Added uncertainty in the estimated temperature due to fixing parameters

Bas des Tombe (bdestombe@gmail.com), October 7, 2020

This note attempts to explain how parameters are fixed and to clarify why the uncertainty increases when fixing parameters. In the example the parameter gamma is fixed.

The parameter uncertainty of C and D increases when γ is fixed. Basically what you are estimating is how well does the linear system describes the observations that are corrected for the effects of the fixed parameter. Thus from the observations you subtract the effect of the fixed gamma (Equation 11). That results in a higher uncertainty of the, now corrected, observations (Equation 13). These observations, with a larger uncertainty, are used to estimate C and D, resulting in a larger uncertainty of C and D. The larger parameter uncertainty results in a wider confidence interval of the estimated temperature.

The following is copied and adjusted from: *des Tombe, B.; Schilperoort, B.; Bakker, M. Estimation of Temperature and Associated Uncertainty from Fiber-Optic Raman-Spectrum Distributed Temperature Sensing. Sensors 2020, 20, 2235.*

Single-Ended Calibration Procedure

In single-ended calibration, the temperature is estimated from Stokes and anti-Stokes intensity measurements with Equation **??**. The parameters that need to be estimated from calibration are γ, $\Delta\alpha$, and C, where C needs to be estimated for each time step. The parameters are estimated from the reference temperature at M locations along the reference sections and at N times. Equation **??** is reorganized to amend it for linear regression. The observation at location m and time n, denoted with $I_{m,n}$, is written as a linear combination of the unknown parameters:

$$I_{m,n} = \frac{1}{T_{m,n}}\gamma - x_m \Delta\alpha - C_n, \quad \text{with} \quad m = 1, 2, .., M \quad \text{and} \quad n = 1, 2, .., N \tag{1}$$

where $T_{m,n}$ is the reference temperature at location m and time n, x_m is the location of point m along the reference sections, and C_n is the constant C of the fiber at time n. In total, there are $N+2$ unknown parameters and MN observations.

The system of N Equation 1 for location m may be written in vector form as:

$$\mathbf{y}_m = \mathbf{X}_m \mathbf{a} + \epsilon_m, \tag{2}$$

where ϵ_m are the residuals between the observed values and the fitted values for location m, and

$$\mathbf{y}_m = \begin{bmatrix} I_{m,1} \\ I_{m,2} \\ \vdots \\ I_{m,N} \end{bmatrix}, \quad \mathbf{X}_m = \begin{bmatrix} \frac{1}{T_{m,1}} & -x_m & -1 & & \\ \frac{1}{T_{m,2}} & -x_m & & -1 & \\ \vdots & \vdots & & & \ddots & \\ \frac{1}{T_{m,N}} & -x_m & & & & -1 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \gamma \\ \Delta\alpha \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \tag{3}$$

The vector $\mathbf{a}$ contains the unknown parameters that are to be estimated. The system of MN equations for all locations may be combined into one system of equations:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \epsilon, \tag{4}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} \tag{5}$$

This system (Equation 4) is solved by minimizing the sum of the squared weighted residuals χ^2:

$$\chi^2 = (\mathbf{y} - \mathbf{Xa})^\intercal \mathbf{W} (\mathbf{y} - \mathbf{Xa}) \tag{6}$$

where $^\intercal$ refers to the transposed matrix and $\mathbf{W}$ is a diagonal matrix given by

$$\mathrm{diag}\,(\mathbf{W}) = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_M \end{bmatrix}, \quad \mathbf{W}_m = \begin{bmatrix} \frac{1}{\sigma^2_{I_{m,1}}} \\ \frac{1}{\sigma^2_{I_{m,2}}} \\ \vdots \\ \frac{1}{\sigma^2_{I_{m,N}}} \end{bmatrix} \tag{7}$$

The variance, $\sigma^2_{I_{m,n}}$, of the distribution of the noise in the observation at location m, time n, is a function of the variance of the noise in the Stokes and anti-Stokes intensity measurements ($\sigma^2_{P_+}$ and $\sigma^2_{P_-}$), and is approximated with (Ku et al., 1966):

$$\sigma^2_{I_{m,n}} \approx \left[\frac{\partial I_{m,n}}{\partial P_{m,n+}}\right]^2 \sigma^2_{P_+} + \left[\frac{\partial I_{m,n}}{\partial P_{m,n-}}\right]^2 \sigma^2_{P_-} \tag{8}$$

$$\approx \frac{1}{P^2_{m,n+}}\sigma^2_{P_+} + \frac{1}{P^2_{m,n-}}\sigma^2_{P_-} \tag{9}$$

The variance of the noise in the Stokes and anti-Stokes intensity measurements is estimated directly from Stokes and anti-Stokes intensity measurements using the steps outlined in Section **??**.

Fixed gamma

From an alternative calibration you have estimated γ and σ^2_γ. The adjusted system of N Equation 1 for location m may be written in vector form as:

$$\mathbf{y}'_m = \mathbf{X}'_m \mathbf{a}' + \epsilon'_m, \tag{10}$$

where ϵ_m are the residuals between the observed values and the fitted values for location m, and

$$\mathbf{y}'_m = \begin{bmatrix} I_{m,1} - \frac{\gamma}{T_{m,1}} \\ I_{m,2} - \frac{\gamma}{T_{m,2}} \\ \vdots \\ I_{m,N} - \frac{\gamma}{T_{m,N}} \end{bmatrix}, \quad \mathbf{X}'_m = \begin{bmatrix} -x_m & -1 & & & \\ -x_m & & -1 & & \\ \vdots & & & \ddots & \\ -x_m & & & & -1 \end{bmatrix}, \quad \mathbf{a}' = \begin{bmatrix} \Delta\alpha \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \tag{11}$$

The variance of $\mathbf{y}'_m$ increases as follows

$$\sigma^2_{I_{m,n}} \approx \left[\frac{\partial I_{m,n}}{\partial P_{m,n+}}\right]^2 \sigma^2_{P_+} + \left[\frac{\partial I_{m,n}}{\partial P_{m,n-}}\right]^2 \sigma^2_{P_-} + \frac{1}{T_{m,n}}\sigma^2_\gamma \tag{12}$$

$$\approx \frac{1}{P^2_{m,n+}}\sigma^2_{P_+} + \frac{1}{P^2_{m,n-}}\sigma^2_{P_-} + \frac{1}{T_{m,n}}\sigma^2_\gamma \tag{13}$$

Confidence Intervals of the Temperature

The uncertainty in the estimated temperature varies along the fiber as the laser pulse attenuates when propagating through the fiber, and varies over time due to varying gains and losses in the DTS device. The two sources that contribute to the uncertainty in the temperature estimate are the uncertainty in the calibrated parameters and the uncertainty associated with the noise in the Stokes and anti-Stokes intensity measurements. The former dominates the uncertainty in the estimated temperature for measurements with longer acquisition times, while the latter dominates measurements with shorter acquisition times. Other sources of possible uncertainty are not taken into account here. These include the uncertainty introduced by the model that relates measured Stokes and anti-Stokes intensities to temperature, and the uncertainty in measured temperatures obtained with external sensors. The latter is generally much smaller than the uncertainty in the DTS temperature from the noise in the Stokes and anti-Stokes intensity measurements.

Estimation of the confidence intervals of the temperature starts with estimating separate probability density functions for the Stokes and anti-Stokes intensity measurements and the calibrated parameters. The probability density functions are propagated through the model using a Monte Carlo sampling procedure following the steps from Joint Committee for Guides in Metrology (2008a) and Joint Committee for Guides in Metrology (2008b). This procedure results in an approximation of the probability density function for the estimated temperature, which is different at each location and varies over time. Various summarizing statistics are computed from the approximate probability density function, including the expected value, the standard deviation, and the confidence intervals. The standard deviation is also called the temperature resolution, but in line with Joint Committee for Guides in Metrology (2008a), the term standard uncertainty is used here. The procedure is explained first for single-ended measurements, followed by the procedure for double-ended measurements.

Single-Ended Measurements

Estimation of the confidence intervals for the temperatures measured with a single-ended setup consists of five steps. First, the variances of the Stokes and anti-Stokes intensity measurements are estimated following the steps in Section **??**. A Normal distribution is assigned to each intensity measurement that is centered at the measurement and using the estimated variance.

- Here, the original $\sigma^2_{I_{m,n}}$ is used that is not corrected for the fixed γ

Second, a multi-variate Normal distribution is assigned to the estimated parameters using the covariance matrix from the calibration procedure presented in Section .

- The observations have a larger variance when a parameter is fixed, thus the covariance matrix contains larger values and the parameters are estimated less certain.

Third, the distributions are sampled, and the temperature is computed with Equation **??**. Fourth, step three is repeated, e.g., 10,000 times for each location and for each time. The resulting 10,000 realizations of the temperatures approximate the probability density functions of the estimated temperature at that location and time. Fifth, the standard uncertainties are computed with the standard deviations of the realizations of the temperatures, and the 95% confidence intervals are computed from the 2.5% and 97.5% percentiles of the realizations of the temperatures.

- Which are wider due to fixing parameter γ.