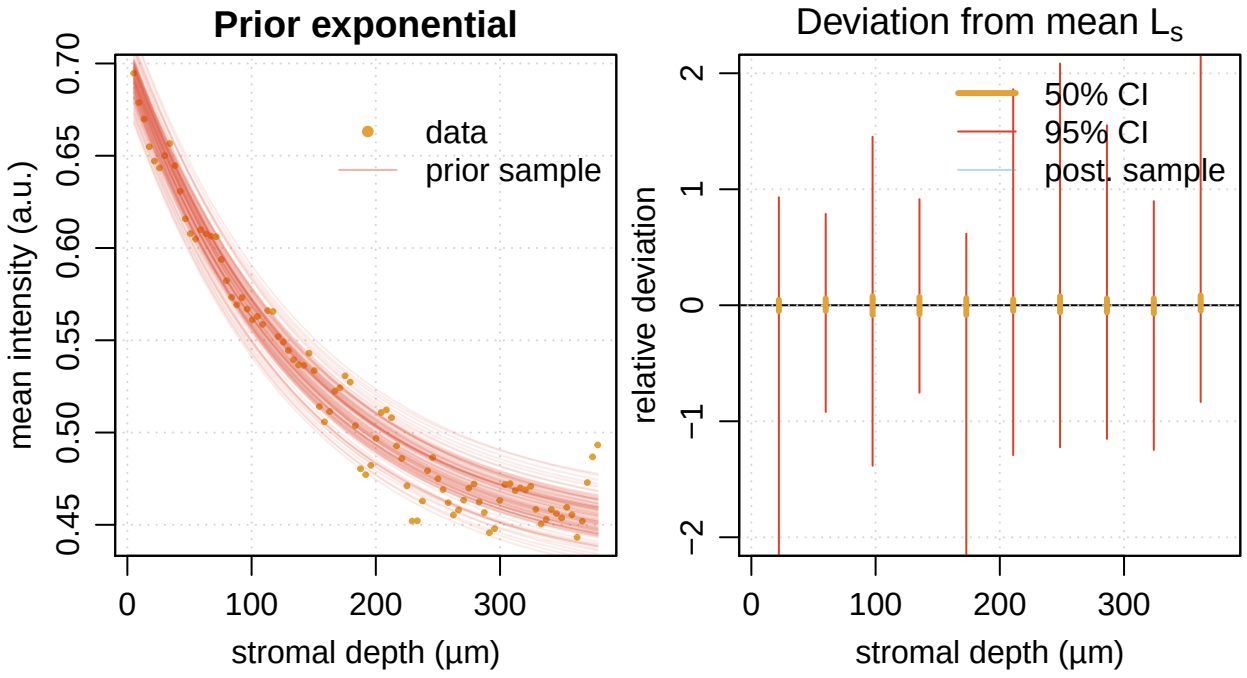

Figure 7: Samples from prior pdf.

The Birge ratio analysis invalidates this model.

```
## Active pts.: 2 / 10
## ndf        : 84
## br         : 1.3
## CI95(br)   : 0.72-1.3
##
```

As can be seen in Fig. 8, the residuals have not been notably improved from the mono-exponential model, and the present design does not enable to fit a sharp rise in the OCT signal at a depth of about 400 μm. An option is to increase the number of control points.

4.2.4 Add control points

The number of control points has been raised to M=15 on a grid covering the extremities of the depts range, and the posterior distribution has been resampled.

```
##
##  ExpGP parameters:
## Inference for Stan model: modFitExpGP.
## 4 chains, each with iter=200; warmup=100; thin=1;
## post-warmup draws per chain=100, total post-warmup draws=400.
##
##            mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## theta[1]   0.43    0.00 0.00   0.43   0.43   0.43   0.44   0.44   302 0.99
## theta[2]   0.27    0.00 0.00   0.26   0.26   0.26   0.27   0.27   324 0.99
## theta[3] 280.63    0.35 7.01 266.89 275.91 280.62 285.05 294.93   395 1.00
## yGP[1]    -0.02    0.01 0.13  -0.31  -0.08  -0.01   0.04   0.28   381 1.00
## yGP[2]    -0.02    0.00 0.07  -0.16  -0.06  -0.01   0.02   0.11   510 1.00
## yGP[3]     0.04    0.00 0.05  -0.06   0.00   0.03   0.06   0.16   399 1.00
## yGP[4]     0.00    0.00 0.04  -0.07  -0.02   0.01   0.03   0.09   281 1.00
## yGP[5]     0.05    0.00 0.05  -0.02   0.02   0.05   0.09   0.16   277 1.01
## yGP[6]     0.06    0.00 0.05  -0.02   0.02   0.06   0.09   0.17   283 0.99
## yGP[7]     0.02    0.00 0.04  -0.06   0.00   0.02   0.04   0.10   404 1.00
## yGP[8]     0.01    0.00 0.04  -0.07  -0.01   0.01   0.03   0.09   525 1.00
## yGP[9]    -0.03    0.00 0.04  -0.13  -0.06  -0.03   0.00   0.04   312 1.00
## yGP[10]   -0.12    0.00 0.06  -0.23  -0.16  -0.12  -0.08   0.00   312 1.00
## yGP[11]   -0.07    0.00 0.06  -0.19  -0.11  -0.06  -0.02   0.03   234 1.00
## yGP[12]   -0.03    0.00 0.05  -0.15  -0.06  -0.03   0.00   0.07   350 0.99
## yGP[13]    0.08    0.00 0.07  -0.03   0.03   0.07   0.12   0.23   389 1.00
## yGP[14]   -0.05    0.00 0.07  -0.21  -0.08  -0.04   0.00   0.05   450 1.00
## yGP[15]    0.66    0.01 0.15   0.36   0.56   0.66   0.76   0.96   221 1.00
## lambda    18.90    0.50 6.14   9.27  14.62  18.43  22.55  35.50   151 1.02
## sigma      1.12    0.00 0.06   0.99   1.08   1.12   1.16   1.25   371 0.99
```

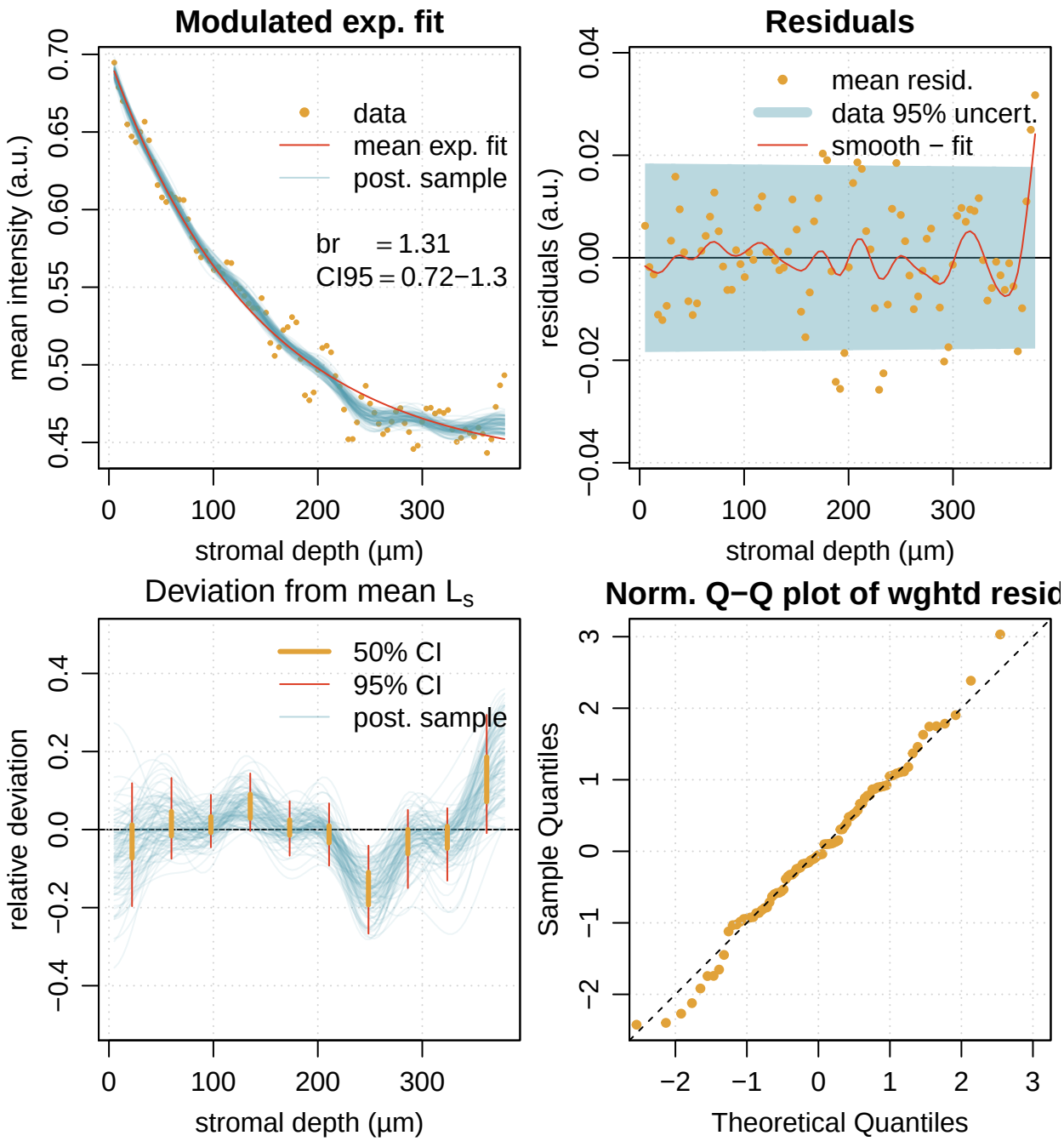

Figure 8: Modulated-exponential fit, residuals, Modulation function and Normal QQ-plot of residuals.

```
## br         1.45    0.01 0.17   1.15   1.34   1.43   1.54   1.85   364 1.00
##
## Samples were drawn using NUTS(diag_e) at Tue Dec  4 15:08:24 2018.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
##
##
## Active pts.: 2 / 15
## ndf        : 84
## br         : 1.2
## CI95(br)   : 0.72-1.3
##
```

The Birge ratio is now acceptable, although there are only 3 active control points. The others do not contribute significantly to the fit. The residuals are also improved (Fig. 9). The model is statistically valid, and the modulation curve can be used for physical interpretation.

References

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. 2008. "Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement (GUM)." 100:2008. Joint Committee for Guides in Metrology, JCGM. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_F.pdf.

Gelman, A., D. Lee, and J. Guo. 2015. "Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization." *J. Educ. Behav. Stat.* 40. American Educational Research Association (AERA): 530–43. https://doi.org/10.3102/1076998615606113.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. 3rd ed. Chapman; Hall/CRC. https://www.crcpress.com/Bayesian-Data-Analysis-Third-Edition/Gelman-Carlin-Stern-Dunson-Vehtari-Rubin/p/book/9781439840955?source=igodigital.

Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15: 1593–1623. http://jmlr.org/papers/v15/hoffman14a.html.

Mallick, H., and N. Yi. 2014. "A New Bayesian Lasso." *Statistics and Its Interface* 7: 571–82. https://doi.org/10.4310/SII.2014.v7.n4.a12.

Park, Trevor, and George Casella. 2008. "The Bayesian Lasso." *J. Am. Stat. Assoc.* 103: 681–86. https://doi.org/10.1198/016214508000000337.

Pernot, Pascal. 2017. "The Parameter Uncertainty Inflation Fallacy." *J. Chem. Phys.* 147 (10): 104102. https://doi.org/10.1063/1.4994654.

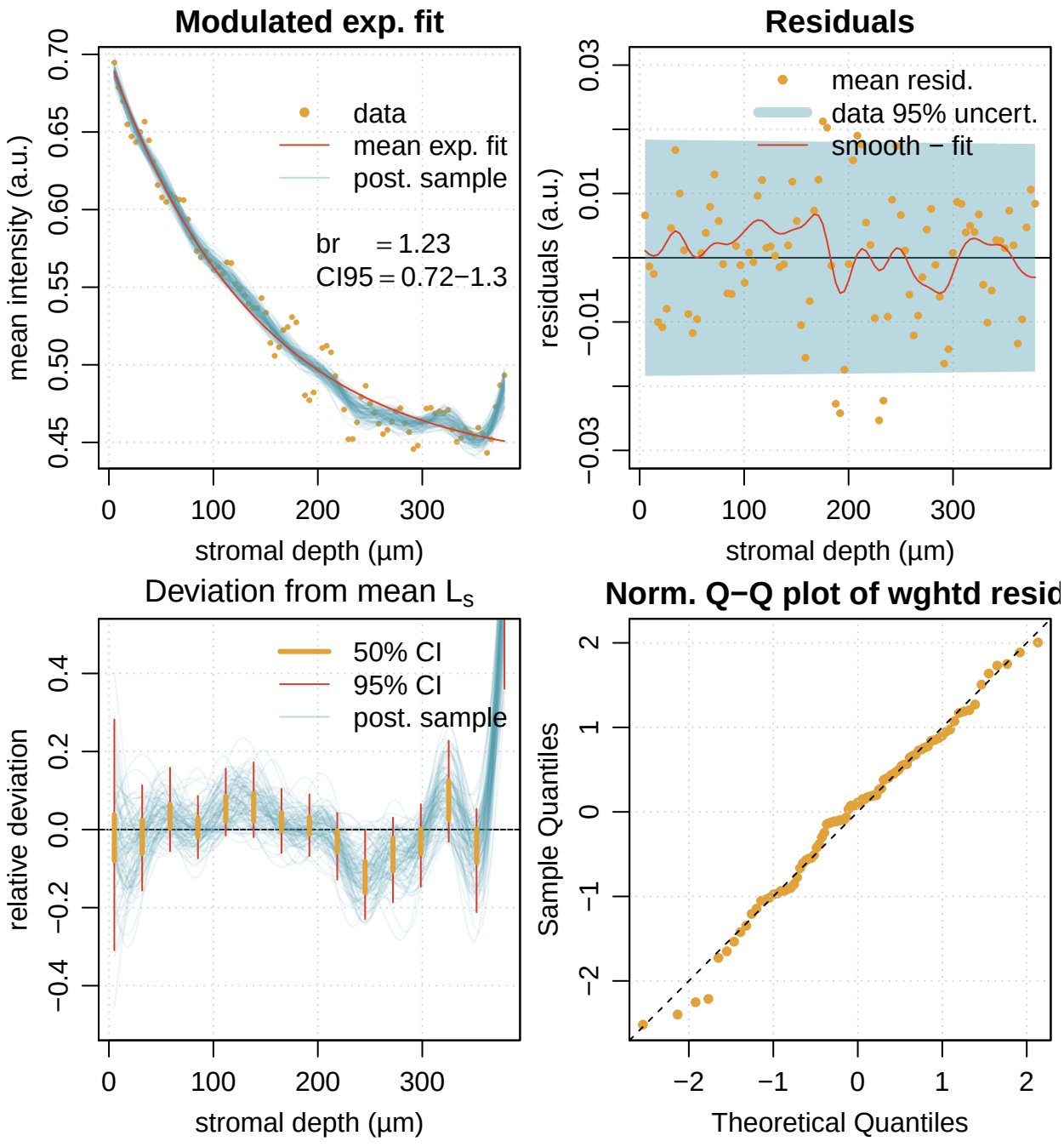

Figure 9: Modulated-exponential fit, residuals, Modulation function and Normal QQ-plot of residuals.

Pernot, Pascal, and Andreas Savin. 2018. "Probabilistic Performance Estimators for Computational Chemistry Methods: The Empirical Cumulative Distribution Function of Absolute Errors." *J. Chem. Phys.* 148: 241707. https://doi.org/10.1063/1.5016248.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Stan Development Team. 2016. *RStan: The R Interface to Stan*. http://mc-stan.org/.