Measurement 92 (2016) 340–351

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

An online two-stage adaptive algorithm for strain profile estimation from noisy and abruptly changing BOTDR data and application to underground mines

G. Soto [a,*], J. Fontbona [a], R. Cortez [a], L. Mujica [b]

[a] Center for Mathematical Modeling (CMM), UMI (2807), UCHILE-CNRS, Chile
[b] Mining Information Communications and Monitoring (Micomo), Fundación Chile, Chile

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form 13 May 2016
Accepted 13 June 2016
Available online 16 June 2016

Keywords:
BOTDR
Brillouin gain spectrum
Strain
Time series
Smoothing
Outliers
Abrupt changes

ABSTRACT

Strain measurement using BOTDR (Brillouin Optical Time-Domain Reflectometry) is nowadays a standard tool for structural health monitoring. In this context, weak data quality and noise, usually owed to defective fiber installation, hinders discriminating actual level shifts from outliers and might entail a biased risk assessment. We propose a novel online adaptive algorithm for strain profile estimation in strain time series with abrupt and gradual changes and missing data. It relies on a convolution filter in Brillouin spectrum domain and a smoothing technique in time domain. In simulated data, the convolution filter is shown to reduce strain measurement uncertainty by up to 8 times the strain resolution. The two-stage method is illustrated with systematic outliers removal from real data of a Chilean copper mine and the improvement of the associated gain spectrum quality by up to 18 dB in SNR terms.

1. Introduction

Distributed fiber sensors technology based on Brillouin scattering has attracted much attention for remarkable features such as its high precision, high spatial resolution and long measuring range [1]. Structural deformation monitoring based on Brillouin Optical Time-Domain Reflectometry (BOTDR) allows the measurement of the longitudinal strain applied along an optical fiber attached to a structure, by means of the Brillouin gain spectrum [2]. If the structure suffers a deformation, a change in the shape of the spectrum measured by the BOTDR system is observed. The behavior of the structure subjected to strength can then be assessed by means of suitable physical and statistical models. Among other uses, this technique has been applied to civil structural monitoring [3,4], roadworks [5,6], and mining [7–9].

It is well known that, when an optical pulse is launched into an optical fiber, some backscattered signals come back to the input end. The BOTDR system measures the power distribution in frequency of the backscattered light at each point along to the optical fiber. This distribution, called Brillouin gain spectrum (BGS), can be seen as a set of Lorentzian-like curves along the distance coordinate. When a longitudinal strain ϵ is applied to the fiber, and as a result of the non-linear interaction between the incident light and acoustic phonons of the crystal lattice of the semiconductor material, the backscattered light is shifted in frequency by an amount ν_B, called Brillouin shift [10,11]. This applied longitudinal strain ϵ is related to the value of the frequency shift [12,13] through the linear relationship $\nu_B(\epsilon) = \nu_B(0)(1 + C\epsilon)$, where $\nu_B(\epsilon)$ stands for the Brillouin frequency shift after strain has been applied, $\nu_B(0)$ is a referential Brillouin frequency shift and C is the proportional coefficient of strain [14].

The frequency shift is standardly estimated by fitting a Lorentzian function to the spectrum at each point of the optical fiber and determining its mode or central value. This fitting procedure is usually done by means of suitable least squares minimization algorithms [15]. This method presents an important drawback: the accuracy of the estimation of the strain-BOTDR can be considerably affected by measurement noise, which can bias the estimation of ν_B and result in multiple outliers in the strain time series. The presence of outliers could prevent from correctly discriminating whether a true level shift has occurred or not, and mislead the instantaneous analysis of time series. Consequences of this could in some cases be serious, especially if the sampling rate of the online strain-BOTDR monitoring is low, since the confirmation or rectification of the measured level shift by means of new measurements

* Corresponding author.
E-mail address: gsoto@dim.uchile.cl (G. Soto).

http://dx.doi.org/10.1016/j.measurement.2016.06.022
0263-2241/

could arrive too late. Therefore, there is strong a necessity for a robust procedure that estimates strain profile consistently over time and along the fiber.

To address this problem, this paper proposes an adaptive algorithm for strain profile estimation from low-quality BOTDR data, in the context of an online procedure. Our methodology is suitable for forecasting the statistical behavior of Brillouin gain spectra with both abrupt and gradual changes over time, and is robust against outliers.

Spectral noise can be seen as a stochastic process following some trivariate underlying joint distribution that depends on two explanatory variables of the BGS, namely the distance x and the frequency ν, and on a time variable t inherent to the time series. We are therefore faced a 3-dimensional estimation problem, which we will address by means of the simultaneous use of two techniques. The first technique consists in a convolution filter on the (x,ν) domain whose main feature is the ability to keep the value of ν_B unchanged at the observed spectral distribution. The second one is an exponential smoothing technique for temporal noise reduction, using a time series model consisting of a linear trend with an additive noise term. In addition, our temporal estimation technique is robust enough to deal with missing data and abrupt changes in the level, a situation often found in practice which affects the performance of standard filtering techniques.

Although some of the aforementioned papers deal with some statistical aspects of Brioullin shift estimation, to our knowledge the only available method which also addresses the problem of global consistent space–time estimation, is the procedure based on 3d Non Local Means (NLM) presented in the paper [16] (which appeared while the present work was in its revision stage). Comparative discussions will be given later at several places, in particular when we will asses the performance and computational cost of our method.

The remainder of this paper is organized as follows. In Section 2 we discuss the effect of noise on the BGS and the methodology proposed to reduce it. We describe the observation data model in Section 3. In Section 4 we propose a pseudo two-dimensional convolution filter to estimate the underlying (unnoisy) spectral distribution in the distance-frequency domain. Its performance is assessed in Section 5 both theoretically and by means of Monte-Carlo simulations. The temporal filtering method based on exponential smoothing is described in Section 6. Section 7 presents the algorithm combining the two previous stages. The algorithm is applied on real data from El Teniente mine (Codelco, Chile) in Section 8 and the obtained results are then discussed. The computational cost of our method in a general sensing framework is discussed in Section 9. Finally, Section 10 brings the conclusions of the paper.

2. The problem of measurement-noise in Brillouin gain spectrum

The intrinsic variations of the strain time series, in combination with suitable forecasting models, are commonly used for structural health monitoring or assessment of structural instabilities. However, the accuracy of these analysis heavily depend on the quality of the measured Brillouin gain spectrum (BGS) which often contains noise. The BOTDR system records a sequence of spectra with a given sampling rate τ, measured in hours. The observed BGS can be seen as a frame of this sequence. A typical noisy BGS obtained by the BOTDR system is shown in Fig. 1, which corresponds to a frame of a sequence of real data taken from a Chilean underground copper mine (a detailed description of this data and the BOTDR system setting used to obtain it will be given in Section 5.3).

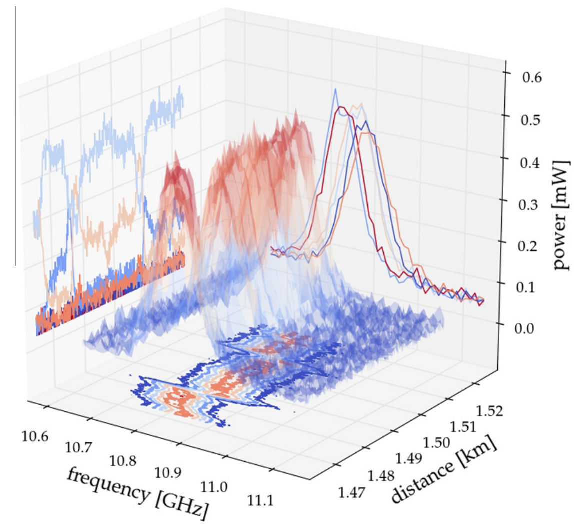

Fig. 1. An observed Brillouin gain spectrum frame, taken from of a sequence of real data registered at El Teniente (Codelco) mine in Chile during 2008. The data description and the BOTDR system setting is detailed in Section 5.3.

As mentioned in the introduction, if the spectral noise variance is large enough, outliers owed to a biased Brillouin frequency shift estimation can appear in the strain time series, which could prevent the correct detection of true level shifts and entail a biased risk assessment, with potentially serious consequences. This motivates the development of statistical tools that simultaneously reduce the effects of the spectral noise and provide coherent observation sequences in presence of temporally varying field conditions.

The approach we propose to reduce the spectral noise is based on separating the problem in two stages, developed in the following sections: (1) applying a filtering technique in the distance-frequency coordinate and (2) applying then a temporal filtering technique. Firstly, the underlying BGS is estimated by applying a pseudo 2-dimensional convolution filter designed in the distance-frequency domain, and then, a temporal smoothing technique is carried out using a locally linear regression model, which maintains over time the coherence of the distance-frequency coordinates of the filtered data.

3. Observation model

We next introduce a simple model for the evolution of the observed variations of the BGS. Let $\mathcal{S}$ and $\mathcal{T}$ be two sets such that $\mathcal{S}\subset\mathbb{R}^2$ and $\mathcal{T}\subset\mathbb{R}$. We consider the observed BGS as a noisy 2-dimensional sequence $M:\mathcal{S}\times\mathcal{T}\rightarrow\mathbb{R}$ defined as

$$M(\mathbf{s},t)=B(\mathbf{s},t)+\epsilon(\mathbf{s},t),\qquad \mathbf{s}\in\mathcal{S},\ t\in\mathcal{T}, \tag{1}$$

where $B(\mathbf{s},t)$ is the original BGS, $(\mathbf{s},t)$ represents the coordinate in the $\mathcal{S}\times\mathcal{T}$ domain, with $\mathbf{s}=(x,\nu)$, and $\varepsilon(\mathbf{s},t)\sim f_\varepsilon$ is an additive noise that follows some trivariate distribution.

Notice we have not yet given a regression model for the temporal behavior of the rock mass itself, which is represented by the sequence $B(\mathbf{s},t)$ and can be specified independently. Such a model will be presented below.

According to the model, the aim of the proposed filtering method is to provide an estimate $\widetilde{B}$ of the original BGS, B, from the observed data, M, separating the task into two fundamental stages: a 2-dimensional filtering technique in the $\mathcal{S}$-domain and a smoothing technique for time series in the $\mathcal{T}$-domain.

4. Filtering in the $\mathcal{S}$-domain

4.1. The problem statement

The problem of the spectral estimation at a fixed time t, can be stated as follows. Let $\mathbf{s} = (x, \nu)$ be the distance-frequency coordinate. The observation model defined in (1) can be rewritten as

$$m(x,\nu) = b(x,\nu) + \epsilon(x,\nu), \qquad x \in \mathcal{X}, \ \nu \in \mathcal{V}, \tag{2}$$

with $\mathcal{S} = \mathcal{X} \times \mathcal{V} \subset \mathbb{R}^2$, where $m(x,\nu)$ is a frame of the sequence $M(\mathbf{s}, t)$ corresponding to the observed BGS measured by the BOTDR system at a fixed time t, and the additive noise $\epsilon \sim f_\epsilon$ follows some bivariate distribution.

In addition to outliers removal, when estimating the original spectrum b from a noisy observation m by applying a 2-dimensional convolution filter, the method we propose will keep unchanged two fundamental features of the data, namely (a) the central frequency ν_d of the Lorentzian distribution at each point x along to the optical fiber, and (b) the spatial resolution of the BOTDR system.

4.2. The convolution filter

It is known that the implementation of a classical low-pass filtering technique is highly affected by the tradeoff between noise reduction and bias. The choice of a suitable value for the smoothing parameter plays a rather crucial role. A too large value can excessively blur some areas of high curvature, which results in a large bias. This means that, if we applied a filter with this feature on the distance domain of a noisy BGS, we then reduce the spatial resolution of the BOTDR system, and consequently, the estimate of strain profile will be biased.

In order to prevent a large bias in the distance-frequency domain, the proposed filter $h(x,\nu;\Gamma_h)$ consists in a Lorentzian function, centered at 0, defined as

$$h(x,\nu;\Gamma_h) = \frac{g_h}{\pi} \frac{\Gamma_h}{\nu^2 + \Gamma_h^2} \tag{3}$$

where g_h is the gain of the filter and Γ_h is a smoothing parameter that controls the filtering quality. This parameter is also called Full Width at Half Maximum (FWHM). The filter (3) is applied in each position x along the fiber, as a 1-dimensional convolution between the (noisy) observed Brillouin distribution and this filter. In other words, we only take into account the frequency domain for denoising this 2-dimensional data, and this is the reason why the proposed filter is referred as a "pseudo 2-dimensional technique". In this manner, our approach allows for reducing the spectral noise without affecting the spatial resolution of the system, since the filter does not smooth areas of high curvature in the distance domain of the BGS.

On the other hand, since the accuracy of the central frequency estimation by Lorentzian fitting depends on how noisy and deformed the spectral distribution is, we have chosen this filter because of a fundamental fact: the convolution between two Lorentzian curves results in a curve with Lorentzian distribution.

Indeed, suppose that the (unknown) original spectrum $b(x,\nu)$ in model (2) can be represented by a family of Lorentzian distributions with local central frequency $\nu_b(x)$, as follows:

$$b(x,\nu) = \frac{g_b(x)}{\pi} \frac{\Gamma_b(x)}{(\nu - \nu_b(x))^2 + \Gamma_b^2(x)} \tag{4}$$

where $\Gamma_b(x)$ and $g_b(x)$ stand for the FWHM and the maximum value of the original spectrum at the distance x along the fiber. We define $\tilde{b}(x,\nu)$, the estimate of the original BGS $b(x,\nu)$ at instant t, in the following way

$$\tilde{b}(x,\nu) = m(x,\nu) * h(x,\nu;\Gamma_h), \tag{5}$$

where $*$ is the convolution operator. Then, applying (3) in (2) according to relationships (5) and (4), we obtain that

$$\tilde{b}(x,\nu) = \beta(x,\nu;\boldsymbol{\theta}_x) + \eta(x,\nu) \tag{6}$$

$$\beta(x,\nu;\boldsymbol{\theta}_x) = \frac{g_\beta(x)}{\pi} \frac{\Gamma_\beta(x)}{(\nu - \nu_\beta(x))^2 + \Gamma_\beta^2(x)}, \tag{7}$$

$$\eta(x,\nu) = h(x,\nu;\Gamma_h) * \epsilon(x,\nu) \tag{8}$$

with

$$\boldsymbol{\theta}_x = [\nu_\beta(x), \Gamma_\beta(x), g_\beta(x)] \tag{9}$$

$$\nu_\beta(x) = \nu_b(x) \tag{10}$$

$$g_\beta(x) = g_h g_b(x) \tag{11}$$

$$\Gamma_\beta(x) = \Gamma_b(x) + \Gamma_h, \tag{12}$$

where $\boldsymbol{\theta}_x$, $\nu_\beta(x)$, $\Gamma_\beta(x)$, and $g_\beta(x)$ stand for the estimated parameter vector, the central frequency, the FWHM and the maximum value of the filtered spectrum $\tilde{b}(x,\nu)$ at the distance x, respectively. From (6) and (7), the estimated spectrum is again given by a Lorentzian distribution plus some noise. Therefore, the characteristic shape of the estimated spectrum does not suffer from meaningful changes. Most importantly, we have $\nu_\beta(x) = \nu_b(x)$, that is, the central frequency value of the Lorentzian curve is kept unchanged and, consequently, the strain associated with the Brillouin frequency shift too. In Fig. 2, we show the result of applying the filter $h(x,\nu;\Gamma_h)$ with $\Gamma_h = 20$ MHz to the spectrum of Fig. 1.

Furthermore, the noise level has been reduced by the convolution $h(x,\nu;\Gamma_h) * \epsilon(x,\nu)$, where the filtering performance is being controlled by the parameter Γ_h. We also observe an increase of the FWHM $\Gamma_\beta(x)$ when compared to the value $\Gamma_b(x)$ of the measured BGS, but this result does not affect, theoretically at least, the estimation of the frequency shift. In Ref. [17], by using different pump pulse widths, experimental results about the impact of the FWHM on the frequency error are shown. Errors do not exceed 1 MHz (10^{-3} GHz), a relatively small value compared to typical values of the frequency shifts, which are the order of 10 GHz.

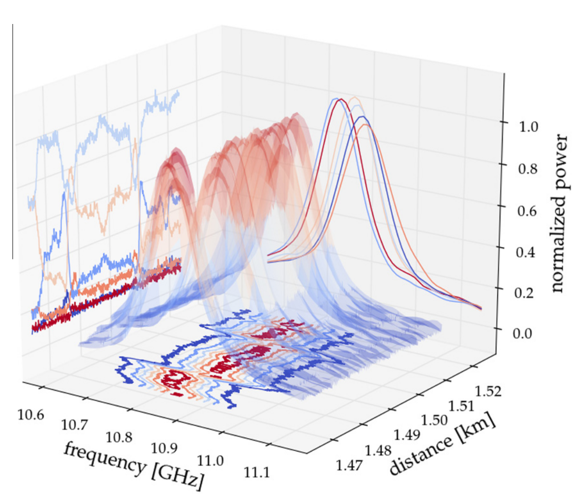

Fig. 2. A filtered Brillouin gain spectrum, resulting from applying the convolution filter with $\Gamma_h = 20$ MHz on the spectrum of Fig. 1.

5. The filtering performance

5.1. Signal-to-noise ratio

Signal-to-noise ratio (SNR) is a good measure of the filtering performance, since it describes how much a signal has been corrupted by noise. Let P_S and P_N be the discrete average power of the Brillouin gain spectrum and the spectral noise, respectively. Then, in decibels (dB), the SNR is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{P_S}{P_N}\right). \tag{13}$$

Note that, if the noise is zero mean, the power of noise is defined by only its variance. In this manner, the SNR can written as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{P_S}{\sigma^2}\right), \tag{14}$$

where σ^2 stands for the noise variance.

This definition will be used later in the assessment of frequency shift estimators respect to sensitivity to noise, and consequently, respect to the impact of noise effects on deformation measurements.

5.2. Estimator performance using synthetic data

It is important to determine quantitatively the magnitude of the filter effect, in order to be able to reduce the possible effects of the spectral noise, such as outliers. The Cramér-Rao theorem (see [18] for background) provides a strict lower bound for the variance of a quantity which is estimated from a set of noisy measurements, and can be applied to determine the (theoretical) minimum uncertainty in the determination of the Brillouin Frequency Shift or deformation.

5.2.1. Cramér-Rao lower bound

Let us assume that a noisy spectrum with central frequency θ was generated by adding to it a Gaussian zero mean noise of variance σ^2. Thus, the Brillouin spectrum can be written as

$$m_i = b_i(\theta) + \sigma\epsilon_i, \quad i = 1,\ldots,n \tag{15}$$

with $\epsilon_i \sim \mathcal{N}(0,1)$, where

$$b_i(\theta) = \frac{g_b}{\pi}\frac{\Gamma_b}{(\nu_i-\theta)^2+\Gamma_b^2}, \tag{16}$$

is the unnoisy Brillouin spectrum, that is, the spectrum without effects induced by the BOTDR system. Then, if we assume that σ, g_b and Γ_b are known, it can be shown that the variance of any unbiased maximum likelihood estimator $\hat{\theta}$ is

$$var(\hat{\theta}) \geqslant \sigma_{CR}^2 \tag{17}$$

where σ_{CR}^2 is called the Cramér-Rao Lower Bound (CRLB) (see [18]), with

$$\sigma_{CR} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}\left(\frac{db_i(\theta)}{d\theta}\right)^2}}. \tag{18}$$

It is worth noting that if $b_i(\theta)$ is the identity function ($b_i(\theta) = \theta$), we have that $var(\hat{\theta}) \geqslant \sigma^2/n$, as might be expected. Now, computing the derivate of $b_i(\theta)$ respect to θ, we finally have

$$\sigma_{CR} = \frac{\pi}{2g_b\Gamma_b\sqrt{\sum_{i=1}^{n}\left(\frac{\nu_i-\theta}{(\nu_i-\theta)^2+\Gamma_b^2}\right)^2}}\sigma. \tag{19}$$

This theoretical bound will be used as a benchmark for assessing the impact of the noise on deformation measurements.

5.2.2. Monte Carlo simulation

A simulation study was conducted to compare the performance of the Lorentzian curve fitting respect to the sensitivity to noise, when previously using, or not, a convolution filter for denoising. The performance is assessed by comparing the Root Mean Squared Error (RMSE) of the estimate $\hat{\theta}$, whose value corresponds to the standard deviation of an unbiased estimator, with the CRLB, that is, the theoretical bound for the standard deviation given by Eq. (19). RMSE is computed as the expected value of the mean squared difference between the estimated values $\hat{\theta}$ and the underlying true value ν_o, using synthetic data $\{m_i : i = 1,\ldots,n\}$ as follows

$$\mathrm{RMSE}(\hat{\theta}) = \left(\mathbb{E}[(\nu_o-\hat{\theta})^2]\right)^{1/2}. \tag{20}$$

The expected value $\mathbb{E}$ is approximated by the Monte Carlo estimate

$$\mathbb{E}[(\nu_o-\hat{\theta})^2] = \frac{1}{R}\sum_{r=1}^{R}(\nu_o-\hat{\theta}_r)^2, \tag{21}$$

where $\hat{\theta}_r$ stands for the rth estimate of true value $\theta = \nu_o$.

Let $\eta_i(\theta)$ be the fitting error, defined as follows

$$\eta_i(\theta) = m_i * h_i(\Gamma_h) - \beta_i(\theta), \tag{22}$$

where m_i stands for the noisy spectrum, h_i stands for the convolution filter defined as

$$h_i(\Gamma_h) = \frac{g_h}{\pi}\frac{\Gamma_h}{(\nu_i-\theta)^2+\Gamma_h^2}. \tag{23}$$

and $\beta_i(\theta)$ is the spectrum model defined as

$$\beta_i(\theta) = \frac{g_h g_b}{\pi}\frac{\Gamma_b+\Gamma_h}{(\nu_i-\theta)^2+(\Gamma_b+\Gamma_h)^2}. \tag{24}$$

Since we assume that the noise is governed by a Normal distribution, the maximum likelihood estimation of θ can be estimated by non-linear least squares as follows

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\sum_{i=1}^{n}\eta_i^2(\theta). \tag{25}$$

Simulation were repeated $R = 1000$ times to make results independent of any particular observation. Noisy spectra are generated using the model given by Eq. (15) using different values of noise variance σ^2. From Eq. (14), and knowing that the power of $\beta_i(\nu_o)$ is

$$P_S = \frac{1}{n}\sum_{i=1}^{n}\beta_i^2(\nu_o), \tag{26}$$

the noise variance is computed as

$$\sigma^2 = P_S 10^{-\frac{\mathrm{SNR}}{10}}. \tag{27}$$

Fixed parameters were set to $g_b = g_h = 1$ by simplicity, and $\Gamma_b = 25$ MHz, according to the BOTDR system setting. The scan interval of simulated spectra was set to 10 MHz in a range of frequencies from 10.62 to 11.11 GHz. In Fig. 3 the result of computing the CRLB on this range of frequencies is shown. Aside from sequences close to the upper or lower bounds of the range, the rest has the same value of CRLB. Therefore, we take a value from that range for simulations. In this manner, we choose $\nu_o = 10.85$ GHz as the value of the (true) Brillouin frequency shift to estimate, because this value belongs to the range of minimum CRLB and it also corresponds to the case when the fiber does not have deformation, according to the BOTDR system setting.

Fig. 4 shows results of the simulation in which the accuracy of the unfiltered and pre-filtered estimator are only comparable for SNR values greater than 2 dB, practically all estimators attaining the CRLB after 5 dB. Although the performance of pre-filtered

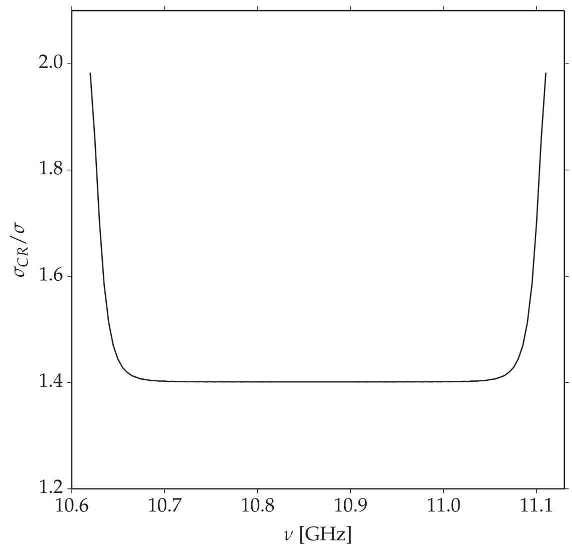

Fig. 3. Cramér-Rao lower bound for $\nu \in [10.62, 11.11]$ GHz.

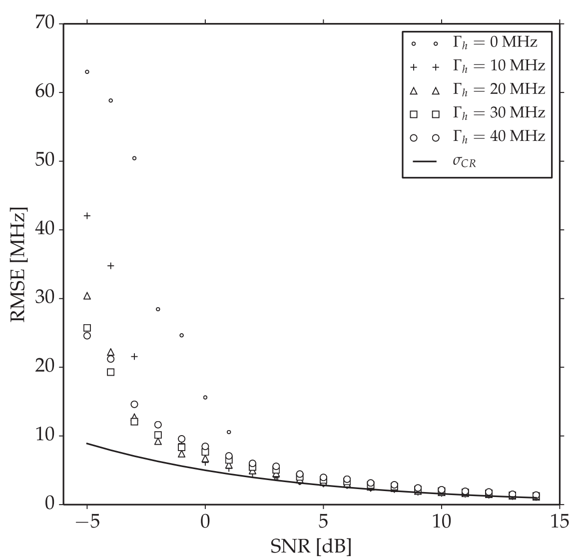

Fig. 4. RMSE curves of the estimated frequency shift respect to the sensitivity to noise using different values of Γ_h. $\Gamma_h = 0$ MHz indicates that the noisy spectra was not filtered.

estimators are marginally worst than the unfiltered estimator for SNR values greater than 2 dB, standard errors are still relatively low. As we indicated above, when the spectral noise variance is large enough, the estimate of ν_o can be biased. This can be seen in the way that RMSE curves move away from the CRLB, for SNR values less than 2 dB, and increasing the bias when SNR values decrease. Pre-filtered estimators yield a RMSE value smaller than 10 MHz for low-SNR values (SNR = -1 dB), while the unfiltered version has more than 20 MHz. In addition, the standard error for SNR = -5 dB is approximately 25 MHz, when data is pre-filtered using $\Gamma_h = 40$ MHz, significantly smaller than the unfiltered estimator whose error reaches over 60 MHz, six times the interval scan. It seems that the accuracy of the estimate of ν_o is affected by the numerical difficulty of the optimization algorithm to find the true optimum value when the underlying regression function $b_i(\theta)$ and noisy data with large variance are far from matching each other. But we have already shown that the standard error can be reduced by means of a previous filtering applied to the noisy spectrum, without fixing the algorithm.

As regards the outliers problem, we have said that these can appear in the strain time series due to a biased Brillouin frequency shift estimation. In order to asses the impact of a biased estimation of $\hat{\theta}$ on the estimated strain $\hat{\epsilon}$, we use also the RMSE. Recalling that the applied longitudinal strain ϵ is related to the value of the frequency shift $\nu_B(\epsilon)$ through the linear relationship $\nu_B(\epsilon) = \nu_B(0)(1 + C\epsilon)$, where $\nu_B(0)$ is a referential Brillouin frequency shift and C is the proportional coefficient of strain, we have

$$\mathrm{RMSE}(\hat{\epsilon}) = \frac{100}{C\nu_B(0)}\mathrm{RMSE}(\hat{\theta}), \tag{28}$$

where $\mathrm{RMSE}(\hat{\epsilon})$ stands for the Root Mean Square Error of the estimated strain, both measured in %, with $C = 4.46$ and $\nu_B(0) = 10.85$ GHz, according with the BOTDR system setting. The scan interval of simulated strain covers a range of values from -0.48% to 0.54%, with a resolution of 0.02%.

Fig. 5 shows the impact of a biased estimation of $\hat{\theta}$ on the estimated strain $\hat{\epsilon}$, whose true value is $\epsilon = 0$%. It is worth noting that the standard error of the unfiltered estimator reaches approximately 0.13%, when SNR = -5 dB. This means that the confidence interval for $\epsilon = 0$% can be $[-0.26, 0.26]$%, if we use a significance level of 95%. This level of uncertainty is a basic characteristic of what constitutes an outlier due to a biased frequency shift estimation. On other hand, the standard error for SNR = -5 dB is approximately 0.05% when data is pre-filtered using $\Gamma_h = 40$ MHz, the corresponding confidence interval significance level being of 95% is $[-0.10, 0.10]$%, that is, smaller than the unfiltered estimator by 0.16%. The accuracy of the estimated strain is thus improved in eight times its resolution.

It is important to bear in mind that the intensity of the light propagated within the fiber tends to be more attenuated the longer the fiber is and, therefore, SNR values tend to decrease along the fiber. Thus, the fact that our method reduces the uncertainty of the strain estimate in low SNR environments, makes in principle the pre-filtered estimator suitable for long fiber configurations, including BOTDR sensing systems of lengths of tens of kilometers commonly found in industrial applications.

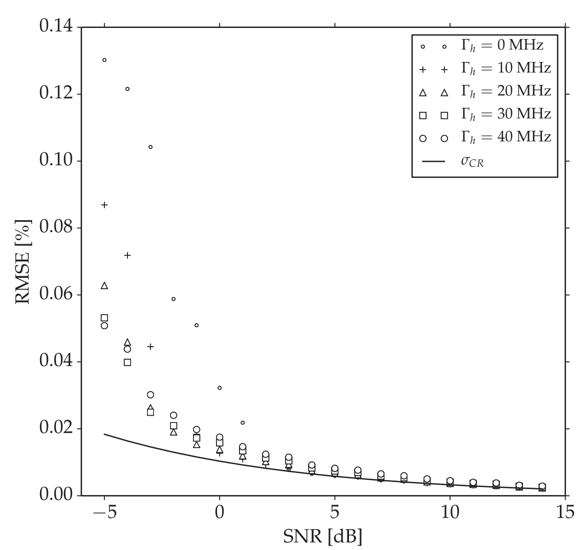

Fig. 5. RMSE curves of the estimated strain respect to the sensitivity to noise using different values of Γ_h. $\Gamma_h = 0$ MHz indicates that the noisy spectra was not filtered.

5.3. Filtering performance using real data

We next assess the estimator when the convolution filter is applied to a real time BGS sequence, obtained in a fiber section installed inside El Teniente mine of Chilean state mining company Codelco.

These measurements were carried out with a sampling rate of $\tau = 4$ h. In total there are 165 frames, which cover 728 h ($\sim$ 30 days) of regular samples, with some missing data. The fiber under analysis is 4.94 m long, and a sampling rate of 10 cm was used, making a total of 494 samples. The frequency resolution of the BOTDR system was set to 10 MHz in a range of frequencies from 10.62 to 11.11 GHz, which means we have 50 samples. Through the linear relationship between strain and frequency shift, the BOTDR system is able to measure deformation in intervals of 0.02%, from −0.48% to 0.54%, with a spatial resolution of 1 m, under this setting.

Unlike the previous assessment, SNR will be used as an indicator of improved data quality as well as of the stability of this improvement over time.

Let P_S and P_N be the discrete average power of the Brillouin gain spectrum and the spectral noise, respectively. These are computed as follows:

$$P_S = \frac{1}{|\mathcal{X}||\mathcal{V}|}\sum_{x\in\mathcal{X}}\sum_{\nu\in\mathcal{V}}\{\beta(x,\nu)\}^2, \tag{29}$$

$$P_N = \frac{1}{|\mathcal{X}||\mathcal{V}|}\sum_{x\in\mathcal{X}}\sum_{\nu\in\mathcal{V}}\{\eta(x,\nu)\}^2. \tag{30}$$

Since values of P_S and P_N are unknown, we will estimate them via Lorentzian-curve fitting using non-linear least squares.

Let $\eta(x,\nu;\boldsymbol{\theta}_x)$ be the fitting error, whose expression is defined as follows

$$\eta(x,\nu;\boldsymbol{\theta}_x) = \tilde{b}(x,\nu) - \beta(x,\nu;\boldsymbol{\theta}_x), \tag{31}$$

where $\tilde{b}(x,\nu)$ stands for the measured spectrum, $\beta(x,\nu;\boldsymbol{\theta}_x)$ is the spectrum model defined in (7) and $\boldsymbol{\theta}_x$ is the parameter model at distance x. If we assume that the noise is governed by a Normal distribution, the maximum likelihood estimation of $\boldsymbol{\theta}_x$ can be estimated by non-linear least squares as follows

$$\boldsymbol{\theta}_x^{LF} = \arg\min_{\boldsymbol{\theta}_x\in\Theta}\sum_{\nu\in\mathcal{V}}\{\eta(x,\nu;\boldsymbol{\theta}_x)\}^2, \tag{32}$$

where $\boldsymbol{\theta}_x^{LF}$ is the estimate of $\boldsymbol{\theta}_x$ by means of Lorentzian-curve fitting in the frequency domain. Taking into account definitions (29) and (30), and the fact that the spectrum and the noise can be estimated by

$$\beta(x,\nu) \simeq \beta(x,\nu;\boldsymbol{\theta}_x^{LF}) \tag{33}$$

$$\eta(x,\nu) \simeq \eta(x,\nu;\boldsymbol{\theta}_x^{LF}) \tag{34}$$

we can now estimate the SNR in Eq. (13), using the expression

$$\mathrm{SNR} \simeq 10\log_{10}\left(\frac{\sum_{x\in\mathcal{X}}\sum_{\nu\in\mathcal{V}}\{\beta(x,\nu;\boldsymbol{\theta}_x^{LF})\}^2}{\sum_{x\in\mathcal{X}}\sum_{\nu\in\mathcal{V}}\{\eta(x,\nu;\boldsymbol{\theta}_x^{LF})\}^2}\right). \tag{35}$$

Fig. 6 shows the filtering performance of the proposed filter for different values of Γ_h, where SNR values are estimated using the sequence of real data from El Teniente mine. It should also be noted that the SNR value varies over time in the sequence, reaching the lowest level at $t = 608$. These variations might have several causes like, for instance, a faulty installation of the optical fiber attached to the monitored structure. The curves show that the proposed filter achieves an improvement at the data quality (high SNR) as the value of Γ_h increases, as well as a reduction of temporal variations of the noise variance, due to a slower SNR decay with respect to unfiltered sequence, especially when using $\Gamma_h \geqslant$ 20 MHz.

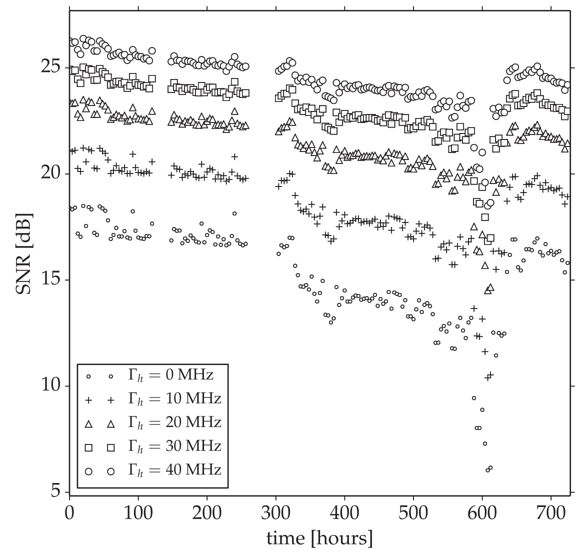

Fig. 6. Variation of the SNR using a filtered BGS sequence with different values of Γ_h. The curve $\Gamma_h = 0$ MHz stands for the variation of the SNR using the unfiltered sequence.

The proposed temporal smoothing technique, for maintaining the spatial coherence of the filtered data over time, is detailed next.

6. Filtering in the $\mathcal{T}$-domain

Recursive methods in time series consist in adaptation of a previous estimate by means of a correction term which depends both on the previous estimate and on a new observation. Due to their simplicity and good performance, they are successfully used for estimation, smoothing and forecasting in time series analysis. In this section, we describe our proposed method for temporal smoothing, based on a particular robust version of the Holt–Winters method. This version is able to estimate the model parameters in presence of abrupt and gradual trends and when observations are missing.

6.1. The Holt–Winters method

Suppose we have an univariate time series y_n, which is observed at $n = 1,\ldots,N$. The observation model is given by

$$y_n = \mu_n + \xi_n, \qquad \xi_n \sim f_\xi, \tag{36}$$

where μ_n stands for the original (unnoisy) observation at time n, and ξ_n is an additive noise that follows a one-dimensional distribution f_ξ. In the case of exponentially weighted moving average (locally constant regression), the value of the smoothed series at time n, $\tilde{\mu}_n$, is the solution of the following minimization problem [19]:

$$\tilde{\mu}_n = \arg\min_{\mu_n\in\Theta}\sum_{k=1}^{n}(1-\alpha)^{n-k}\{y_k - \mu_n\}^2 \tag{37}$$

where α is a parameter taking values between 0 and 1 which controls the degree of smoothing.

It is possible to show that the solution of the optimization problem (37) can be solved by means of the following recursive equation

$$\tilde{\mu}_n = \alpha y_n + (1-\alpha)\hat{\mu}_{n|n-1}, \tag{38}$$

where $\hat{\mu}_{n|n-1}$, the one-step-ahead forecast of y_n based on previous observations $\{y_k\}_{k=1}^{n-1}$, is obtained as

$$\hat{\mu}_{n|n-1} = \tilde{\mu}_{n-1}. \tag{39}$$

The recursive Eq. (38) suggests that, if the underlying mean is subject to large changes, α should be taken close to 1 so as to quickly attenuate the effect of old observations. However, if α is too close to 1, $\tilde{\mu}_n$ is subject to much higher random variations, because of an under-smooth estimated mean. The rate at which information is discounted is characterized by the asymptotic sample length (ASL), defined by

$$\text{ASL} = \frac{1}{\alpha}, \tag{40}$$

and corresponds to the number of samples per length unit of the data window within which the contribution of the error values is significant. Clearly, when $\alpha = 0$, the length of the data window becomes infinite and we obtain the standard least squares method. If we use values of α between 0.01 and 0.1, values of ASL lie in the range of 10–100 time samples, which corresponds to 40–400 h of data, according to the BOTDR system setting.

In order to improve the performance of the method in presence of trended time series, Holt [20] and Winters [21] proposed to include a local linear trend variable, with regression model of the general form $u_t + v_t t$, by solving the following minimization problem:

$$(\tilde{u}_n, \tilde{v}_n) = \arg\min_{u_n, v_n \in \Theta} \sum_{k=1}^{n} (1-\alpha)^{n-k} \{y_k - (u_n + v_n k\tau)\}^2. \tag{41}$$

The recursive equations to solve (41) are

$$\tilde{u}_n = \alpha y_n + (1-\alpha)\hat{u}_{n|n-1} \tag{42}$$

$$\tilde{v}_n = \beta \frac{\tilde{u}_n - \tilde{u}_{n-1}}{\tau} + (1-\beta)\tilde{v}_{n|n-1} \tag{43}$$

where $\tilde{u}_n$ y $\tilde{v}_n$ are the level and trend, respectively, τ is the sampling rate and $\alpha = \beta$ in general, according to (41); nevertheless, different values for smoothing the two parameters are commonly used in practice. In a similar way as for exponential smoothing, the smoothing parameters α and β take values between zero and one. Then, $\hat{u}_{n|n-1}$ and $\hat{v}_{n|n-1}$, the one-step-ahead forecast of y_n and of $\dot{y}_n$, respectively, are obtained as

$$\hat{u}_{n|n-1} = \tilde{u}_{n-1} + \tau\tilde{v}_{n-1} \tag{44}$$

$$\hat{v}_{n|n-1} = \tilde{v}_{n-1}. \tag{45}$$

where $\tilde{u}_{n-1}$ and $\tilde{v}_{n-1}$ are the level and trend estimated at time $n-1$, respectively.

Extending this method to the multidimensional case, where $M(\mathbf{s}, n)$ denotes the BGS measured by the BOTDR system at time n, Eqs. (42) and (43) represent an adaptive filtering technique to estimate $B(\mathbf{s}, n)$, the original (unnoisy) BGS, when changes in the level of the observed sequence are gradual. Then, the observation model (1) can be rewritten as

$$M_n = B_n + \varepsilon_n, \qquad \varepsilon_n \sim f_\varepsilon \tag{46}$$

where $M_n \equiv M(\mathbf{s}, n), B_n \equiv B(\mathbf{s}, n)$, and $\varepsilon_n \equiv \varepsilon(\mathbf{s}, n)$.

Rewriting (42) and (43), the estimation of B_n at time n is carried out using the following update equations:

$$\widetilde{B}_n = \alpha M_n + (1-\alpha)\widehat{B}_{n|n-1} \tag{47}$$

$$\widetilde{D}_n = \beta \frac{\widetilde{B}_n - \widetilde{B}_{n-1}}{\tau} + (1-\beta)\widehat{D}_{n|n-1} \tag{48}$$

where $\widetilde{B}_n$ denotes the estimation of B_n and $\widetilde{D}_n$ its temporal variation. Finally, the one-step-ahead forecast value of B_n based on previous observations $\{B_k\}_{k=1}^{n-1}$, $\widehat{B}_{n|n-1}$, and the forecast of its temporal variation, $\widehat{D}_{n|n-1}$, are given by

$$\widehat{B}_{n|n-1} = \widetilde{B}_{n-1} + \tau\widetilde{D}_{n-1} \tag{49}$$

$$\widehat{D}_{n|n-1} = \widetilde{D}_{n-1}. \tag{50}$$

6.2. Extended method for missing data

Sometimes, BOTDR data are missed, for instance because of an accidental break of the optical fiber produced by mining activities in the nearby area. If the smoothing parameters are taken to be constant over time, this may result in estimation errors, due to inadequate weights on older observations in the update equations.

To solve this problem, we use the extended Holt–Winters method introduce in [22] to deal with time series observed at irregular time intervals. If we assume that M_{t_n} is the observed BGS measured by the BOTDR system at time t_n, Eq. (46) can be rewritten as

$$M_{t_n} = B_{t_n} + \varepsilon_{t_n}, \qquad \varepsilon_{t_n} \sim f_\varepsilon \tag{51}$$

where B_{t_n} is the original (unnoisy) BGS and ε_{t_n} is a zero-mean additive noise at time t_n. Then, the forecasting model is defined as

$$\widehat{B}_{t_n|t_{n-1}} = \widetilde{B}_{t_{n-1}} + (t_n - t_{n-1})\widetilde{D}_{t_{n-1}} \tag{52}$$

$$\widehat{D}_{t_n|t_{n-1}} = \widetilde{D}_{t_{n-1}}, \tag{53}$$

where $\widehat{B}_{t_n|t_{n-1}}$ stand for the one-step-ahead forecast value of B_{t_n} based on the estimates of BGS at time t_{n-1}, and $\widehat{D}_{t_n|t_{n-1}}$ is the forecast of its temporal variation. Note that the factor $t_n - t_{n-1}$ weights temporal variations according to the time-distance between two consecutive measurements.

To update the parameters of model (51), we need to modify the recursive Eqs. (47) and (48) so that smoothing parameters can also be updated as needed. In this manner, update equations are rewritten as follows

$$\widetilde{B}_{t_n} = \alpha_{t_n} M_{t_n} + (1-\alpha_{t_n})\widehat{B}_{t_n|t_{n-1}} \tag{54}$$

$$\widetilde{D}_{t_n} = \beta_{t_n} \frac{\widetilde{B}_{t_n} - \widetilde{B}_{t_{n-1}}}{t_n - t_{n-1}} + (1-\beta_{t_n})\widehat{D}_{t_n|t_{n-1}}, \tag{55}$$

where $\widetilde{B}_{t_0} = M_{t_0}$ and $\widetilde{D}_{t_0} = 0$ are initial conditions.

Taking into account the time span $t_n - t_{n-1}$ between two consecutive observations, and the sampling rate τ, we update the smoothing parameters as follows

$$\alpha_{t_n} = \frac{\alpha_{t_{n-1}}}{(1-\alpha)^{\frac{t_n - t_{n-1}}{\tau}} + \alpha_{t_{n-1}}} \tag{56}$$

$$\beta_{t_n} = \frac{\beta_{t_{n-1}}}{(1-\beta)^{\frac{t_n - t_{n-1}}{\tau}} + \beta_{t_{n-1}}}. \tag{57}$$

Initial conditions are fixed at $\alpha_{t_0} = \alpha$ and $\beta_{t_0} = \beta$. Clearly, if observations are regularly sampled, with $t_n - t_{n-1} = \tau$, then $\alpha_{t_n} = \alpha$ and $\beta_{t_n} = \beta, \forall t_n$.

Parameters α and β are commonly know as forgetting factors. Values of α and β are chosen looking for trade-off between convergence velocity of the algorithm and variance of the estimation.

The main assumption underlying this methodology is that the parameters describing the data are either constant or slowly time-varying. Therefore, abrupt level shifts in the strain time series (see e.g. Fig. 7) cannot be suitably predicted, in which case the estimates may fail. We next deal with this problem, proposing a modification of the update equations based on an analytical tool that detects unusual or abrupt changes.

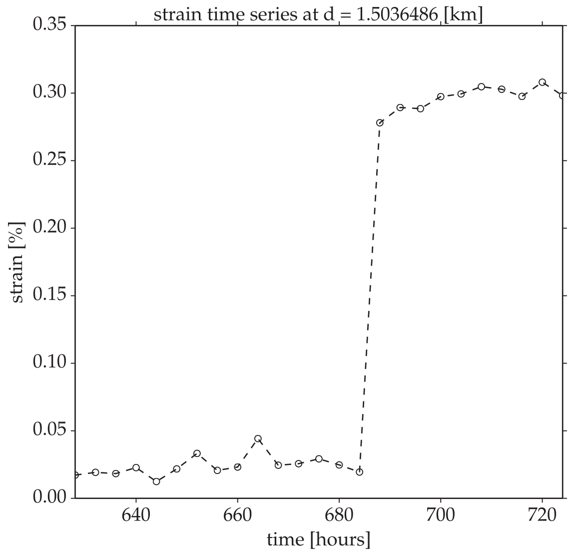

Fig. 7. A level shift in the strain time series.

6.3. Extended method for abrupt changes detection

By abrupt change we understand a time instant at which parameters suddenly change, from some constant, stationary, or slowly time-varying regime, to another. It should also be noted that although abrupt changes commonly imply changes with large magnitude, many change detection problems are concerned with the detection of small changes [23]. This notion serves as a basis to the corresponding formal mathematical problem statement, and to the formal derivation of algorithms for change detection.

Let us suppose that a strain-level shift at distance x along the optical fiber occurs at time t_n. If we analyze and compare the Brillouin gain spectra $B_{t_{n-1}}$ and B_{t_n}, the Lorentzian distribution of B_{t_n} at distance x is abruptly shifted at a frequency value proportional to the level shift. The difference of spectra $B_{t_n} - B_{t_{n-1}}$ will show some energy variation at x, as shown in Fig. 8. We propose to use the

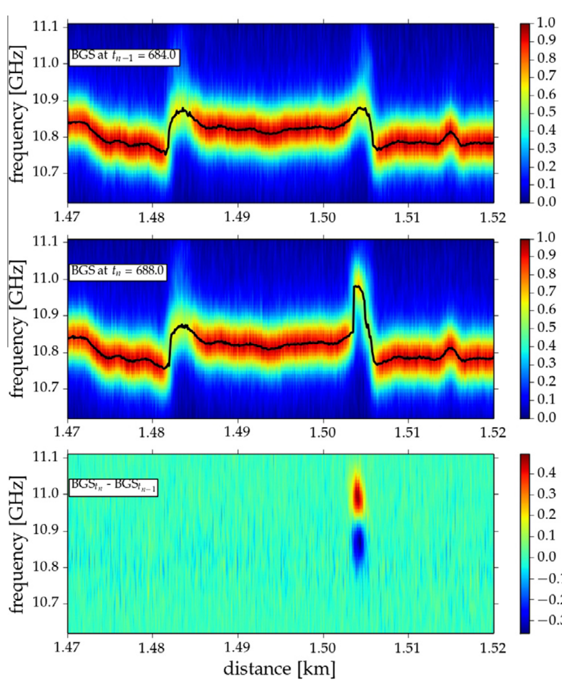

Fig. 8. Abrupt change detection using the Brillouin gain spectrum.The black line denotes the curve of frequency shifts obtained at each distance x.

forecast error $\widehat{E}_{t_n}(x,\nu)$ as discriminant information to design a suitable level-shift indicator, able to detect when the abrupt change occurrences. The forecast error provides information on whether the forecast model has been able to track intrinsic variations of a time series. It is defined as follows:

$$\widehat{E}_{t_n}(x,\nu) \equiv \widehat{E}(\mathbf{s},t_n) = M(\mathbf{s},t_n) - \widehat{B}(\mathbf{s},t_n|t_{n-1}). \tag{58}$$

with

$$\widehat{B}(\mathbf{s},t_n|t_{n-1}) = \widetilde{B}(\mathbf{s},t_{n-1}) + (t_n - t_{n-1})\widetilde{D}(\mathbf{s},t_{n-1}), \tag{59}$$

where $\widehat{B}(\mathbf{s},t_n|t_{n-1})$ stands for the one-step-ahead forecast of $B(\mathbf{s},t_n)$ given t_{n-1}. When changes are gradual and forecast errors are assumed homoskedastic, $\widehat{E}_{t_n}(x,\nu)$ should be a zero-mean noise with constant variance, of similar order as the measurement noise variance. However, an abrupt level-shift at some point of the optical fiber should result in an observed forecast error $\widehat{E}_{t_n}(x,\nu)$ of increased variance, with respect to the usual situation.

Using this fact, we propose a level shift indicator ψ_{t_n} based on the positive forecast error $\widehat{E}^+_{t_n}(x,\nu)$, defined by

$$\widehat{E}^+_{t_n}(x,\nu) = \begin{cases} \widehat{E}_{t_n}(x,\nu) & \widehat{E}_{t_n}(x,\nu) \geqslant 0, \\ 0 & \widehat{E}_{t_n}(x,\nu) < 0. \end{cases} \tag{60}$$

We choose to measure the positive forecast error since it indicates the new central frequency and thus provides more information than, for instance, the absolute value of the error. Nevertheless, for the purpose of the present algorithm, we could also use shift indicators defined in terms of the absolute value or of other positive functions of the forecast error.

Given the definition of $\widehat{E}^+_{t_n}(x,\nu)$, and in order to detect level shifts based on their time-varying spectral variations, we consider first its sample standard deviation (ssd) $S_\nu(x)$ at each point x in the frequency domain:

$$S_\nu(x) = \sqrt{\frac{1}{|\mathcal{V}|-1}\sum_{\nu\in\mathcal{V}}\left(\widehat{E}^+_{t_n}(x,\nu) - A_\nu(x)\right)^2}, \tag{61}$$

where $A_\nu(x)$ denotes an average in the frequency domain, so that

$$A_\nu(x) = \frac{1}{|\mathcal{V}|}\sum_{\nu\in\mathcal{V}}\widehat{E}^+_{t_n}(x,\nu). \tag{62}$$

Under the assumption that zones with large variations are those where abrupt changes are detected, $S_\nu(x)$ is able to quantify and locate shifts of the Lorentzian-like distribution that are relatively abrupt, when compared to those of other points along the fiber. In order to have values between 0 and 1, we now define $\Psi_{t_n}(x)$ as the normalized version of Eq. (61), given by

$$\Psi_{t_n}(x) = \frac{S_\nu(x)}{S}, \tag{63}$$

where S is

$$S = \sqrt{\frac{1}{|\mathcal{X}||\mathcal{V}|-1}\sum_{(x,\nu)\in\mathcal{S}}\left(\widehat{E}^+_{t_n}(x,\nu) - A\right)^2}, \tag{64}$$

with A defined by

$$A = \frac{1}{|\mathcal{X}||\mathcal{V}|}\sum_{(x,\nu)\in\mathcal{S}}\widehat{E}^+_{t_n}(x,\nu). \tag{65}$$

Since we actually do not need to locate the zone where the level shift was detected, the level shift indicator ψ_{t_n}, finally defined as

$$\psi_{t_n} = \sqrt{\frac{1}{|\mathcal{X}|-1}\sum_{x\in\mathcal{X}}(\Psi_{t_n}(x) - \bar{\psi}_{t_n})^2}, \tag{66}$$

with

$$\bar{\psi}_{t_n} = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\Psi_{t_n}(x)$$

reduces the information contained in the change detection function $\Psi_{t_n}(x)$ into a simple real-valued scalar function.

As regards the update equations of the model parameters, we will add a level shift condition of the form $\psi_{t_n} < u\psi_{t_{n-1}}$, where $u > 0$ is a detection threshold to be fixed. If the latter condition is true, parameters are updated according to Eqs. (54) and (55). If the condition is false, parameters are updated according to the following rule

$$\widetilde{B}_{t_n}(x, v) = M^*_{t_n}(x, v) \tag{67}$$

$$\widetilde{D}_{t_n}(x, v) = \widetilde{D}_{t_{n-1}}(x, v), \tag{68}$$

where $M^*_{t_n}(x, v)$ is the BGS after being filtered using the convolution filter in the $\mathcal{S}$-domain. This rule is based on the assumption that only the level can vary abruptly in the time series and that the underlying trend function must be continuous over time.

7. Proposed algorithm

In summary, our online two-stage algorithm for strain profile estimation with outliers removal, level shift detection and possibly missing observations consists in the following steps:

- **Requirements:** $\Gamma_h, \alpha, \beta, u$.
- **Step 0:** Adjustment of initial conditions:
  - $\widetilde{B}(\mathbf{s}, t_0) = M^*(\mathbf{s}, t_0)$
  - $\widetilde{B}(\mathbf{s}, t_1) = \frac{M^*(\mathbf{s},t_1)+M^*(\mathbf{s},t_0)}{2}$
  - $\widetilde{D}(\mathbf{s}, t_1) = \frac{\widetilde{B}(\mathbf{s},t_1)-\widetilde{B}(\mathbf{s},t_0)}{t_1-t_0}$
  - $\alpha_{t_1} = \alpha[(1-\alpha)^{\frac{t_1-t_0}{\tau}} + \alpha]^{-1}$
  - $\beta_{t_1} = \beta[(1-\beta)^{\frac{t_1-t_0}{\tau}} + \beta]^{-1}$
- **Step 1:** $n = 2$.
- **Step 2:** Reading of data at time t_n: $M(\mathbf{s}, t_n)$
- **Step 3:** Filtering in the distance-frequency domain: $M^*(\mathbf{s}, t_n) = M(\mathbf{s}, t_n) * h(\mathbf{s})$
- **Step 4:** Normalization of the filtered BGS: $M^N(\mathbf{s}, t_n) = \frac{M^*(\mathbf{s},t_n)-M^*_{min}}{M^*_{max}-M^*_{min}}$, such that $0 \leqslant M^N(\mathbf{s}, t_n) \leqslant 1$, where $M^*_{min} = \min\{M^*(\mathbf{s}, t_n)\}$ and $M^*_{max} = \max\{M^*(\mathbf{s}, t_n)\}$. Since the temporal noise variance varies at each observation, an affine transformation is applied, so that the filtered BGS $M^*(\mathbf{s}, t_n)$ varies between zero and one, in the $\mathcal{S}$-domain.
- **Step 5:** Estimation of the forecast error according to Eq. (58) and computation of the level shift indicator (66).
- **Step 6:** Parameters updating. This step checks if the condition $\psi_{t_n} < u\psi_{t_{n-1}}$ is true or false:
  - If it is true, parameters are updated according to Eqs. (54) and (55).
  - If it is false, parameters are updated according to Eqs. (67) and (68).
- **Step 7:** Updating of the smoothing parameters by means of Eqs. (56) and (57).
- **Step 8:** Set $n \leftarrow n + 1$ and repeat steps from 2 to 8.

For each time and position along the fiber, the strain value can then be obtained from the Brillouin frequency shift estimated by fitting a Lorentzian function.

8. Results

In this Section, we present experimental results obtained with a Python implementation of the proposed algorithm. The method is applied to the real time BGS sequence obtained in a fiber section installed inside El Teniente mine of Chilean state mining company Codelco, (the description of which was given in Section 5.3). Besides, we use the following parameters to configure the algorithm: $\Gamma_h = 40$ MHz, $\alpha = \beta = 0.1$ (10 time samples, 40 h) and $u = 2$.

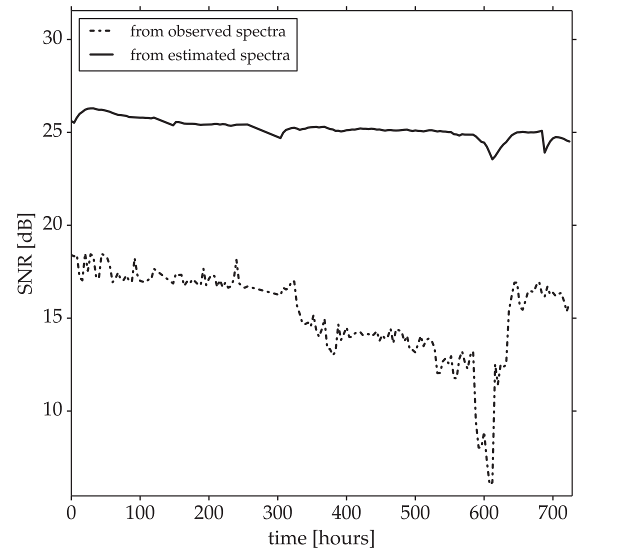

Fig. 9. The algorithm performance using the SNR computed spectrum-by-spectrum.

The algorithm performance is measured using the SNR using definition (13). Fig. 9 shows the performance of the proposed algorithm. SNR values are computed spectrum-by-spectrum. Note that the SNR values of filtered spectra improved in at least 8 dB upon those of observed spectra. In particular, the SNR value of the estimated BGS at $t = 608$ is larger by about 18 dB that the corresponding value for the noisy BGS. Also, SNR values tend to be temporally constant in the sequence. Fig. 10 offers a visual comparison between a noisy and an estimated BGS. Note that the quality of processed data is consistently superior compared to the noisy data.

By means of a visual comparison between the noisy and the filtered BGS at $t = 608$, Fig. 11 moreover shows the robustness of the proposed algorithm against outliers. The effect of the filtering over time on the strain profile is shown in Fig. 12: outliers are removed while the characteristic shape of the strain profile is kept.

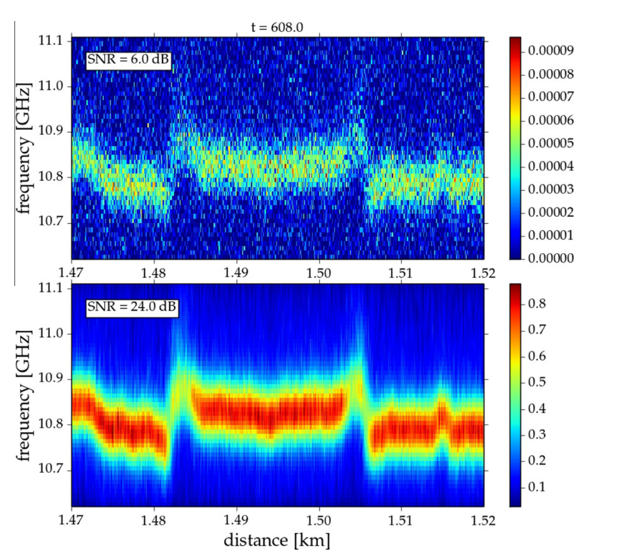

Fig. 10. A comparison between the observed and the estimated BGS at $t = 608$.

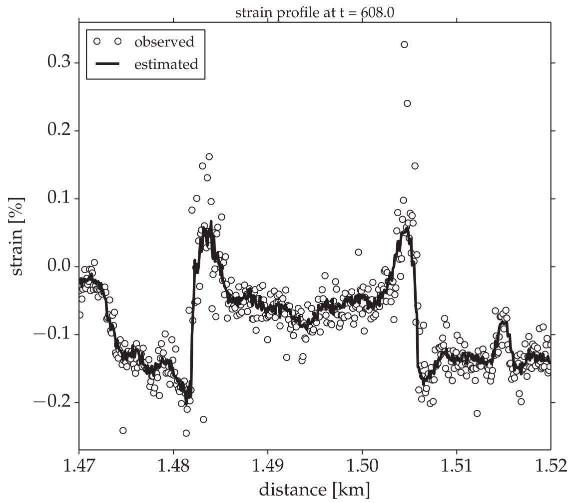

Fig. 11. A comparison between strain profiles obtained from the observed and the estimated BGS at $t = 608$.

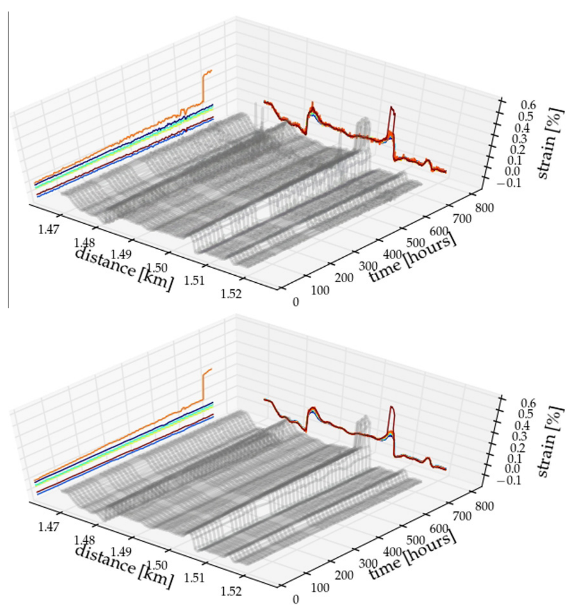

Fig. 12. Temporal smoothing. Top: Observed strain profile. Bottom: Estimated strain profile.

A comparison between strain time series extracted spectrum-by-spectrum from noisy and from estimated strain by Lorentzian fitting is shown in Fig. 13. Together with a systematic and time-consistent removal of outliers, we observe how the algorithm adapts itself to the gradual and abrupt changes of the time series.

9. Computational cost and performance

Thanks to the convolution theorem, the spectral filtering can be done using Fourier Transform (FT) at each position on the fiber and each time step. By using FFT (Fast Fourier Transform) the 1-dimensional convolution at each position x can moreover be computed with a computational complexity of $O(|\mathcal{X}||\mathcal{V}|\log|\mathcal{V}|)$, where $|\mathcal{X}|$ and $|\mathcal{V}|$ stand for the number of sample frequencies and positions, respectively (as opposite to a $O(|\mathcal{X}||\mathcal{V}|^2)$ complexity using standard FT.).

The algorithm is recursive in the time domain and the corresponding matrix operations are carried out element by element.

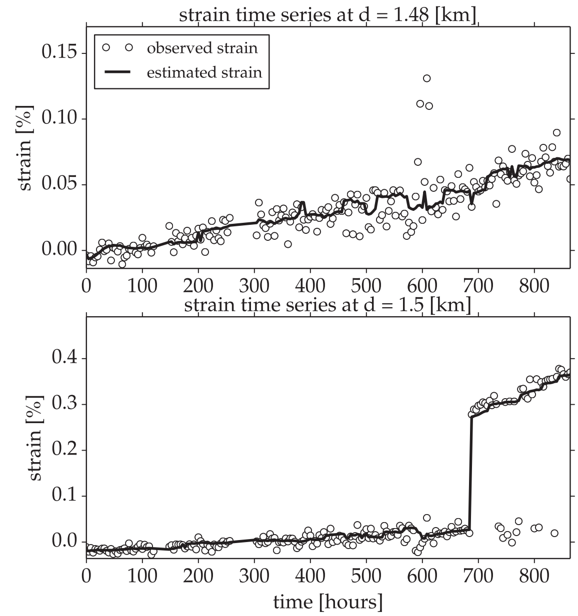

Fig. 13. A comparison between time series extracted spectrum-by-spectrum from noisy and estimated strain. The line denotes the time series extracted from the estimated sequence. Circles stand for the time series extracted from the noisy sequence.

These operations consists in multiplications and sums, whose computational complexity is $O(|\mathcal{X}||\mathcal{V}|)$.

Hence, the computational cost is of order of $O(|\mathcal{X}||\mathcal{V}|\log|\mathcal{V}|)$ for each time step, and this which grants that the global computational cost will scale in a reasonable way when our method is applied to sensing fiber systems much longer that the one considered in Section 8. In particular, if $|\mathcal{V}| \ll |\mathcal{X}|$ or if $|\mathcal{V}|$ is fixed, the complexity depends linearly on $|\mathcal{X}|$. The recursive nature of the algorithm grants on the other hand that the complexity also grows linearly with time.

The processing time of our algorithm was compared with the method proposed in [16], which is based on a 3-dimensional NLM (Non Local Means) smoothing technique. These authors used a data matrix of $|\mathcal{X}| = 100{,}000 \times |\mathcal{V}| = 200$ points taken from a BOTDA sensor, considering 10 consecutive frames, that is, $100{,}000 \times 200 \times 10$ points. Using a conventional computer with a 3.5 GHz processor and 8 GB RAM, they report in that setting a processing time of about 4 min.

For a data matrix of the same size, that is, with $100{,}000 \times 200$ simulated points, and considering 10 consecutive frames as well, our recursive algorithm analyzed the data in about 5 s, using a conventional computer with a 3.1 GHz processor and 4 GB RAM.

Unfortunately, we were not able to compare the two methods simultaneously in terms of computational cost and estimations results on real data, for a measuring configuration similar to the one considered in [16] (we do not dispose by the moment of such a sensing fiber facility).

Nevertheless, the theoretical and simulation results of Section 5 suggest that a significant relative quality improvement should be expected in SNR terms, irrespective of the initial quality of the signal at each point of the fiber, and in particular of the length of the fiber to which our method is applied.

It is also important to keep in mind that, in addition to the data size and the degree of accuracy required to estimate the strain profile, the choice of a method should also take into account the relation between the temporal evolution scale of the underlying physical process under observation, and the data acquisition time scale.

10. Conclusions

We have proposed an easily implementable two-stage adaptive algorithm for strain profile estimation from noisy and abruptly changing BOTDR data. We have shown that data quality of a sequence of noisy Brillouin gain spectra is significantly improved by our proposed method. A pseudo 2-dimensional convolution filter, represented by a Lorentzian function centered at zero with smoothing parameter Γ_h, was proposed for filtering the observed BGS in the distance-frequency domain. The main characteristic of this filtering is that it does not affect the value of the central frequency of the Lorentzian function measured by the BOTDR system and, as a consequence, the value of the strain associated with the Brillouin frequency shift is preserved. This filtering technique was used to remove outliers of the strain extracted from the BGS.

For filtering in the time domain, we have proposed a time series model to estimate the variations of the sequence of BGS during the analysis period. Since the deformation produced in the rock mass is mainly gradual over time, we have proposed to track the changes using a time series model at the level of spectrum, consisting of a linear trend with an additive noise term.

The model parameters were estimated by means of updating equations, in the context of an exponential moving average method with two smoothing parameters. The original updating equations of the Holt–Winters method were modified so that the algorithm be robust when strain level shifts are observed and when data is missed. The information provided by the forecast error was used to design a level shift indicator to update the parameters in a suitable way.

The algorithm performance was assessed by computing a spectrum-by-spectrum SNR value, associated with each Brillouin gain spectrum of the sequence. The proposed methodology has improved data quality, increased the SNR in the whole sequence and made its noise variations steady over time. Moreover, our method is able to consistently discriminate genuine strain level shifts from outliers due to the situation of low-quality data.

Since our approach allows for reducing the uncertainty of the strain estimate in low SNR environments, our algorithm should in principle be applicable even in long fiber configurations, where the SNR values of the original signal might decrease over the distance. This, together with the different analysis and test we performed on our method, suggest that it can in principle be used for similar purposes as the procedure presented in [16] (despite we could not test our method in similar long range fiber configurations).

One of the relevant features of the NLM method used in [16], which makes it a suitable solution for robust filtering, is the use of geometric information of the images and their time evolution, in order to estimate without bias areas of both low and high curvature. In our method, the geometric features of the data are dealt with by modeling in specific ways the frequency, distance and time variables.

This distinguished treatment of the variables is also the reason why we are able to, for instance, identify frequency shift time-trends at each position of the fiber, which is of significant value for risk forecasting, and to do our analysis in a fast and computationally very efficient way. The presented estimation procedure should therefore represent a viable alternative to the method in [16], especially when samples are taken at short time intervals, when the time evolution of the process under analysis is not slow and/or if the processing time is too large to study particular transient behaviors.

We have shown that the proposed algorithm is suitable for efficiently estimating the statistical behavior of strain time series with both abrupt and gradual changes over time, and in a robust way against high levels of measurement noise. In particular, this methodology is useful when monitoring small sections of rock mass by means of optical fiber sensors and BOTDR technology. Our algorithm should thus contribute to a more accurate risk assessment in mining activities, based on the statistical time analysis of rock mass behavior, but we also expect it to be useful in more standard sensing configurations.

In principle, it should also be possible to extend the use of this algorithm to other problems or sensoring systems, where the accurate and robust measurement of spatially distributed physical parameters is required over periods of time.

Acknowledgements

This work was partially supported by BASAL-Conicyt Center for Mathematical Modeling, Chile and by MICOMO S.A company, Research Contracts MIC-001/2007 and MIC-020/2009. The authors also thank an anonymous referee for comments and question which allowed us to improve the presentation of our results, and for drawing their attention to Ref. [16].

References

[1] Y. Zhang, C. Yu, X. Fu, D. Li, W. Jia, W. Bi, An improved Newton algorithm based on finite element analysis for extracting the Brillouin scattering spectrum features, Measurement 51 (2014) 310–314.
[2] T. Kurashima, T. Horiguchi, H. Izumita, S. Furukawa, Y. Koyadama, Brillouin optical-fiber time domain reflectometry, IEICE Trans. Commun. E76-B (1993) 382–390.
[3] X.F. Zhang, Z.H. Lv, X.W. Meng, F. Jiang, Q. Zhang, Application of optical fiber sensing real-time monitoring technology using in Ripley landslide, Appl. Mech. Mater. 610 (2014) 199–204.
[4] C. Gao, F. Wang, C.L. Li, Application of BOTDR in railway fence intrusion detecting system, Appl. Mech. Mater. 71–78 (2011) 4278–4281.
[5] G. Li, G. Wang, Y. Sun, N. Guo, The application of BOTDR on health diagnosis in the Wangjiatan bridge, in: International Symposium on Optoelectronic Technology and Application, International Society for Optics and Photonics, 2014 (pp. 92972M-92972M).
[6] K. Komatsu, K. Fujihashi, M. Okutsu, Application of optical sensing technology to the civil engineering field with optical fiber strain measurement device (BOTDR), Proc. SPIE 4920, vol. 352, 2002.
[7] N. Shiqing, G. Qian, Application of distributed optical fiber sensor technology based on BOTDR in similar model test of backfill mining, Proc. Earth Planet. Sci. 2 (2011) 34–39.
[8] H. Naruse et al., Application of a distributed fibre optic strain sensing system to monitoring changes in the state of an underground mine, Meas. Sci. Technol. 18 (2007) 3002.
[9] S. Wang, L. Luan, Analysis on the security monitoring and detection of mine roof collapse based on BOTDR technology, in: 2nd International Conference on Soft Computing in Information Communication Technology, Atlantis Press, 2014.
[10] N. Shibata, R.G. Waarts, R.P. Braun, Brillouin-gain spectra for single-mode fibers having pure-silica, GeO_2-doped, and P_2O_5-doped cores, Opt. Lett. 12 (1987) 269–271.
[11] T. Horiguchi, T. Kurashima, M. Tateda, Tensile strain dependence of Brillouin frequency shift in silica optical fibers, IEEE Photon. Technol. Lett. 1 (1989) 107–108.
[12] T. Horiguchi, K. Shimizu, T. Kurashima, M. Tateda, Y. Koyamada, Development of a distributed sensing technique using Brillouin scattering, J. Lightwave Technol. 13 (1995) 1296–1302.
[13] H. Minshuang, C. Weimin, H. Shanglian, W. Xingqiang, Theoretical analysis of distributed optic fiber strain sensor based on Brillouin scatter, Opto-electron. Eng. 22 (1995) 11–36.
[14] J. Gao, B. Shi, W. Zhang, H. Zhu, Monitoring the stress of the post-tensioning cable using fiber optic distributed strain sensor, Measurement 39 (2006) 420–428.
[15] C. Zhang, Y. Yang, A. Li, Application of Levenberg–Marquardt algorithm in the Brillouin spectrum fitting, in: Seventh International Symposium on Instrumentation and Control Technology, International Society for Optics and Photonic, 2008.
[16] M.A. Soto, J.A. Ramírez, L. Thévenaz, Intensifying the response of distributed optical fibre sensors using 2D and 3D image restoration, Nat. Commun. 7 (2016) 1–11.
[17] M. Soto, L. Thévenaz, Modeling and evaluating the performance of Brillouin distributed optical fiber sensors, Opt. Exp. 21 (2013) 31347–31366.
[18] Y. Pawitan, In All Likelihood: Statistical Modelling and Inference Using Likelihood, Oxford Science Publications, Great Clarendon Street, Oxford OX2 6DP, 2001.
[19] S. Gelper, R. Fried, C. Croux, Robust forecasting with exponential and Holt–Winters smoothing, J. Forecast 29 (2010) 285–300.

[20] C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, ONR Res. Memor. 52 (1957).
[21] P. Winters, Forecasting sales by exponentially weighted moving averages, Market. Sci. 6 (1960) 324–342.
[22] T. Cipra, Exponential smoothing for irregular data, Appl. Math. 51 (2006) 597–604.
[23] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.