Estimation of temperature and associated uncertainty from fiber-optic raman-spectrum distributed temperature sensing

Tombe, Bas des; Schilperoort, Bart; Bakker, Mark

Estimation of Temperature and Associated Uncertainty from Fiber-Optic Raman-Spectrum Distributed Temperature Sensing

Bas des Tombe *[ID], **Bart Schilperoort** [ID] and **Mark Bakker** [ID]

Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628CN Delft, The Netherlands; B.Schilperoort@tudelft.nl (B.S.); mark.bakker@tudelft.nl (M.B.)
* Correspondence: b.f.destombe@tudelft.nl

check for updates

Abstract: Distributed temperature sensing (DTS) systems can be used to estimate the temperature along optic fibers of several kilometers at a sub-meter interval. DTS systems function by shooting laser pulses through a fiber and measuring its backscatter intensity at two distinct wavelengths in the Raman spectrum. The scattering-loss coefficients for these wavelengths are temperature-dependent, so that the temperature along the fiber can be estimated using calibration to fiber sections with a known temperature. A new calibration approach is developed that allows for an estimate of the uncertainty of the estimated temperature, which varies along the fiber and with time. The uncertainty is a result of the noise from the detectors and the uncertainty in the calibrated parameters that relate the backscatter intensity to temperature. Estimation of the confidence interval of the temperature requires an estimate of the distribution of the noise from the detectors and an estimate of the multi-variate distribution of the parameters. Both distributions are propagated with Monte Carlo sampling to approximate the probability density function of the estimated temperature, which is different at each point along the fiber and varies over time. Various summarizing statistics are computed from the approximate probability density function, such as the confidence intervals and the standard uncertainty (the estimated standard deviation) of the estimated temperature. An example is presented to demonstrate the approach and to assess the reasonableness of the estimated confidence intervals. The approach is implemented in the open-source Python package "dtscalibration".

Keywords: distributed temperature sensing; DTS; fiber optic; Raman; Stokes; temperature; calibration; uncertainty; confidence intervals

1. Introduction

Temperatures can be measured with distributed temperature sensing (DTS) along optical fibers that may extend to several kilometers with a sub-meter resolution [1,2]. The application of DTS is of great value to characterize thermal dynamics at a scale that corresponds to, for example, many geophysical processes. Heat tracer tests with DTS have been used to estimate wind speed [3,4], evaporation [5,6], soil moisture [7], groundwater-surface water interaction [1,8], and groundwater flow [9,10]. The specification sheets of DTS measurement systems list the uncertainty in the temperature estimated under ideal conditions. In practice, the uncertainty in the temperature estimate varies along the fiber, is different for each setup, and is temperature-dependent. Therefore, the uncertainty in the temperature estimate should be estimated separately for every DTS measurement.

DTS systems that estimate temperature from Raman-backscatter measurements shoot a laser pulse through a fiber-optic cable, which is scattered back to the DTS system by inhomogeneities in the fiber. Most of the backscattered laser has the same wavelength as the emitted laser (Rayleigh

scattering), but a small fraction has different wavelengths (Raman scattering). The detectors in DTS systems measure the intensity of the backscatter at two distinct wavelengths: Stokes (-Raman) and anti-Stokes (-Raman) scatter. The temperature at the point of reflection is estimated from these two types of scatter. Stokes scatter has a longer wavelength than the laser and its intensity does not vary much with temperature, while anti-Stokes scatter has a shorter wavelength than the laser and its intensity varies significantly with temperature. The location of the measurement along the fiber is estimated from the time between sending the laser pulse and receiving the scatter. Temperature along the fiber is estimated from the measured intensities of the Stokes and anti-Stokes scatter by calibrating to reference sections with a known temperature. In practice, these fiber sections are submerged in water baths of which the temperature is continuously measured with a separate temperature sensor. The water can be mixed with small pumps in an attempt to equalize the temperature of the water in the baths. Sequential temperature measurements require continuous calibration due to varying gains and losses in the DTS system. Detailed calibration procedures are available in the literature [11–14]. The uncertainty in the temperature estimates is strongly affected by the calibration procedure [2,15].

Attenuation of light propagating through a fiber depends on its wavelength and therefore is different for Stokes and anti-Stokes scatter. The key in DTS calibration is to differentiate the attenuation from the temperature effects at the point of reflection. There are two types of setups to estimate temperature from Stokes and anti-Stokes scatter: single-ended and double-ended. In single-ended setups, only one end of the fiber is connected to the DTS system, and the difference in attenuation between the Stokes and the anti-Stokes scatter is approximated as constant over the fiber. However, this difference in attenuation is also affected by sharp bends, connectors, and bad splices, which may result in a shift in the temperature if not accounted for. In double-ended setups, measurements are carried out from both ends of the fiber to estimate the difference in attenuation between the Stokes and anti-Stokes scatter at each point along the fiber. In such setups, measurements from both ends of the fiber must be combined, including their uncertainty.

In this paper, a new calibration procedure is presented for the temperature estimates for single-ended and double-ended setups, including estimates of the uncertainty in the form of confidence intervals. First, a brief review is given of how temperature can be estimated from Stokes and anti-Stokes intensity measurements. Next, the calibration steps are outlined for single-ended and double-ended setups. The spatial and temporal variability in the uncertainty of the temperature are demonstrated by an example of the double-ended calibration procedure. The new calibration procedure is implemented in the open-source Python package "dtscalibration". This article concludes with a conclusion and a discussion with practical tips to reduce the uncertainty in the temperature.

2. Estimation of Temperature from Stokes and Anti-Stokes Scatter

The equations that relate the Stokes and anti-Stokes intensity measurements to temperature along a fiber are reviewed here briefly; a comprehensive resource that covers DTS theory is Hartog [16]. The path of a laser pulse emitted by a DTS system through a fiber is shown in Figure 1. The laser module emits a short laser pulse that is approximated here as an impulse with energy E_0. The pulse travels with the speed of light, v, through the fiber, during which it attenuates with $\exp\left(-\int_0^x \alpha_1 \, dx\right)$, where α_1 is the attenuation at the wavelength of the laser pulse. A small fraction, $S(x)R(x)\Delta x$, of the energy of the laser pulse is scattered back over section Δx, where S is the scattering-loss coefficient for the scattering at the wavelength of the laser pulse, and R is the capture fraction, which depends on the fiber composition (e.g., diameter and numerical aperture [17]). The backscatter pulse is $2\Delta t$ wide and is attenuated again on its way back to the detector. The intensity measured by the detector, P from scattering at x (between times $2x/v - \Delta t$ and $2x/v + \Delta t$) is given by [16] (Equation (3.5)):

$$P(x,t) = \tfrac{1}{2}E_0 \eta v S R \exp\left(-\int_0^x 2\alpha_1 \, dx\right) \tag{1}$$

where η corrects for the sensitivity of the detector and the attenuation between the detector and where the fiber end is connected to the DTS system [18].
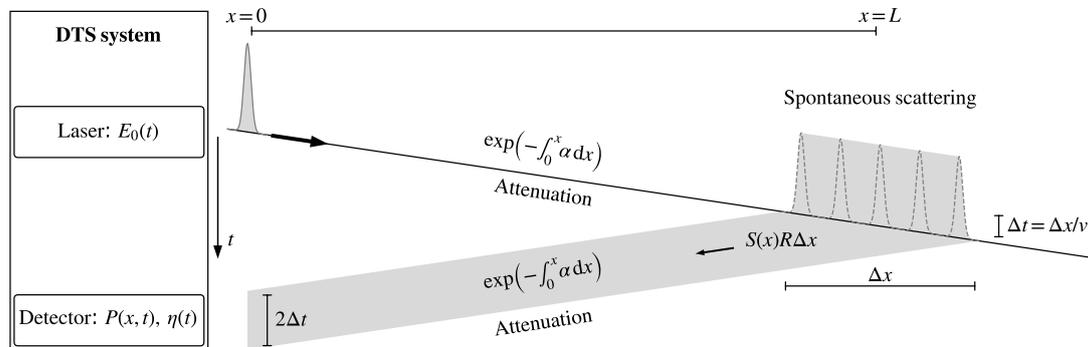

Figure 1. Path of a laser pulse.

In practice, the backscatter is not returned as a pulse, but is dispersed, resulting in spatially correlated measurements. The spreading of the pulse is caused by the finite width of the emitted laser pulse, the measurement response of the detector, and intermodal dispersion of the pulse while propagating through the fiber. The spreading of the pulse may be experimentally estimated following Simon et al. [19]. Most of the scattered energy has the same wavelength as the emitted laser pulse (Rayleigh scattering), but a small part has different wavelengths. The scattering-loss coefficient S_+ for the Stokes wavelength and S_- for the anti-Stokes wavelength are given by [16,20]:

$$S_+(x,t) = \frac{K_+}{\lambda_+^4} \frac{\exp\left[\gamma/T(x,t)\right]}{\exp\left[\gamma/T(x,t)\right] - 1} \tag{2}$$

$$S_-(x,t) = \frac{K_-}{\lambda_-^4} \frac{1}{\exp\left[\gamma/T(x,t)\right] - 1} \tag{3}$$

where K_+ and K_- correct for the fraction of the molecules in the fiber that scatter at the Stokes and anti-Stokes wavelengths [21], λ is the wavelength, T is the temperature along the fiber in Kelvin, and γ is the sensitivity of the Stokes and anti-Stokes scattering to temperature in Kelvin and depends on the fiber material.

The Stokes and anti-Stokes power (P_+ and P_-, respectively) measured at the detector(s) in a DTS system are given by:

$$P_+(x,t) = \tfrac{1}{2}E_0\eta_+ vS_+ R \exp\left[-\int_0^x (\alpha_l + \alpha_+) \, dx\right] \tag{4}$$

$$P_-(x,t) = \tfrac{1}{2}E_0\eta_- vS_- R \exp\left[-\int_0^x (\alpha_l + \alpha_-) \, dx\right] \tag{5}$$

where α_+ and α_- are the attenuation as a function of x at the Stokes and anti-Stokes wavelengths. Most DTS systems use the ratio of Equation (4) over Equation (5) [20]:

$$\frac{P_+(x,t)}{P_-(x,t)} = \frac{\eta_+(t)K_+/\lambda_+^4}{\eta_-(t)K_-/\lambda_-^4} \exp\left(-\int_0^x \Delta\alpha(x')dx'\right) \exp\left[\gamma/T(x,t)\right] \tag{6}$$

where Equations (2) and (3) are substituted for S_+ and S_-, respectively, and $\Delta\alpha$ $\left(\text{m}^{-1}\right)$ is the differential attenuation, $\Delta\alpha(x) = \alpha_+ - \alpha_-$. Taking the logarithm of Equation (6) gives:

$$I(x,t) = -C(t) - \int_0^x \Delta\alpha(x')\mathrm{d}x' + \frac{\gamma}{T(x,t)} \tag{7}$$

where

$$I(x,t) = \ln\left(\frac{P_+(x,t)}{P_-(x,t)}\right) \tag{8}$$

$$C(t) = \ln\left(\frac{\eta_-(t)K_-/\lambda_-^4}{\eta_+(t)K_+/\lambda_+^4}\right) \tag{9}$$

where C is the lumped effect of the difference in gain at $x = 0$ between Stokes and anti-Stokes intensity measurements and the dependence of the scattering intensity on the wavelength. An equation for T may be obtained from Equation (7) as:

$$T(x,t) = \frac{\gamma}{I(x,t) + C(t) + \int_0^x \Delta\alpha(x')\mathrm{d}x'} \tag{10}$$

The temperature along the fiber can now be estimated from the Stokes and anti-Stokes intensity measurements, I, if the terms γ, C, and $\int_0^x \Delta\alpha(x')\mathrm{d}x'$ are known. These terms are estimated by calibration to reference sections.

3. Integrated Differential Attenuation

The differential attenuation, $\Delta\alpha$, is different for each fiber type, varies along a fiber, and may change at sharp fiber bends and fiber connections. The integrated differential attenuation ($\int_0^x \Delta\alpha(x')\mathrm{d}x'$) differs per setup and must be estimated experimentally to differentiate it from a shift in the temperature. It is estimated differently in single-ended setups than in double-ended setups.

3.1. Single-Ended Measurements

In single-ended setups, Stokes and anti-Stokes intensity is measured from a single end of the fiber. The differential attenuation is assumed constant along the fiber so that the integrated differential attenuation may be written as [11]:

$$\int_0^x \Delta\alpha(x')\mathrm{d}x' \approx \Delta\alpha x \tag{11}$$

The temperature can now be written from Equation (10) as:

$$T(x,t) \approx \frac{\gamma}{I(x,t) + C(t) + \Delta\alpha x} \tag{12}$$

The parameters γ, $C(t)$, and $\Delta\alpha$ must be estimated from calibration to reference sections, as discussed in Section 5. The parameter C must be estimated for each time and is constant along the fiber. Confidence intervals for the estimated temperature are discussed in Section 7.1. When jumps in the integrated differential attenuation are expected, the measurements can be split into sections with a constant differential attenuation. Each additional section requires its own reference section to estimate the magnitude of the jump in the integrated differential attenuation.

3.2. Double-Ended Measurements

In double-ended setups, Stokes and anti-Stokes intensity is measured in two directions from both ends of the fiber. The forward-channel measurements are denoted with subscript F, and the backward-channel measurements are denoted with subscript B. Both measurement channels start at a different end of the fiber and have opposite directions, and therefore have different spatial coordinates.

The first processing step with double-ended measurements is to align the measurements of the two measurement channels so that they have the same spatial coordinates. The spatial coordinate x (m) is defined here positive in the forward direction, starting at 0 where the fiber is connected to the forward channel of the DTS system; the length of the fiber is L. Consequently, the backward-channel measurements are flipped and shifted to align with the forward-channel measurements. Alignment of the measurements of the two channels is prone to error because it requires the exact fiber length [15]. Depending on the DTS system used, the forward channel and backward channel are measured one after another by making use of an optical switch, so that only a single detector is needed. However, it is assumed in this paper that the forward channel and backward channel are measured simultaneously, so that the temperature of both measurements is the same. This assumption holds better for short acquisition times with respect to the time scale of the temperature variation, and when there is no systematic difference in temperature between the two channels. The temperature may be computed from the forward-channel measurements (Equation (10)) with:

$$T_F(x,t) = \frac{\gamma}{I_F(x,t) + C_F(t) + \int_0^x \Delta\alpha(x')\mathrm{d}x'} \tag{13}$$

and from the backward-channel measurements with:

$$T_B(x,t) = \frac{\gamma}{I_B(x,t) + C_B(t) + \int_x^L \Delta\alpha(x')\mathrm{d}x'} \tag{14}$$

where $C_F(t)$ and $C_B(t)$ are the parameter $C(t)$ for the forward-channel and backward-channel measurements, respectively. $C_B(t)$ may be different from $C_F(t)$ due to differences in gain, and difference in the attenuation between the detectors and the point the fiber end is connected to the DTS system (η_+ and η_- in Equation (9)). The calibration procedure presented in van de Giesen et al. [12] approximates $C(t)$ to be the same for the forward and backward-channel measurements, but this approximation is not made here.

Parameter $A(x)$ is introduced to simplify the notation of the double-ended calibration procedure and represents the integrated differential attenuation between locations x_1 and x along the fiber. Location x_1 will be selected in Section 6.

$$A(x) = \int_{x_1}^x \Delta\alpha(x')\mathrm{d}x' \tag{15}$$

so that Equations (13) and (14) may be written as:

$$T_F(x,t) = \frac{\gamma}{I_F(x,t) + D_F(t) + A(x)} \tag{16}$$

$$T_B(x,t) = \frac{\gamma}{I_B(x,t) + D_B(t) - A(x)} \tag{17}$$

where

$$D_F(t) = C_F(t) + \int_0^{x_1} \Delta\alpha(x')\mathrm{d}x' \tag{18}$$

$$D_B(t) = C_B(t) + \int_{x_1}^L \Delta\alpha(x')\mathrm{d}x' \tag{19}$$

Parameters D_F and D_B must be estimated for each time and are constant along the fiber, and parameter A must be estimated for each location and is constant over time. The calibration procedure is discussed in Section 6. T_F and T_B are separate approximations of the same temperature at the same time. The estimated T_F is more accurate near $x = 0$ because that is where the signal is strongest. Similarly, the estimated T_B is more accurate near $x = L$. A single best estimate of the temperature is obtained from the weighted average of T_F and T_B as discussed in Section 7.2.

4. Estimation of the Variance of the Noise in the Intensity Measurements

The Stokes and anti-Stokes intensities are measured with detectors, which inherently introduce noise to the measurements. Knowledge of the distribution of the measurement noise is needed for a calibration with weighted observations (Sections 5 and 6) and to project the associated uncertainty to the temperature confidence intervals (Section 7). Two sources dominate the noise in the Stokes and anti-Stokes intensity measurements [16] (p. 125). Close to the laser, noise from the conversion of backscatter to electricity dominates the measurement noise. The detecting component, an avalanche photodiode, produces Poisson-distributed noise with a variance that increases linearly with the intensity. The Stokes and anti-Stokes intensities are commonly much larger than the standard deviation of the noise, so that the Poisson distribution can be approximated with a Normal distribution with a mean of zero and a variance that increases linearly with the intensity. At the far-end of the fiber, noise from the electrical circuit dominates the measurement noise. It produces Normal-distributed noise with a mean of zero and a variance that is independent of the intensity. However, in this paper the sum of the two noise sources is approximated with a single Normal distribution with a variance that is independent of the intensity. This approximation holds better for small setups with little attenuation and DTS systems with avalanche photodiodes with a lower gain. In Appendix A, a procedure is presented to compute the intensity-dependent variance of the noise for when this approximation does not hold.

The variance of the noise in the Stokes ($\sigma_{P_+}^2$) and anti-Stokes ($\sigma_{P_-}^2$) measurements are estimated by fitting Equations (4) and (5) through the Stokes and anti-Stokes intensity measurements of the reference sections, after which the distribution of the residuals is used as an estimate of the distribution of the noise [22]. Fiber sections that are used for calibration have an approximately spatially uniform reference temperature ($T_r(t)$), so that Equations (4) and (5) are each expressed in a term that varies along the reference section but is constant over time ($H(x)$) and a term that varies with time ($G(t)$) but is constant for a reference section:

$$P_+(x,t) = G_+(t)H_+(x) \tag{20}$$

$$P_-(x,t) = G_-(t)H_-(x) \tag{21}$$

with

$$G_+(t) = E_0(t)\eta_+(t)\frac{K_+}{\lambda_+^4}\frac{\exp\left[\gamma/T_r(t)\right]}{\exp\left[\gamma/T_r(t)\right]-1} \tag{22}$$

$$G_-(t) = E_0(t)\eta_-(t)\frac{K_-}{\lambda_-^4}\frac{1}{\exp\left[\gamma/T_r(t)\right]-1} \tag{23}$$

$$H_+(x) = \tfrac{1}{2}vR\exp\left(-\int_0^x\left[\alpha_1(x')+\alpha_+(x')\right]\mathrm{d}x'\right) \tag{24}$$

$$H_-(x) = \tfrac{1}{2}vR\exp\left(-\int_0^x\left[\alpha_1(x')+\alpha_-(x')\right]\mathrm{d}x'\right) \tag{25}$$

Values for $G(t)$ and $H(x)$ are obtained by minimizing the residuals between the Stokes and anti-Stokes intensity measurements and Equations (20) and (21) with a standard routine. Here, the "minimize" routine of the Python package scipy.optimize [23] is used, and is implemented in the package "dtscalibration" (Section 8). $G(t)$ and $H(x)$ are, of course, highly correlated (Equations (20) and (21)), but that is not relevant here as only the product of $G(t)$ and $H(x)$ is of interest. The residuals between the fitted product of $G(t)$ and $H(x)$ and the Stokes and anti-Stokes intensity measurements are attributed to the noise from the detectors. The variance of the residuals is used as a proxy for the variance of the noise in the Stokes and anti-Stokes intensity measurements. The consequence of a non-uniform temperature of the reference sections on the estimated variance is described in Section 10.

5. Single-Ended Calibration Procedure

In single-ended calibration, the temperature is estimated from Stokes and anti-Stokes intensity measurements with Equation (12). 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 (7) 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:

$$I_{m,n} = \frac{1}{T_{m,n}}\gamma - x_m\Delta\alpha - C_n, \quad \text{with} \quad m = 1, 2, ..., M \quad \text{and} \quad n = 1, 2, ..., N \tag{26}$$

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 (26) for location m may be written in vector form as:

$$\mathbf{y}_m = \mathbf{X}_m\mathbf{a} + \boldsymbol{\epsilon}_m, \tag{27}$$

where $\boldsymbol{\epsilon}_m$ are the residuals between the observed values and the fitted values for location m, and

$$\mathbf{y}_m = \begin{bmatrix} I_{m,1} \\ I_{m,2} \\ \vdots \\ I_{m,N} \end{bmatrix}, \quad \mathbf{X}_m = \begin{bmatrix} \frac{1}{T_{m,1}} & -x_m & -1 & & \\ \frac{1}{T_{m,2}} & -x_m & & -1 & \\ \vdots & \vdots & & & \ddots \\ \frac{1}{T_{m,N}} & -x_m & & & -1 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \gamma \\ \Delta\alpha \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \tag{28}$$

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:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\epsilon}, \tag{29}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_M \end{bmatrix} \tag{30}$$

This system (Equation (29)) is solved by minimizing the sum of the squared weighted residuals χ^2:

$$\chi^2 = (\mathbf{y} - \mathbf{X}\mathbf{a})^\mathsf{T}\mathbf{W}(\mathbf{y} - \mathbf{X}\mathbf{a}) \tag{31}$$

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

$$\text{diag}\,(\mathbf{W}) = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_M \end{bmatrix}, \quad \mathbf{W}_m = \begin{bmatrix} \frac{1}{\sigma^2_{I_{m,1}}} \\ \frac{1}{\sigma^2_{I_{m,2}}} \\ \vdots \\ \frac{1}{\sigma^2_{I_{m,N}}} \end{bmatrix} \tag{32}$$

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 [24]:

$$\sigma^2_{I_{m,n}} \approx \left[\frac{\partial I_{m,n}}{\partial P_{m,n+}}\right]^2 \sigma^2_{P_+} + \left[\frac{\partial I_{m,n}}{\partial P_{m,n-}}\right]^2 \sigma^2_{P_-} \tag{33}$$

$$\approx \frac{1}{P^2_{m,n+}}\sigma^2_{P_+} + \frac{1}{P^2_{m,n-}}\sigma^2_{P_-} \tag{34}$$

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 4.

The sum of the squared weighted residuals may be minimized with a standard routine. In this paper, a custom sparse implementation of the weighted least squares routine of the Python package Statsmodels is used [25], which returns the optimal parameter set, the covariance matrix of the optimal parameter set, and the residuals (this is part of the Python package "dtscalibration", which is described in Section 8). The number of reference sections required for calibration is discussed in Section 10.2.

6. Double-Ended Calibration Procedure

In double-ended calibration, the temperature is estimated using Stokes and anti-Stokes intensity measurements from both ends of the fiber. The temperature is estimated from Equations (16) and (17). The unknown parameters γ, $D_F(t)$, $D_B(t)$, and $A(x)$ are obtained from calibration to reference sections, similar to the single-ended calibration procedure. x_1 is chosen as the first location of the first reference section, so that the value for A at that location equals 0 (Equation (15)). There are M locations along the reference sections where the temperature is measured at N times. In total there are $2N + M$ unknowns: parameter γ, N parameters D_F, N parameters D_B, and $M - 1$ parameters A (since $A_1 = 0$). Equations (16) and (17) are reorganized to amend them for linear regression as:

$$\left.\begin{aligned} I_{F,m,n} &= \frac{1}{T_{m,n}}\gamma - D_{F,n} - A_m \\ I_{B,m,n} &= \frac{1}{T_{m,n}}\gamma - D_{B,n} + A_m \end{aligned}\right\} \quad \text{with} \quad m = 1,2,...,M \quad \text{and} \quad n = 1,2,...,N \tag{35}$$

where $D_{F,n}$ is the constant D_F at time n, $D_{B,n}$ is the constant D_B at time n A_m is the constant A at location m, and $A_1 = 0$. The system of $2N$ equations for location m may be written as:

$$\mathbf{y}_m = \mathbf{X}_m \mathbf{a} + \boldsymbol{\epsilon}_m, \tag{36}$$

where

$$\mathbf{y}_m = \begin{bmatrix} \mathbf{F}_m \\ \mathbf{B}_m \end{bmatrix}, \quad \mathbf{X}_m = \begin{bmatrix} \mathbf{R}_m & -\mathbf{I}(N,N) & \mathbf{0}(N,N) & -\mathbf{Q}_m \\ \mathbf{R}_m & \mathbf{0}(N,N) & -\mathbf{I}(N,N) & \mathbf{Q}_m \end{bmatrix}, \tag{37}$$

$$\mathbf{F}_m = \begin{bmatrix} I_{F,m,1} \\ I_{F,m,2} \\ \vdots \\ I_{F,m,N} \end{bmatrix}, \quad \mathbf{B}_m = \begin{bmatrix} I_{B,m,1} \\ I_{B,m,2} \\ \vdots \\ I_{B,m,N} \end{bmatrix}, \quad \mathbf{R}_m = \begin{bmatrix} \frac{1}{T_{m,1}} \\ \frac{1}{T_{m,2}} \\ \vdots \\ \frac{1}{T_{m,N}} \end{bmatrix} \tag{38}$$

where $\mathbf{I}$ is an identity matrix with its number of rows and columns between brackets, $\mathbf{0}$ is a matrix of zeros with its number of rows and columns between brackets, and $\boldsymbol{\epsilon}_m$ are the residuals at location m. The matrix $\mathbf{Q}_m$ is defined differently for the first location than for the other locations. For $m = 1$,

$\mathbf{Q}_m$ is a matrix of zeros size N by $M - 1$. For $m > 1$, $\mathbf{Q}_m$ is a matrix of zeros of size N by $M - 1$ except for column $m - 1$, which contains ones. The vector $\mathbf{a}$ contains the unknown parameters, has length $2N + M$, and is given by:

$$\mathbf{a} = \begin{bmatrix} \gamma \\ \mathbf{a}_{D,\mathrm{F}} \\ \mathbf{a}_{D,\mathrm{B}} \\ \mathbf{a}_A \end{bmatrix}, \quad \mathbf{a}_{D,\mathrm{F}} = \begin{bmatrix} D_{\mathrm{F},1} \\ D_{\mathrm{F},2} \\ \vdots \\ D_{\mathrm{F},N} \end{bmatrix}, \quad \mathbf{a}_{D,\mathrm{B}} = \begin{bmatrix} D_{\mathrm{B},1} \\ D_{\mathrm{B},2} \\ \vdots \\ D_{\mathrm{B},N} \end{bmatrix}, \quad \mathbf{a}_A = \begin{bmatrix} A_2 \\ A_3 \\ \vdots \\ A_M \end{bmatrix} \tag{39}$$

The weights of the observations at location m are given by:

$$\mathrm{diag}\,(\mathbf{W}_m) = \begin{bmatrix} 1/\sigma^2\,[\mathbf{F}_m] \\ 1/\sigma^2\,[\mathbf{B}_m] \end{bmatrix} \tag{40}$$

where the elements of $\sigma^2\,[\mathbf{F}_m]$ and $\sigma^2\,[\mathbf{B}_m]$ are approximated with Equation (34) applied to the forward and backward channels, respectively.

The equations for all M locations are gathered in a single set of equations given by Equations (29) and (30), where $\mathbf{y}_m$, $\mathbf{X}_m$, $\mathbf{a}$ are given by Equations (37)–(39). This is a system of $2MN$ equations for $M + 2N$ unknown parameters. The system is solved by minimizing χ^2 as given by Equation (31), where $\mathbf{W}$ is given by Equation (32) with Equation (40) for $\mathbf{W}_m$.

The set of estimated parameters contains estimates of A for the locations along the reference sections only. An estimate of A and its variance for locations outside the reference sections are required to estimate the temperature outside of the reference sections. An estimate of A for location p and time n outside the reference sections is obtained by setting $T_\mathrm{F} = T_\mathrm{B}$ using Equations (16) and (17), which gives:

$$A_{p,n} = \frac{I_{\mathrm{B},p,n} - I_{\mathrm{F},p,n}}{2} + \frac{D_{\mathrm{B},n} - D_{\mathrm{F},n}}{2} \tag{41}$$

The variance of $A_{p,n}$ is approximated with:

$$\sigma^2\,[A_{p,n}] \approx \tfrac{1}{4}\left(\sigma^2\,[I_{\mathrm{B},p,n}] + \sigma^2\,[I_{\mathrm{F},p,n}] + \sigma^2\,[D_{\mathrm{F},n}] + \sigma^2\,[D_{\mathrm{B},n}] - 2\sigma\,[D_{\mathrm{F},n}, D_{\mathrm{B},n}]\right) \tag{42}$$

$$\approx \tfrac{1}{4}\left(\frac{1}{P_{\mathrm{B},m,n+}^2}\sigma_{P_{\mathrm{B}+}}^2 + \frac{1}{P_{\mathrm{B},m,n-}^2}\sigma_{P_{\mathrm{B}-}}^2 + \frac{1}{P_{\mathrm{F},m,n+}^2}\sigma_{P_{\mathrm{F}+}}^2 + \frac{1}{P_{\mathrm{F},m,n-}^2}\sigma_{P_{\mathrm{F}-}}^2 \right.$$
$$\left. + \sigma^2\,[D_{\mathrm{F},n}] + \sigma^2\,[D_{\mathrm{B},n}] - 2\sigma\,[D_{\mathrm{F},n}, D_{\mathrm{B},n}]\right) \tag{43}$$

where $\sigma_{P_{\mathrm{B}+}}^2$, $\sigma_{P_{\mathrm{B}-}}^2$, $\sigma_{P_{\mathrm{F}+}}^2$, and $\sigma_{P_{\mathrm{F}-}}^2$ are estimated directly from the Stokes and anti-Stokes intensity measurements of the forward and the backward channels using the procedure presented in Section 4, the terms $\sigma^2\,[D_{\mathrm{F},n}]$ and $\sigma^2\,[D_{\mathrm{B},n}]$ are parameter uncertainties from the diagonal of the covariance matrix, and the term $\sigma\,[D_{\mathrm{F},n}, D_{\mathrm{B},n}]$ is the square-root of the covariance between $D_{\mathrm{F},n}$ and $D_{\mathrm{B},n}$ from the covariance matrix. A single estimate of A_p and $\sigma^2\,[A_p]$ at location p is obtained using the inverse temporal-variance weighted mean:

$$A_p = \sigma^2\,[A_p]\left(\sum_{n=1}^{N} \frac{A_{p,n}}{\sigma^2\,[A_{p,n}]}\right) \tag{44}$$

$$\sigma^2\,[A_p] = \frac{1}{\displaystyle\sum_{n=1}^{N} 1/\sigma^2\,[A_{p,n}]} \tag{45}$$

The number of reference sections required for calibration is discussed in Section 10.2.

7. 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 JCGM [26,27]. 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 JCGM [26], the term standard uncertainty is used here. The procedure is explained first for single-ended measurements, followed by the procedure for double-ended measurements.

7.1. 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 4. A Normal distribution is assigned to each intensity measurement that is centered at the measurement and using the estimated variance. Second, a multi-variate Normal distribution is assigned to the estimated parameters using the covariance matrix from the calibration procedure presented in Section 5. Third, the distributions are sampled, and the temperature is computed with Equation (12). 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.

7.2. Double-Ended Measurements

Double-ended setups require four additional steps to estimate the confidence intervals for the temperature. First, the variances of the Stokes and anti-Stokes intensity measurements of the forward and backward channels are estimated following the steps in Section 4. A Normal distribution is assigned to each intensity measurement that is centered at the measurement and using the estimated variance. Second, a multi-variate Normal distribution is assigned to the estimated parameters using the covariance matrix from the calibration procedure presented in Section 6. Third, Normal distributions are assigned for A for each location outside of the reference sections. These distributions are centered around A_p and have variance $\sigma^2\left[A_p\right]$ given by Equations (44) and (45). Fourth, the distributions are sampled and $T_{\mathrm{F},m,n}$ and $T_{\mathrm{B},m,n}$ are computed with Equations (16) and (17), respectively. Fifth, step four is repeated to compute, e.g., 10,000 realizations of $T_{\mathrm{F},m,n}$ and $T_{\mathrm{B},m,n}$ to approximate their

probability density functions. Sixth, the standard uncertainties of $T_{\mathrm{F},m,n}$ and $T_{\mathrm{B},m,n}$ ($\sigma\left[T_{\mathrm{F},m,n}\right]$ and $\sigma\left[T_{\mathrm{B},m,n}\right]$) are estimated with the standard deviation of their realizations. Seventh, for each realization i the temperature $T_{m,n,i}$ is computed as the weighted average of $T_{\mathrm{F},m,n,i}$ and $T_{\mathrm{B},m,n,i}$:

$$T_{m,n,i} = \sigma^2\left[T_{m,n}\right]\left(\frac{T_{\mathrm{F},m,n,i}}{\sigma^2\left[T_{\mathrm{F},m,n}\right]} + \frac{T_{\mathrm{B},m,n,i}}{\sigma^2\left[T_{\mathrm{B},m,n}\right]}\right) \tag{46}$$

where

$$\sigma^2\left[T_{m,n}\right] = \frac{1}{1/\sigma^2\left[T_{\mathrm{F},m,n}\right] + 1/\sigma^2\left[T_{\mathrm{B},m,n}\right]} \tag{47}$$

The best estimate of the temperature $T_{m,n}$ is computed directly from the best estimates of $T_{\mathrm{F},m,n}$ and $T_{\mathrm{B},m,n}$ as:

$$T_{m,n} = \sigma^2\left[T_{m,n}\right]\left(\frac{T_{\mathrm{F},m,n}}{\sigma^2\left[T_{\mathrm{F},m,n}\right]} + \frac{T_{\mathrm{B},m,n}}{\sigma^2\left[T_{\mathrm{B},m,n}\right]}\right) \tag{48}$$

Alternatively, the best estimate of $T_{m,n}$ can be approximated with the mean of the $T_{m,n,i}$ values. Finally, the 95% confidence interval for $T_{m,n}$ are estimated with the 2.5% and 97.5% percentiles of $T_{m,n,i}$.

8. Python Implementation

The presented calibration procedure is implemented in the Python package "dtscalibration" [28]. It is open source, has a BSD 3-or-later license and is available online at https://github.com/dtscalibration/python-dts-calibration, together with installation instructions, examples, and documentation. The package reads DTS measurement files into an object, which has several calibration and plotting methods. Calibration and calculation of confidence intervals can be conducted within 10 lines of Python code, as is demonstrated in the examples of Section 9 and Supplementary Materials. Several routines are implemented to read Stokes and anti-Stokes intensity measurement files from the following DTS devices: AP Sensing CP320, Sensornet Oryx, Sensornet Halo, SensorTran 5100, Silixa Ultima, Silixa XT. The package inherits many functions (e.g., visualization, parallel computing) from xarray [29] so that the code base remains small. Most computations are performed by Dask [30] in chunks and in parallel, so that gigabytes of DTS measurement data can be processed on a personal computer with limited memory.

9. Example

An example of a double-ended setup is presented to demonstrate the spatial and temporal variability in the uncertainty of the temperature. This example attempts to estimate the uncertainty of the temperature along the entire fiber for a given acquisition time. The uncertainty in the estimated temperature can easily be reduced by increasing the acquisition time, but that limits the ability to observe temporal variation in temperature. Alternatively, the uncertainty can be reduced by increasing the sampling distance, but that limits the ability to observe spatial variation in temperature.

9.1. Setup and Data Collection

A schematic representation of the setup is shown in Figure 2. Sections of the fiber are submerged in water baths with a measured temperature as listed in Table 1. A uniform temperature in all water baths is desired so that the temperature measured with an external temperature sensor resembles the temperature of the fiber. Therefore, an aluminum bath filled with water is placed in a cold climate room, and a second one is placed in a warm climate room. The air in the climate rooms circulates around the water baths and the air temperature is kept constant with a maximum variation of 0.5 °C to achieve a uniform temperature in the water baths. A cooler turns on if the air temperature in the climate rooms is 0.3 °C above its target temperature, and a heater turns on if the air temperature in the

climate rooms is 0.2 °C below the target temperature. Both baths contain two coils of fiber. Another coil of fiber is placed in a cooler box filled with water, without temperature control, but with an aquarium air pump to attempt to mix the water and achieve a uniform temperature. All coils of fiber are lifted from the bottom of the baths with plastic spacers so that the coils are approximately centered. The setup was left to rest to achieve a stable and uniform temperature in the water baths before starting the measurements.
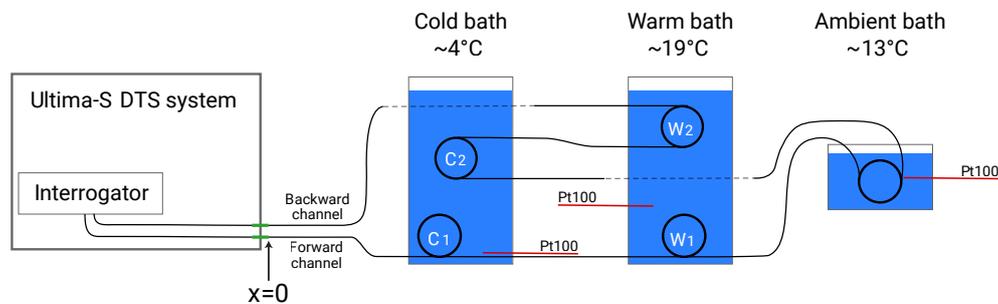
Figure 2. The setup of the example.

Table 1. Fiber sections submerged in water baths.

Name	Fiber Section (m)	Average Temperature (°C)	Number of Measurement Locations	Notes
Cold 1	7.5–17.0	4.35	37	Used for calibration
Warm 1	24.0–34.0	18.52	39	Used for calibration
Ambient	40.0–50.0	12.62	39	
Cold 2	70.0–80.0	4.35	39	
Warm 2	85.0–95.0	18.52	39	

The DTS system used in this example is a Silixa Ultima-S DTS system (Hertfordshire, UK) that measures fibers up to 2 km with a factory-reported spatial resolution of approximately 30 cm. The system is configured to measure the Stokes and anti-Stokes intensity with an acquisition time of 2 s in the forward direction and 2 s in the backward direction, every 12.7 cm along 100 m of fiber (Figure 2). The system's specification sheet lists several values for the standard uncertainty (referred to as the temperature resolution) for different single-ended setups using the built-in calibration routine. According to the specification sheet, the standard uncertainty is 0.34 °C if the fiber is shorter than 500 m, sampled every 12.7 cm, and measured for a second, which translates roughly to a standard uncertainty of 0.17 °C for an acquisition time of 4 s. The used fiber (j-BendAble made by j-fiber GmbH, Jena, Germany) is an OM3 fiber with a germanium-doped core with a diameter of 50 μm and a silica cladding with a diameter of 125 μm. The temperature in the warm- and cold-water baths is measured by the DTS system with two Pt100 RTDs, and the temperature in the ambient water bath is measured with a Pt100 RTD using a Fluke 1524 Handheld Thermometer.

9.2. Estimation of the Temperature and the Associated Uncertainty

The variance of the noise in the Stokes and anti-Stokes intensity measurements of the forward and backward channels were estimated using measurements from the two reference sections ('Cold 1' and 'Warm 1' in Table 1) with the procedure described in Section 4. A scatter plot with the residuals of the intensity measurements of the forward channel is shown in Appendix B. The residuals of the Stokes intensity measurements are plotted on the horizontal axis and the residuals of the anti-Stokes intensity measurements on the vertical axis. The residuals may be approximated with a Normal distribution and show no clear correlation (Pearson correlation coefficient is 0.02).

Double-ended calibration was performed to estimate the optimal parameter values and their covariance following the procedure described in Section 6. The calibration uses measurements from the two reference sections (Table 1). Instead of using all the measurements from the measurement period of one day, 1000 time samples were used, evenly spaced over a single day ($N = 1000$). Inherent to DTS systems, neighboring measurement locations are spatially correlated (Section 2). To reduce this correlation, every other measurement location was disregarded so that the distance between the used measurement locations is 25.4 cm, which is close to the spatial resolution and results in a total of 193 measurement locations. The number of locations along the reference sections that are used for calibration is 76 ($M = 76$).

The 95% confidence intervals are estimated with the approach described in Section 7. The estimated temperature and its confidence interval at the first time step are shown in Figure 3a. The difference between the estimated temperature and reference temperature is shown in Figure 3b with markers, where the confidence interval minus the estimated temperature at the first time step is shown with a blue fill. For the estimation of the confidence intervals, a Monte Carlo sample size of 10,000 was used; doubling the sample size for this measurement setup did not change the 95% confidence intervals significantly, but it did smooth the edges of the confidence intervals (i.e., the roughness of the edges of the blue fill in Figure 3b). The percentage of reference temperature measurements that fall within the 95% confidence interval of the estimated temperature for all 1000 time steps are listed for each section in Table 2. For the ambient bath, only 92.3% of the temperature measurements fall within the 95% confidence interval.
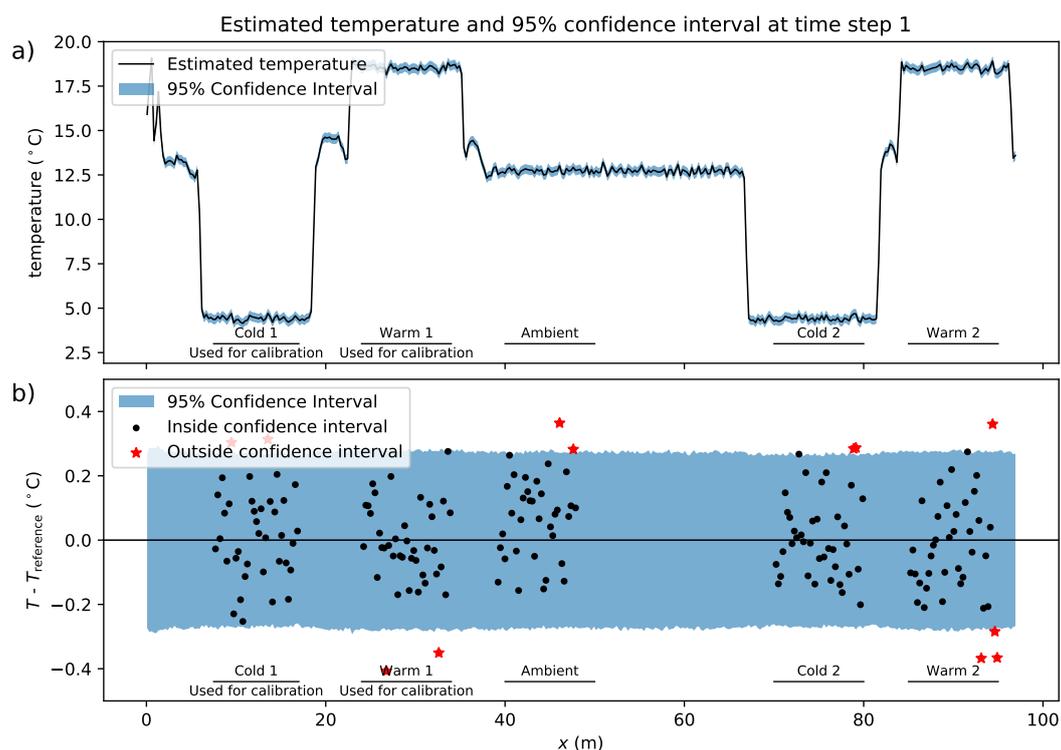

Figure 3. (**a**) Temperature with its 95% confidence intervals at the first time step. (**b**) Differences between the estimated temperature and the reference temperature at the first time step.

Table 2. Percentage of reference temperature within estimated 95% confidence intervals.

Cold 1	Warm 1	Ambient	Cold 2	Warm 2	Total
95.6%	95.0%	92.3%	94.7%	94.3%	94.4%

The mean difference between the reference temperature and the estimated temperature for all times is shown for each location along the reference sections in Figure 4 with orange markers. The estimated temperature of the ambient bath is on average 0.06 °C above the corresponding reference temperature. This bias may be the result of either the external temperature sensor not representing the temperature of the fiber due to local temperature variations in the water bath, or the external temperature sensor is calibrated differently. This could have been confirmed by simultaneously submerging all temperature sensors in a water bath and comparing their measurements, but this was not done here. The mean difference of reference section Warm 2 shows a spatial correlation between the mean differences, which can indicate non-uniform temperatures in the water bath. The spatial variability in the standard deviation of the differences between the reference temperature and the estimated temperature is shown in Figure 4 with blue markers. The standard uncertainty is estimated for each location from the Monte Carlo samples from all times and is shown with a black line. It varies with the temperature, with slightly smaller values for the fiber sections in the cold baths and slightly larger values for the fiber sections in the warm baths. The standard deviation of the difference between the reference temperature and the estimated temperature for all times is shown with blue markers and is well described at each location with the standard uncertainty.
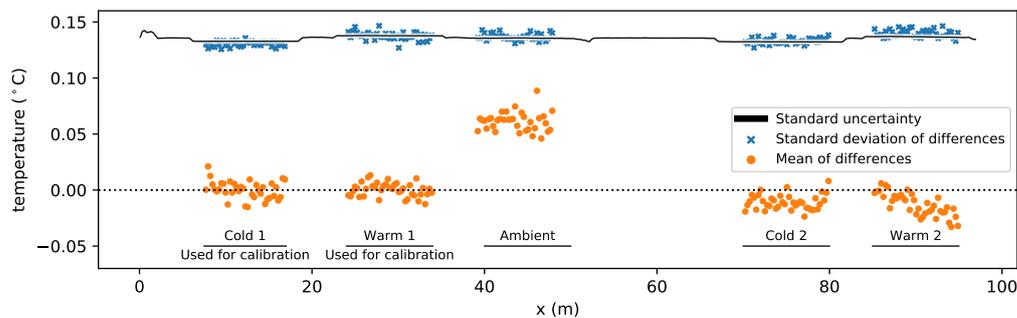

Figure 4. Spatial variation of the standard uncertainty of the estimated temperature, and the mean and standard deviation of the differences between the estimated and reference temperature.

The mean difference between the reference temperature and the estimated temperature for all reference locations is plotted over time in Figure 5 with orange markers. Each marker represents the mean of 193 differences. The mean of the differences are slightly above zero probably caused by the bias in the ambient bath temperature. The standard uncertainty at each time is computed from the Monte Carlo samples from all locations and is shown with a black line. The standard deviation of the difference between the reference and estimated temperature for all locations is shown with blue markers, which roughly follow the estimated standard uncertainties. The increase in standard uncertainty after 11.5 h coincides with a strong decrease in Stokes and anti-Stokes intensity of the forward and backward channels (not shown here), caused by either a decrease in laser strength or a decrease in sensitivity of the detector.
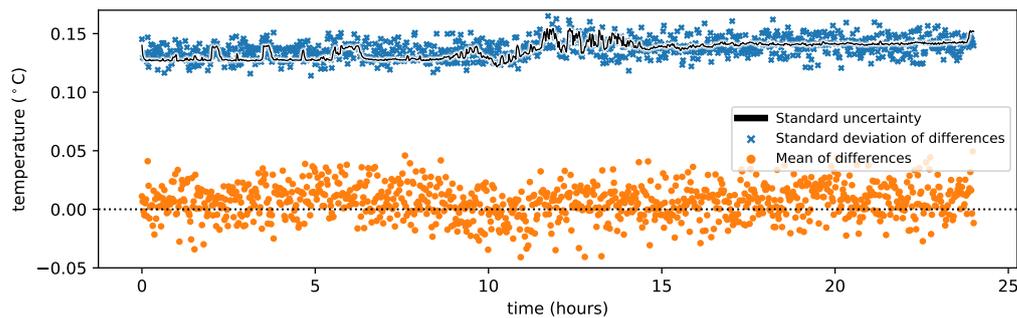
Figure 5. Temporal variation of the standard uncertainty of the estimated temperature, and the mean and standard deviation of the differences between the estimated and reference temperature.

9.3. Effect of Parameter Uncertainty

As stated, the uncertainty of the temperature is a function of the uncertainty in the estimated parameters and the uncertainty in the Stokes and anti-Stokes intensity measurements. The estimation of the standard uncertainty of the temperature was repeated except for that the covariance matrix of the optimal parameters was not propagated to the Monte Carlo set, so that the standard uncertainty consists of only the uncertainty introduced by the noise in the Stokes and anti-Stokes intensity measurements (Section 7.2). The standard uncertainty of the temperature was recalculated, and was on average 0.001 °C smaller than the original values and did not show any patterns or trends. Hence, it is concluded that for this experiment the contribution of the parameter uncertainty to the standard uncertainty is small compared to the uncertainty introduced by the noise in the Stokes and anti-Stokes intensity measurements.

9.4. Effect of Difference in Reference Temperatures

As stated in Section 6, reference sections at two different temperatures are required to estimate the parameters that relate Stokes and anti-Stokes intensity measurements to temperature and only one reference section is needed if its temperature varies sufficiently over time. To test the consequence of calibrating to reference sections with a variation in temperature that is too small, an additional calibration is performed with two reference sections at the same temperature: Cold 1 and Cold 2 (Table 1). 95% of the reference temperatures were between 4.33 and 4.39 °C. Values for the $A(x)$ parameters along the reference sections are correctly estimated, but the parameters γ, $D_F(t)$, and $D_B(t)$ are not. Their estimated values are of the right order of magnitude, and the temperature of the cold-water baths is estimated well with those parameters (not shown). The temperatures outside the 4.33–4.39 °C range are not correctly estimated. The 95% confidence intervals for the estimated temperature outside the reference sections are enormously wide, because of the large parameter uncertainty. The standard uncertainty of the $A(x)$ estimates are the same as in the first calibration configuration, but the standard uncertainty of the γ, $D_F(t)$, and $D_B(t)$ estimates are two orders of magnitude larger than their estimated values. As expected, this calibration configuration failed due to the lack of difference in temperature between the reference sections and the lack of temperature variation of the reference sections.

10. Discussion

10.1. Improved Temperature Estimation for Double-Ended Setups

The estimation of the uncertainty of the temperature from Stokes and anti-Stokes intensity measurements requires several approximations. In this section, the practical implications of these approximations are assessed. Single-ended measurements have several inherent drawbacks compared to double-ended measurements. The differential attenuation is approximated to be constant, step

changes are neglected (e.g., fiber splices, sharp bends), and fiber sections with a different differential attenuation are not accounted for (e.g., coiled up fiber sections, different fiber types). This includes sharp bends in 'bend-tolerant' fiber [14,31]. The step losses in Stokes and anti-Stokes intensity measurements can be corrected manually, if they are identified and the temperature is the same at either side of the step loss [32]. Uncorrected step losses result in a bias in the estimated temperature that is not expressed in the confidence intervals. Commercially available DTS systems usually include an internal reference fiber section for calibration purposes. The use of the internal reference section is not recommended for single-ended measurements because of the step change in the integrated differential attenuation at the connector. The type of the internal fiber is most likely different, with a different differential attenuation than the fiber connected to the DTS system. Hence, it is recommended to use double-ended setups rather than single-ended setups.

The proposed calibration procedure to estimate the integrated differential attenuation in double-ended setups differs from the two-step procedure of van de Giesen et al. [12]. They estimate the differential attenuation integrated between neighboring measurement locations, average them over time for the entire fiber, and sum them to estimate the integrated differential attenuation. In Section 6 of this paper, estimation of the integrated differential attenuation is achieved in one step, and formulated so that it can be used in weighted linear regression.

The presented calibration procedure can lead to better temperature estimates compared to existing calibration procedures. Parameters that are time invariant are kept constant and are estimated from all available data, which improves their estimation. By weighing Stokes and anti-Stokes measurements with their uncertainty, the estimation of parameters is less affected by measurement noise along reference sections with a low signal strength, i.e., at the end of the fiber.

Furthermore, weighted averaging of the temperature from the forward and backward-channel measurements results in better temperature estimates than arithmetic averaging or using either of the individual channel measurements. Consider the synthetic temperature measurement of a double-ended setup shown in Figure 6. The approximated standard uncertainty is computed using the procedure outlined in Section 7.2 and is shown for the forward-channel temperature measurements in orange and for the backward channel in blue. The standard uncertainty of the inverse-variance weighted mean is shown with the solid black line and is much smaller along the entire fiber.
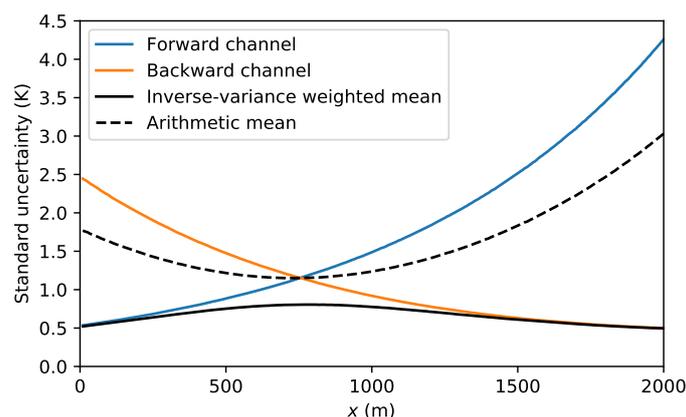

Figure 6. Synthetic example of the standard uncertainty of the estimated temperature using arithmetic mean and the inverse-variance weighted mean.

10.2. Calibration to Reference Sections

The presented calibration procedures require either one or two reference sections with a different temperature. In single-ended setups, two reference sections are needed to differentiate between γ/T_r and $C(t) + x\Delta\alpha$. In double-ended setups, two reference sections are needed to differentiate between γ/T_r and $D_\mathrm{F}(t)$, and between γ/T_r and $D_\mathrm{B}(t)$. Only one reference section is needed in single-ended

and double-ended setups if the reference temperature varies sufficiently over time. To examine the minimum temperature variation that is needed when using one reference section, the analysis of the example of Section 9 is repeated using only the reference section in the ambient water bath, which varies between 12.53 and 12.73 °C. This temperature variation of only 0.2 °C proved to be sufficient for this setup. The contribution of the parameter uncertainty to the standard uncertainty is again small com1pared to the uncertainty introduced by the noise in the Stokes and anti-Stokes intensity measurements. As expected, the likely bias in the estimated temperature of the ambient bath that was apparent in Figure 4 translates to a bias in the estimated temperature (not shown). For example, the temperature estimates along the sections in the cold and warm water baths are 0.03 and 0.10 °C too low, respectively. The minimum temperature difference needed for calibration depends on many factors, including: the matrix solver and its settings, the length of the reference sections, the Stokes and anti-Stokes intensity, and the noise in the Stokes and anti-Stokes intensity measurements.

All existing calibration procedures rely on the temperature of reference sections measured with external sensors. Any deviation of the fiber temperature from the temperature measured with an external sensor introduces errors. Problems that are commonly occurring with DTS calibration are ill-defined positions of the reference sections and non-uniform temperature of the water baths. Both can be discovered at different stages of the calibration procedure. They may introduce a bias in the estimated temperature, which is not accounted for in the estimation of the standard uncertainty and confidence intervals. Therefore, an additional external temperature sensor is recommended for identification of a bias in the reference temperature. A bias may then be identified by, for example, composing a figure of the time-averaged difference between the estimated temperature and the reference temperature, similar to Figure 4. A non-uniform temperature of the reference sections also negatively affects the parameter estimation by introducing an error in the coefficient matrix, in the first column of $\mathbf{X}_m$ in Equations (30) and (37). Investigation of the contribution of the parameter uncertainty to the uncertainty of the estimated temperature is explained in Section 9.3. A non-uniform temperature of the reference sections may also result in an overestimation of the noise in the Stokes and anti-Stokes intensity measurements. The variance of the noise in the intensity measurements are estimated from the residuals between Equations (20) and (21) and the intensity measurements assuming a uniform reference temperature (Section 4). A non-uniform temperature of a reference section increases the residuals, which are incorrectly attributed to the noise from the detector. A comparison of the estimated variance of different reference sections can indicate the location of non-uniform temperature sections. Alternatively, trends in the time-averaged differences between the estimated and reference temperature can also indicate a non-uniform temperature of the reference sections, of which an example is shown with orange markers in Figure 4.

11. Conclusions

A new approach is presented to calibrate temperature from Stokes and anti-Stokes intensity measurements and to provide a confidence interval for the estimated temperature. The uncertainty in the estimated temperature is caused by the noise from the Stokes and anti-Stokes detectors, and the uncertainty in the calibrated parameters that relate the Stokes and anti-Stokes intensity measurements to temperature. Estimation of the confidence interval for the estimated temperature requires an estimation of the distribution of the noise from the Stokes and anti-Stokes detectors and a multi-variate distribution of the parameters that relate the Stokes and anti-Stokes intensity measurements to temperature. All these distributions are propagated with Monte Carlo sampling to approximate the probability density function of the temperature, which is different at each location and varies over time. Various summarizing statistics can be computed from the approximated probability density function, such as standard uncertainties and confidence intervals.

Several improvements were made to existing calibration procedures to reduce the uncertainty in the estimated temperature. The integrated differential attenuation differs per setup and must be experimentally estimated to differentiate it from a shift in temperature. The integrated differential

attenuation for double-ended setups is formulated such that it can be used directly in linear regression and the uncertainty in the parameter estimation is reduced. The parameter estimation is further improved by making use of the fact that some parameters are time-invariant, and to use all available data to estimate their value. The linear regression accounts for spatial and temporal variation in the signal intensity, so that large-intensity measurements have a larger weight in the parameter estimation than the small-intensity measurements.

In double-ended setups, the temperature is estimated from measurements made from both ends of the fiber. Close to the ends of the fiber, the difference in Stokes and anti-Stokes intensity between the forward and backward measurements is large, resulting in a large difference in the uncertainty of the temperature estimated from the forward and backward measurements. The estimated temperature is a weighted average of the temperatures estimated from the forward and backward-channel measurements. Compared to unweighted averaging, this reduces the uncertainty in the estimated temperature. The uncertainty of the temperature that is estimated with the proposed calibration procedure is assessed in an example. The estimated temperature and 95% confidence intervals adequately represent the temperature of the reference sections.

The calibration procedure is implemented in "dtscalibration", an open-source Python package that is freely available under the BSD 3-or-later license from https://github.com/dtscalibration/python-dts-calibration. The package contains the new calibration procedures for both single-ended and double-ended setups, can compute confidence intervals of the estimated temperature, and includes several routines for visualization of the results.

- Dataset: https://doi.org/10.4121/uuid:71b5c3c2-4105-4f4f-bd1e-d7c56732a665
- Dataset license: GPL-3.0-or-later

Author Contributions: Conceptualization, B.d.T., B.S., and M.B.; methodology, B.d.T., B.S., and M.B.; software, B.d.T. and B.S.; validation, B.d.T. and B.S.; formal analysis, B.d.T. and M.B.; investigation, B.d.T. and B.S.; resources, B.d.T., B.S., and M.B.; data curation, B.d.T.; writing—original draft preparation, B.d.T.; writing—review and editing, B.d.T., B.S., and M.B.; visualization, B.d.T.; supervision, M.B.; project administration, B.d.T. and B.S.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Appendix A. Intensity-Dependent Variance of the Noise in the Intensity Measurements

The sources of noise dominating the noise in Stokes and anti-Stokes intensity measurements are discussed in the introductory paragraph of Section 4. For larger setups with more attenuation and DTS systems with a larger avalanche photodiode gain, $\sigma^2_{P_+}$ and $\sigma^2_{P_-}$ may be expressed as a linear function of the intensity,

$$\sigma^2_{P_+} = \beta_+ P_+ + \sigma^2_{P_+,\text{bg}} \tag{A1}$$

$$\sigma^2_{P_-} = \beta_- P_- + \sigma^2_{P_-,\text{bg}} \tag{A2}$$

where $\sigma^2_{P,\text{bg}}$ is the variance of the background noise from the electrical circuit, and β_+ and β_- are the coefficients that relate the intensity to the variance of the noise from the avalanche photodiode.

An algorithm is proposed to obtain β_+, β_-, $\sigma^2_{P_+,\text{bg}}$, and $\sigma^2_{P_-,\text{bg}}$ from the residuals between the fitted product of $G(t)$ and $H(x)$ and the Stokes and anti-Stokes intensity measurements. The residuals

must be known for a wide range of intensities to estimate the unknowns, e.g., with a reference section at the beginning and end of the fiber. First, start with an array with the Stokes intensity measurements and an array with the residuals. The arrays share a temporal and a spatial dimension. Second, sort both arrays by the Stokes intensity measurements. Third, bin both sorted arrays, and compute per bin the mean of the Stokes intensity measurements and the variance of the residuals. Finally, fit Equation (A1) to the two computed values per bin to estimate β_+ and $\sigma^2_{P_+,\text{bg}}$. And repeat the algorithm for the anti-Stokes intensity measurements.

Appendix B. Correlation Stokes and Anti-Stokes Residuals

The variance in the Stokes and anti-Stokes intensity measurements is estimated from measurements of the reference sections used for calibration (Table 1) with the procedure presented in Section 4. The residuals between the Stokes intensity measurements and the fitted curve are plotted against the residuals between the anti-Stokes intensity measurements and the fitted curve in Figure A1. The Pearson correlation coefficient between the two is 0.02. The fitted Normal distribution of the Stokes and anti-Stokes residuals are shown on the top and right axes, respectively. These residuals are used as proxy for the noise from the detector.
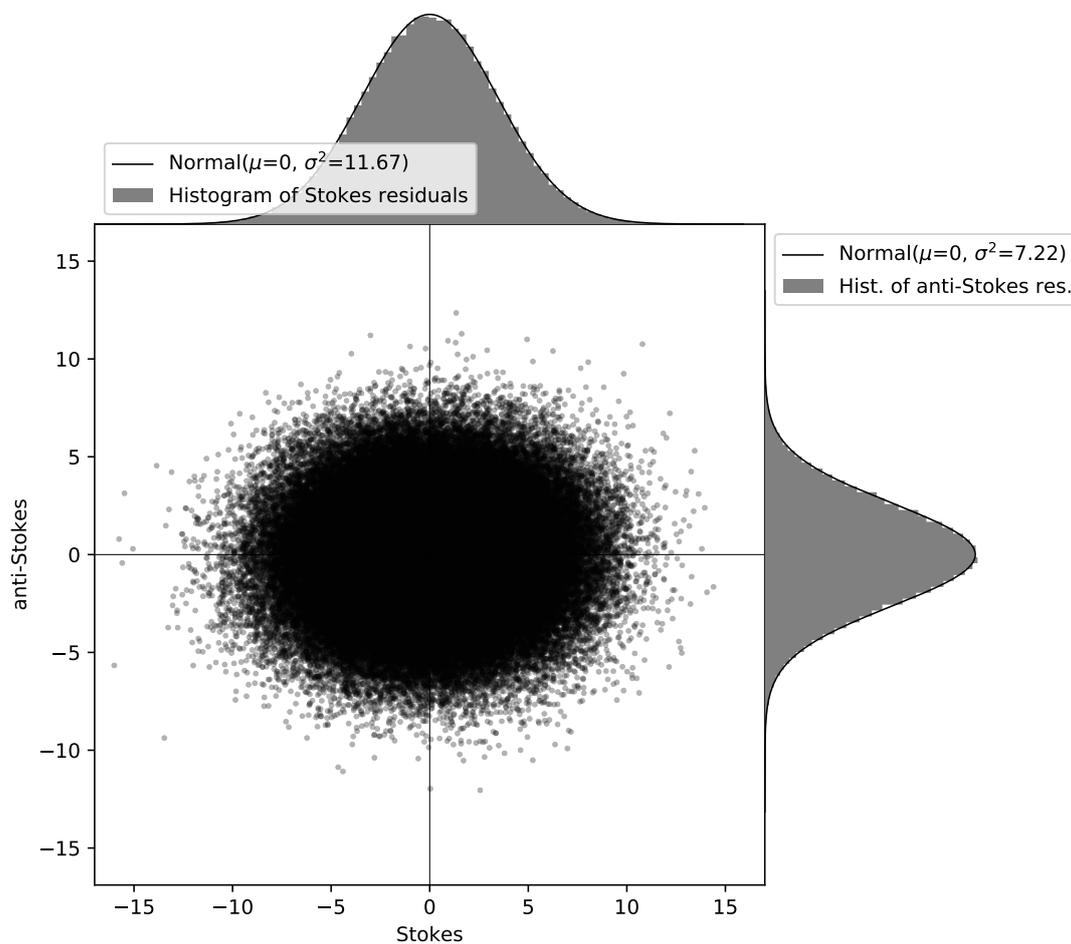

Figure A1. Stokes residuals plotted against the anti-Stokes residuals. The Stokes and anti-Stokes intensity recorded by the DTS instrument have arbitrary units that are linearly related to the power of the scattered signals.

References

1. Selker, J.S.; Thévenaz, L.; Huwald, H.; Mallet, A.; Luxemburg, W.; van de Giesen, N.; Stejskal, M.; Zeman, J.; Westhoff, M.; Parlange, M.B. Distributed fiber-optic temperature sensing for hydrologic systems. *Water Resour. Res.* **2006**, *42*, 1–8. [CrossRef]
2. Tyler, S.W.; Selker, J.S.; Hausner, M.B.; Hatch, C.E.; Torgersen, T.; Thodal, C.E.; Schladow, S.G. Environmental temperature sensing using Raman spectra DTS fiber-optic methods. *Water Resour. Res.* **2009**, *45*, 1–11. [CrossRef]
3. Sayde, C.; Thomas, C.K.; Wagner, J.; Selker, J. High-resolution wind speed measurements using actively heated fiber optics. *Geophys. Res. Lett.* **2015**, *42*, 10064–10073. [CrossRef]
4. Van Ramshorst, J.G.V.; Coenders-Gerrits, M.; Schilperoort, B.; van de Wiel, B.J.H.; Izett, J.G.; Selker, J.S.; Higgins, C.W.; Savenije, H.H.G.; van de Giesen, N.C. Wind speed measurements using distributed fiber optics: A windtunnel study. *Atmos. Meas. Tech. Discuss.* **2019**, *2019*, 1–21. [CrossRef]
5. Euser, T.; Luxemburg, W.M.J.; Everson, C.S.; Mengistu, M.G.; Clulow, A.D.; Bastiaanssen, W.G.M. A new method to measure Bowen ratios using high-resolution vertical dry and wet bulb temperature profiles. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 2021–2032. [CrossRef]
6. Schilperoort, B.; Coenders-Gerrits, M.; Luxemburg, W.; Jiménez Rodríguez, C.; Cisneros Vaca, C.; Savenije, H. Technical note: Using distributed temperature sensing for Bowen ratio evaporation measurements. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 819–830. [CrossRef]
7. Steele-Dunne, S.C.; Rutten, M.M.; Krzeminska, D.M.; Hausner, M.; Tyler, S.W.; Selker, J.; Bogaard, T.A.; van de Giesen, N.C. Feasibility of soil moisture estimation using passive distributed temperature sensing. *Water Resour. Res.* **2010**, *46*, 1–12. [CrossRef]
8. Lowry, C.S.; Walker, J.F.; Hunt, R.J.; Anderson, M.P. Identifying spatial variability of groundwater discharge in a wetland stream using a distributed temperature sensor. *Water Resour. Res.* **2007**, *43*, 1–9. [CrossRef]
9. Bakker, M.; Caljé, R.; Schaars, F.; van der Made, K.J.; de Haas, S. An active heat tracer experiment to determine groundwater velocities using fiber optic cables installed with direct push equipment. *Water Resour. Res.* **2015**, *51*, 2760–2772. [CrossRef]
10. Bense, V.F.; Read, T.; Bour, O.; Le Borgne, T.; Coleman, T.; Krause, S.; Chalari, A.; Mondanos, M.; Ciocca, F.; Selker, J.S. Distributed Temperature Sensing as a downhole tool in hydrogeology. *Water Resour. Res.* **2016**, *52*, 9259–9273. [CrossRef]
11. Hausner, M.B.; Suárez, F.; Glander, K.E.; Giesen, N.V.d.; Selker, J.S.; Tyler, S.W. Calibrating Single-Ended Fiber-Optic Raman Spectra Distributed Temperature Sensing Data. *Sensors* **2011**, *11*, 10859–10879. [CrossRef] [PubMed]
12. Van de Giesen, N.; Steele-Dunne, S.C.; Jansen, J.; Hoes, O.; Hausner, M.B.; Tyler, S.; Selker, J. Double-ended calibration of fiber-optic raman spectra distributed temperature sensing data. *Sensors* **2012**, *12*, 5471–5485. [CrossRef] [PubMed]
13. Krause, S.; Blume, T. Impact of seasonal variability and monitoring mode on the adequacy of fiber-optic distributed temperature sensing at aquifer-river interfaces. *Water Resour. Res.* **2013**, *49*, 2408–2423. [CrossRef]
14. Hilgersom, K.; van Emmerik, T.; Solcerova, A.; Berghuijs, W.; Selker, J.; van de Giesen, N. Practical considerations for enhanced-resolution coil-wrapped distributed temperature sensing. *Geosci. Instrum. Methods Data Syst.* **2016**, *5*, 151–162. [CrossRef]
15. McDaniel, A.; Tinjum, J.M.; Hart, D.J.; Fratta, D. Dynamic Calibration for Permanent Distributed Temperature Sensing Networks. *IEEE Sens. J.* **2018**, *18*, 2342–2352. [CrossRef]
16. Hartog, A.H. *An Introduction to Distributed Optical Fibre Sensors*; CRC Press: Boca Raton, FL, USA, 2017.
17. Eriksrud, M.; Mickelson, A. Application of the backscattering technique to the determination of parameter fluctuations in multimode optical fibers. *IEEE J. Quantum Electron.* **1982**, *18*, 1478–1483. [CrossRef]
18. Fukuzawa, T.; Shida, H.; Oishi, K.; Takeuchi, N.; Adachi, S. Performance improvements in Raman distributed temperature sensor. *Photonic Sens.* **2013**, *3*, 314–319. [CrossRef]
19. Simon, N.; Bour, O.; Lavenant, N.; Porel, G.; Nauleau, B.; Pouladi, B.; Longuevergne, L. A Comparison of Different Methods to Estimate the Effective Spatial Resolution of FO-DTS Measurements Achieved during Sandbox Experiments. *Sensors* **2020**, *20*, 570. [CrossRef]
20. Bolognini, G.; Hartog, A. Raman-based fibre sensors: Trends and applications. *Opt. Fiber Technol.* **2013**, *19*, 678–688, doi:10.1016/j.yofte.2013.08.003. [CrossRef]

21. Davey, S.; Williams, D.; Ainslie, B.; Rothwell, W.; Wakefield, B. Optical gain spectrum of GeO$_2$-SiO$_2$ Raman fibre amplifiers. *IEE Proc. J Optoelectron.* **1989**, *136*, 301–306. [CrossRef]

22. Richter, P. Estimating Errors in Least-Squares Fitting. *NASA Telecommun. Data Acquis. Prog. Rep.* **1995**, *42*, 107–137.

23. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272, doi:10.1038/s41592-019-0686-2. [CrossRef] [PubMed]

24. Ku, H.H. Notes on the use of propagation of error formulas. *J. Res. Natl. Bur. Stand.* **1966**, *70*, 263–273

25. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28–30 June 2010.

26. Joint Committee for Guides in Metrology. *JCGM 100: Evaluation of Measurement Data—Guide for the Expression of Uncertainty in Measurement (GUM)*; Technical Report; JCGM: Paris, France, 2008.

27. Joint Committee for Guides in Metrology. *JCGM 101: Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement (GUM)"—Propagation of Distributions Using a Monte Carlo Method*; Technical Report; JCGM: Paris, France, 2008.

28. Des Tombe, B.F.; Schilperoort, B. Dtscalibration Python Package for Calibrating Distributed Temperature Sensing Measurements, 2020, v0.7.4. Available online: https://zenodo.org/record/3627843#.XpUvaJkRWUk (accessed on 14 April 2020).

29. Hoyer, S.; Hamman, J. xarray: ND labeled Arrays and Datasets in Python. *J. Open Res. Softw.* **2017**, *5*, 1–6. [CrossRef]

30. Rocklin, M. Dask: Parallel Computation with Blocked algorithms and Task Scheduling. In Proceedings of the 14th Python in Science Conference, Austin, TX, USA, 6–12 July 2015.

31. Remouche, M.; Georges, F.; Meyrueis, P. Flexible Optical Waveguide Bent Loss Attenuation Effects Analysis and Modeling Application to an Intrinsic Optical Fiber Temperature Sensor. *Opt. Photonics J.* **2012**, *2*, 1–7. [CrossRef]

32. Hausner, M.B.; Kobs, S. Identifying and correcting step losses in single-ended fiber-optic distributed temperature sensing data. *J. Sens.* **2016**, *2016*, 1–10. [CrossRef]