Journal of Applied Geophysics 183 (2020) 104201

Contents lists available at ScienceDirect

Journal of Applied Geophysics

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

Take the *Eh*? train: Distributed Acoustic Sensing (DAS) of commuter trains in a Canadian City

Robert J. Ferguson [a,*], Matthew A.D. McDonald [b], David J. Basto [c]

[a] Department of Geoscience, University of Calgary, 2500 University Drive NW, Calgary T2N 1N4, Canada
[b] Fotech Solutions Inc, Suite 280, 700 6th Avenue SW, Calgary T2P 0T8, Canada
[c] The City of Calgary - Information Technology Services, Information Technology Infrastructure Services, Canada

ARTICLE INFO

Article history:
Received 10 November 2019
Received in revised form 18 October 2020
Accepted 20 October 2020
Available online 22 October 2020

Keywords:
Tracking
Commuter train
Distributed Acoustic Sensing (DAS)
Optical fibre
Telecom fibre
Rayleigh scattering
Kalman filter

ABSTRACT

We attach a Distributed Acoustic Sensing (DAS) system to an existing telecom fibre that follows the Red Line of the City of Calgary Light Rail Transit (LRT). The City does not have a Global Positioning Satellite (GPS) system to track the position of trains on the Red Line, and very few trains in the fleet are even GPS equipped. Therefore, we propose DAS tracking as an alternative to the retrofit and development of a GPS-based system or as a companion to a future system. Trains on the Red Line register as intensity peaks in the DAS soundfield, and we deduce the DAS distance between the City Hall LRT station (the origin) and the Tuscany station (the terminus) for all trains on the system by tracking intensity peaks.

To estimate DAS position from DAS distance, we use the speedometer logs and GPS-position logs from one of the few GPS-equipped trains on the Red Line. We track this train with DAS during three trips from City Hall station to Tuscany station and then obtain the logs for those trips. Analysis of the logs shows that more than 98% of the GPS positions are either null-valued or are repeats of previous values. This paucity of actual positions suggests that the effective GPS positions lie on a very coarse grid. We convert the coarse GPS-positions into GPS distances and interpolate them onto the dense speedometer-grid using a Kalman filter. The heading along the Redline changes more slowly than its distance so we approximate a unit heading-vector on the dense grid. We show that the product of the Kalman GPS-distance and the unit heading-vector gives an optimal GPS heading-vector on the dense grid. Latitude and longitude at points along the dense grid correspond to the north-south and east-west components of the GPS heading-vector, respectively.

We Kalman filter the DAS distances onto the dense-grid and we find also that this minimizes a number of DAS noise sources. With the Kalman DAS-distances and Kalman GPS-positions on the same dense-grid, we now have an optimal relationship that is time independent. For as long as the fibre and the Red Line themselves remain stationary, all DAS distances obtained from tracking future trains defines unique latitude and longitude positions in realtime.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The City of Calgary Light Rail Transit (LRT) system consists of an east-west line called the Blue Line and a north-south line called the Red Line (Fig. 1). The fleet of cars that link to form the trains on this system were acquired by the City between 1979 and 2013; Fig. 2 is an example of an early unit that is still in service.

GPS logs are available from some newer trains, but the City has no system in place to consume, interpret, or broadcast train-positions. Our inspection of the GPS position-data from the GPS-equipped trains shows that position is impacted negatively by reduced satellite visibility (Sasani et al., 2016; Bento et al., 2017) due to tunnels and urban canyon effects (Gadziński, 2018, for example) - we find that more than 98% of the GPS positions are either null or are repeats of previous values, and the average distance between reliable GPS positions is more than 50 m. All GPS-based systems are exposed to signal-degrading circumstances like deliberate *spoofing* of the GPS signal (Jafarnia-Jahromi et al., 2012; Henningsen et al., 2018; Ganin et al., 2019) so we feel that the development of tracking companion to GPS for commuter trains is justifiable.

As a companion to a future GPS control-system for the Red Line, or perhaps as alternative, we connect a Distributed Acoustic Sensing (DAS) system to the existing City of Calgary telecom fibre running beneath the Red Line and track the position and heading of all trains in realtime. Note, rail applications of geophysics tend to concentrate

* Corresponding author.
E-mail addresses: rjfergus@ucalgary.ca (R.J. Ferguson), matt.mcdonald@fotechsolutions.com (M.A.D. McDonald), david.basto@calgary.ca (D.J. Basto).

https://doi.org/10.1016/j.jappgeo.2020.104201
0926-9851/© 2020 Elsevier B.V. All rights reserved.

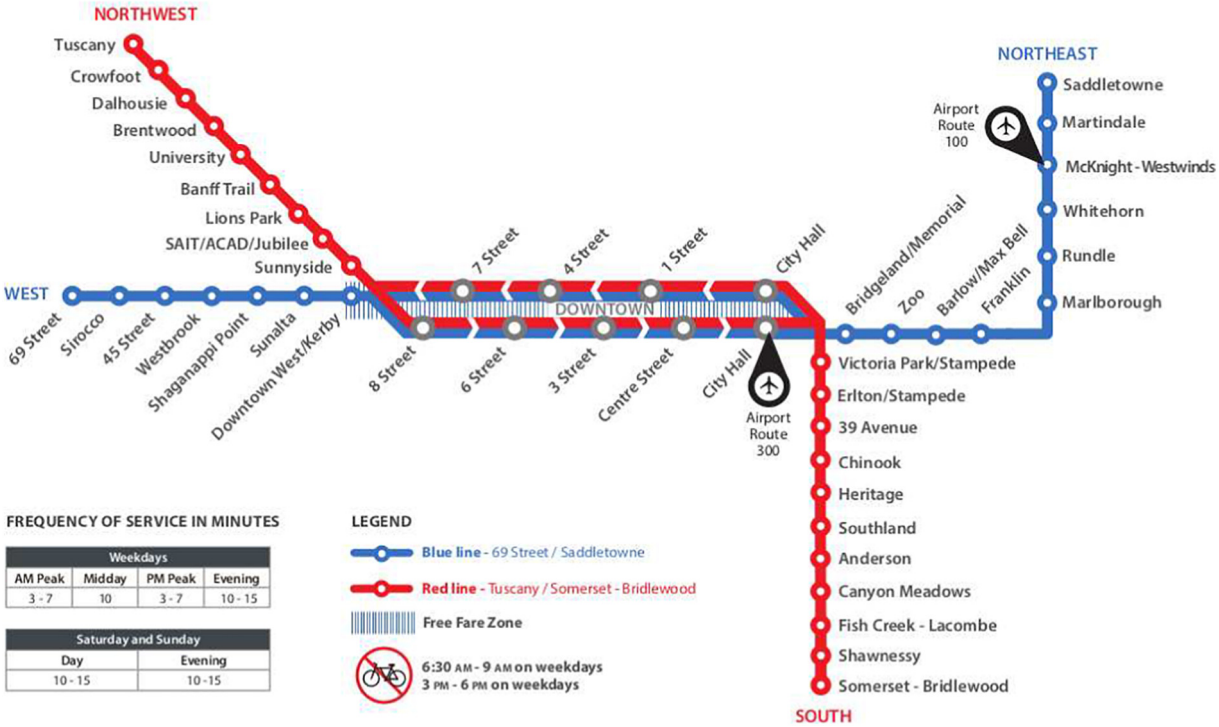

Fig. 1. Map of the City of Calgary LRT system. Data and results for this study correspond to the Red Line between City Hall and Tuscany stations.

Fig. 2. Car 2005 from the City of Calgary Light Rail Transit (LRT) system. This car is one of the original fleet purchased between 1979 and 1985.

on trackbed monitoring with ground-penetrating radar (GPR) (Khakiev et al., 2014a, 2014b, for example) and seismic methods (Pérez-Santisteban et al., 2011, for example), so to monitor the actual traffic above the roadbed rather than the roadbed itself is, we accept, is somewhat unusual geophysically.

DAS systems are sensitive to strain changes along optical fibres so street-level disturbances like trains, vehicle traffic, and construction works reveal their position and velocity when close to a DAS fibre. The identification of intercity rail (rail for freight between cities) defects reduces derailment events (Mohammadi et al., 2019, for example) and DAS has found application in this realm (He et al., 2019, for example). Bespoke fibre is usually installed for DAS tracking of inner-city trains, and distance and velocity are measured rather than position (He et al., 2019).

DAS attached to existing telecom infrastructure is a fairly recent geophysical innovation (Wiesmeyr et al., 2020, for example). In earthquake seismology for example, Martin et al. (2017) use the Stanford University telecom fibre and compare DAS and conventional passive and active seismic recordings. Biondi et al. (2017) use the same array to show that DAS reliably captures P-wave, S-wave and surface-wave kinematics.

Oilfield DAS applications like marine 3D seismic (Mateeva et al., 2014; Hornman, 2017, for example) and Vertical Seismic Profiling (Mateeva et al., 2014; Munn et al., 2017, for example) being the most common DAS applications. Increasingly, however, DAS acquired at street-level is used to investigate anthropomorphic soundfields. Aktas et al. (2017) use DAS to secure physical assets from threats in cities. Dou et al. (2017) use interferometry of DAS traffic-signatures to detect changes in the geotechnical attributes of soils. Iida et al. (2017) show how buildings are monitored using DAS, and threats to pipeline infrastructure are monitored using DAS by Tejedor et al. (2017). Zeng et al. (2017) install fibre at Garner Valley, California and generate cross-correlation DAS functions using traffic on CA Highway 74 as a source soundfield for geological imaging. Even a human walker is detectable at ground level using DAS (Ku and Duckworth, 2013; Wu et al., 2014). Papp et al. (2016) propose the use of existing telecom fibre installed near rail tracks and they present a real-time algorithm that associates DAS intensity peaks with trains.

The City of Calgary telecom system forms an extensive grid throughout Calgary, and a bundle of fibre follows the Red Line. Fig. 3 is a map of part of the Red Line (labelled "Track" on Fig. 3) between the SAIT and Lions Park stations.

The telecom fibre (labelled "Fibre" in Fig. 3) follows the Red Line closely between City Hall station and Tuscany station except for the occasional meanders (an example meander is circled on Fig. 3). Meanders accumulate between City Hall and Tuscany so we must correct any

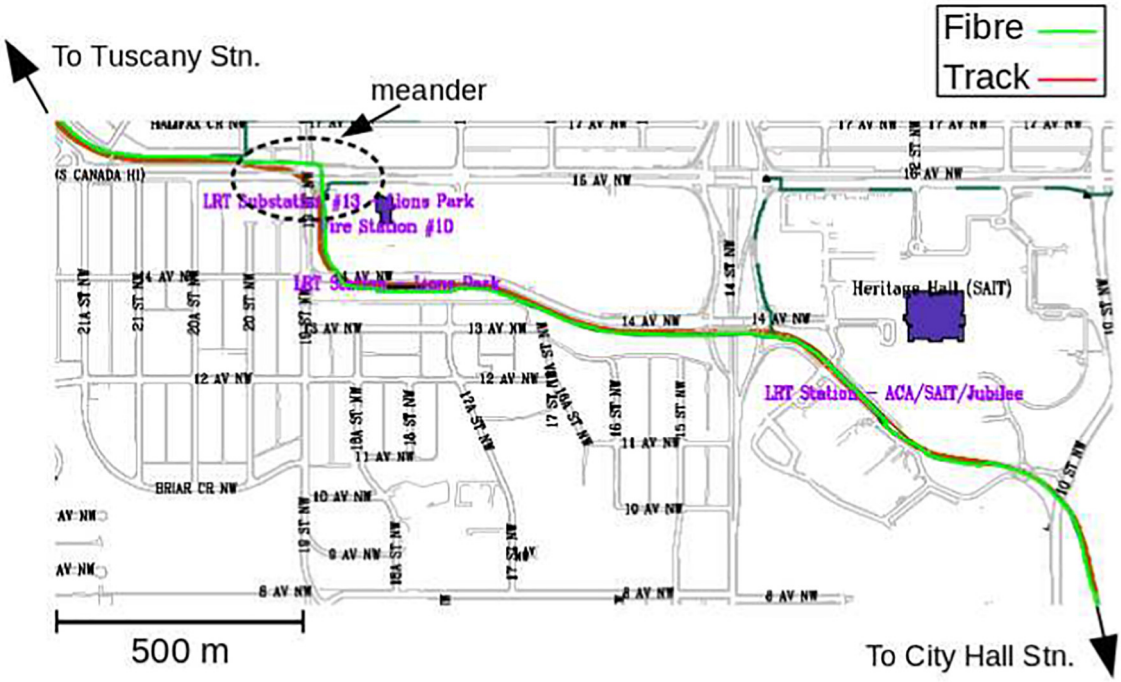

Fig. 3. Map of the City of Calgary telecom fibre and the LRT Red Line between Lions Park station and Sunnyside station (see Fig. 1 for an overview of the Red Line). With the exception of loops of slack fibre and short meanders (a meander is circled), the telecom fibre follows the Red Line from City Hall station to Tuscany station.

distance measured along the fibre because the fibre is longer than the Red Line. Though careful measurement using engineering drawings might be sufficient to deduce a correction, the presence of loops of slack fibre of unknown length compounds the problem.

Rather than correct the length of the DAS fibre to match the length of the Red Line, we determine a relationship between distance along the fibre and the latitude and longitude of the Red Line. We first identify a GPS-equipped four-car test train and track it using DAS over three trips of the north-west leg of the Red Line between City Hall station and Tuscany station. The differing numerical grids between the GPS logs and the DAS distances plus a number of DAS / GPS noise-sources, complicate the DAS / GPS relationship.

We reconcile the grids and reduce the noise by Kalman filtering (Kalman, 1960) the DAS distances (d_{DAS}) with the speedometer logs for each car for all three trips. Analysis indicates that the first trip of the three is the least noisy, and that d_{DAS} correspond most closely to the first two cars of the test train.

The resulting Kalman DAS-distances (k_{DAS}) and the speedometer logs lie on the same numerical grid, but the GPS coordinate logs lie on a much coarser effective-grid due to large numbers of null-values and repeated-values. Kalman filtering does not reconcile the GPS logs with the actual grid because the speedometer data are scalar valued and positive. Instead, we construct a unit heading-vector ($\hat{u}$) for every point on the actual grid by simple interpolation. The vertical and horizontal components of $\hat{u}$ correspond to the latitude and longitude directions from the GPS logs, respectively. We use meridian arcs (Rapp, 1991, for example) to convert the valid GPS positions to GPS distances (d_{GPS}), and then we Kalman filter d_{GPS} onto the same grid as k_{DAS}.

The product $k_{GPS}\hat{u}$ gives the distance and direction along the entire Red Line; distance north-south is the vertical component of $k_{GPS}\hat{u}$, and distance east-west is its horizontal component. We use a simple argument to show that $k_{GPS}\hat{u}$ gives distance and heading that is always more accurate than those from $d_{GPS}\hat{u}$. Conversion of the north-south and east-west distances from $k_{GPS}\hat{u}$ back to latitude λ_k and longitude θ_k (subscript k indicates a Kalman-derived coordinate) respectively gives us position on the same grid as k_{DAS}.

We then construct a three-component vector $[k_{DAS}, \lambda_k, \theta_k]$ using d_{DAS} and the GPS / speedometer logs from the first trip and from the first two cars (they have the least noise). The set of $[k_{DAS}, \lambda_k, \theta_k]$ vectors spanning the entire Red Line give, effectively, a look-up table whereby each DAS distance k_{DAS} corresponds directly to a unique $[\lambda, \theta]$ position independently of time. As long as the fibre and the Red Line remain stationary, we show that $[k_{DAS}, \lambda_k, \theta_k]$ gives $[\lambda, \theta]$ on the Red Line for any future DAS-measurement. We find that the accuracy of our $[k_{DAS}, \lambda_k, \theta_k]$ relationship is ± 30 m north-south and $\pm$ 60 m east-west and this compares well with ± 15 m north-south and $\pm$ 15 m east-west that we obtain from the GPS logs.

2. Materials and methods

A typical City of Calgary fibre installation is given in Fig. 4. The fibre is usually buried about 1 m below street level in a high-density polyethylene (HDPE) conduit encased in approximately 300 mm by 420 mm of permeable fill and backfilled to the street-level with soil. The permeable-fill and soil backfill are highly worked materials so we assume that the telecom fibre lies in the conduit within a homogeneous and well-consolidated medium.

A moving train generates a sub-surface soundfield of seismic waves and there is concern at first is that the fibre might be too deeply buried to register a DAS signal. However, even in seismically-unfriendly conditions like melting permafrost, Ferguson (2015) show that seismic waves traverse 1 m into the subsurface easily; and indeed, we capture a strong train-signal in this study.

2.1. Distributed Acoustic Sensing (DAS)

A Distributed Acoustic Sensing (DAS) system launches a finite-width pulse of laser light into telecom fibre rather than an encoded signal; a photodetector located at the same end of the fibre as the laser, registers the intensity I of the Rayleigh scattered light as a function of time t (Barnoski and Personick, 1978; Nakazawa et al., 1981; Rogers and Handerek, 1992). If the fibre is well-coupled to the subsurface, then

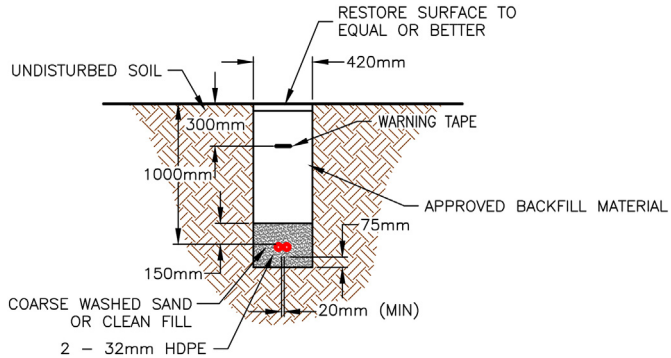

Fig. 4. A typical City of Calgary fibre installation standard (Moroz and McGhee, 2016, with permission). The fibre burial-depth of 1 m (m) in homogeneous backfill suggests good coupling between the fibre and the soundfield at street level.

street-level disturbances, like transportation noise, strain the fibre and I changes at points $x = t'/2/\nu$ along the fibre where ν is the speed of light in the fibre.

Following Rogers and Handerek (1992), for example, the Rayleigh scattered pulses from positions x_0 and x_1 along the fibre due to a source pulse at $t' = 0$ are

$$a_0(t') = r_0 e^{i\omega\left[\frac{2x_0}{\nu}+t'\right]} \tag{1}$$

and

$$a_1(t') = r_1 e^{i\omega\left[\frac{2x_1}{\nu}+t'\right]}, \tag{2}$$

where r_0 and r_1 are the amplitude of Rayleigh scattering from x_0 and x_1 respectively, and ω is the frequency of the laser. The scattered waveforms a_0 and a_1 combine at the photodetector as

$$a_r = a_0 + a_1, \tag{3}$$

where their combined intensity $I = a_r a_r^*$ is

$$I(x') = r_0^2 + r_1^2 + 2r_0 r_1 \cos\left(2\frac{\omega}{\nu}\varepsilon(x')\right), \tag{4}$$

for $x_0 \le x' \le x_1$ and $\varepsilon = x_1 - x_0$ (Rogers and Handerek, 1992).

If the fibre is undisturbed between one pulse and then a subsequent pulse at time Δt_P, where $0 \le t' < \Delta t_P$, the distance ε between Rayleigh scatter points x_0 and x_1 is constant and $I(\Delta t_P, x')$ for the second pulse is identical to $I(0, x')$ for the first pulse.

Strains applied to the fibre between the two pulses, however, change the distance ε between x_0 and x_1 and eq. 4 becomes time-variable for the second pulse according to

$$I(\Delta t_P, x') = r_0^2 + r_1^2 + 2r_0 r_1 \cos\left(2\frac{\omega}{\nu}\varepsilon(\Delta t_P, x')\right) \tag{5}$$

(Nakazawa et al., 1981; Rogers and Handerek, 1992).

For a sequence of pulses at times $t_P = \Delta t_P[0,1,2,\cdots]$ and a sequence of distances $x = \nu t'/2 \Rightarrow \Delta x[0,1,2,\cdots]$ along the fibre, I is proportional to the strains ε according to

$$I(t_P, x) \propto \cos\left(2\frac{\omega}{\nu}\varepsilon(t_P, x)\right). \tag{6}$$

Eq. 6 gives I variation entirely as a function of changes in ε for constant ω and ν; implementation of eq. 6 in actual DAS systems requires very stable lasers and fibres so that ω and ν remain constant. Also, eq. 6 is non-linear in ε so we restrict $2\omega/\nu\varepsilon(t_P, x) < \pi\nu/2\omega$ to avoid harmonic interference.

Table 1 gives the acquisition parameters of our Fotech Helios system. Distance interval Δx satisfies $\Delta x < w\nu/4$ to avoid aliasing the pulse; the number of samples per-pulse n_P satisfies $2[n_P - 1]\Delta x/\nu < \Delta t_P$ to capture I for all x for the current pulse prior to emitting the next pulse, and $2[n_P - 1]\Delta x/\nu = \max\{d_{DAS}\}$ restricts I to within the fibre length $0 \le x \le \max\{d_{DAS}\}$. The maximum fibre-length $\max\{d_{DAS}\}$ is a parameter that we estimate by trial-and-error.

Our gauge length (GL) and spatial pulse-width (w_s) satisfy $w_s > 0.56$ GL according to the Dean et al. (2017) criteria for vertical seismic profiling (VSP), where surface-waves rather than P-waves dominate our DAS soundfields (Krylov, 2017, for example). Our DAS fibre follows the Redline closely with some departures so we expect the dominant strain to propagate along the fibre, and for this case Dean et al. (2015) show that n *notch* frequencies f_n occur in the recorded DAS spectrum according to

$$f_n = n\frac{V_{sw}}{\mathrm{GL}}, \tag{7}$$

where V_{sw} is velocity. Surface-wave dispersion complicates the $f_n(\mathrm{GL})$ relationship and we use eq. 7 here in a qualitative sense only. Plots of the notch frequencies f_n for $1 \le n \le 3$ in Fig. 5 show that they lie within the seismic spectrum (0–100 Hz) for $[0 \le V_{sw} \le 500]$ m/s approximately. If $V_s < 500$ m/s (approximately) for the redline, then GL = 4 m might introduce significant noise in our DAS signal. Fig. 4 indicates homogeneity, however, in the medium beneath the Redline so V_{sw} and f_n, therefore, should be stationary. Stationary f_n distorts I in a consistent way and this suggests that strain local to the fibre causes the variation in I for $V_{sw} < 500$ m/s rather than GL-related noise. For $V_{sw} > 500$ m/s, we expect f_n to lie outside of the seismic frequency-band and not cause noise issues.

Intensity $I(t_P, x')$ in Fig. 5 corresponds to a portion (region Dal* in Fig. 9) of the the optical fibre buried beneath the Red Line between our DAS system at City Hall station ($x' = 0$) and Tuscany station ($x' \approx 19.5$ km). Strain associated with an eastbound train is visible as strong I beginning at 1120 s (Fig. 6, top-right corner); it decelerates to a stop at 13280 m where it remains stationary for about 30 s. At 1180 s the train accelerates and continues eastbound.

Noise contaminates $I(t_P, x')$ (Fig. 6) so we apply a filter operator L to $I(t_P, x')$ according to

$$I(t_P, d_{DAS}) = \langle L I(t_P, x)\rangle\ (t_P, d_{DAS}), \tag{8}$$

where L passes I values that are characteristic of Red Line trains along the fibre.

We then apply filter operator M to enhance train characteristics that are unique to specific distances d_{DAS} along the line according to

$$I(t, d_{DAS}) = \langle M I(t_P, d_{DAS})\rangle\ (t, d_{DAS}). \tag{9}$$

Note, filter operators L and M are usually proprietary to DAS system manufacturers so we don't reveal the specific details of our filters here.

Table 1
Table of DAS acquisition parameters. We choose acquisition parameters so that $\Delta x < w\nu/4$, $2[n_P - 1]\Delta x/\nu < \Delta t_P$ and $[n_P - 1]\Delta x = \max\{d_{DAS}\}$, where $max\{d_{DAS}\}$ is the apparent length of the fibre. Gauge length GL = 4 m ensures the spatial pulse width $w_s > 0.56$ GL following Dean et al. (2017).

Parameter	Value	Units	Description
w_t	80	ns	Temporal pulse-width
w_s	15.6509	m	Spatial pulse-width
η	1.5324		Fibre refractive index
ν	1.9564×10^8	m/s	Light speed in the fibre
GL	4	m	Gauge length
Δt_P	0.4	ms	Time between pulses
Δx	65.2	cm	Distance between intensity measurements
n_P	32,181		Number of intensity samples per pulse

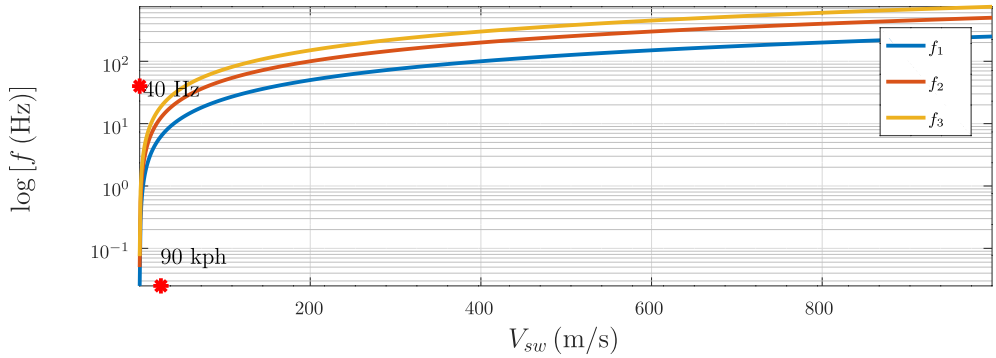

Fig. 5. Notch frequencies f_1, f_2 and f_3 for gauge lengths $GL = 4$. The horizontal axis spans the expected range of surface wave velocity. The train speed maximum of 90 kph is annotated as well as 40 Hz.

Please see Wiesmeyr et al. (2020) for an excellent summary of similar filter operators that they implement.

Fig. 7 shows the corresponding $I(t,x)$ for Fig. 6. The noise near 13,200 m corresponds to a pump station, and the remaining noise is associated with road traffic. Otherwise, I associated with the train is easily identifiable as linear intensity peaks.

2.1.1. d_{DAS} picks

At each distance d_{DAS} along the fibre, we compute in realtime a short-term, moving average along the time axis that we compare to a long-term average. We set the length of the short-term average to be slightly greater than the length of the train divided by the local speed of the train. Our long-term average is usually three times as long as the short-term average. A short-term / long-term ratio sufficiently greater than one indicates the presence of a train at that distance and time value is stored as a *pick* for that train for each d_{DAS}.

Picks associated with the eastbound train are annotated in Fig. 7. The picks overlap in distance at this station and as well as at the other stations (Figs. 10 and 11 for example). The overlaps combine with other noise phenomena to cause d_{DAS} error, and as we describe later in the paper, we use Kalman filtering to reconcile d_{DAS} for each train with its true position.

Our DAS distances (d_{DAS}) from each trip do not lie on the same time grid as the $[\lambda,\theta]$ positions from the GPS logs. Moreover, an ambiguity regarding which car corresponds to the maximum intensity of the DAS soundfield adds noise to our d_{DAS} picks. Application of a Kalman filter (Kalman, 1960) to d_{DAS} allows us to use the speedometer readings (α) to both reduce noise in the picks, and to interpolate d_{DAS} onto the same time grid as $[\lambda,\theta]$.

2.2. DAS interpolation by Kalman filter: theory

If the physical location of an object (one of our cars, for example) is known at time $t = 0$, then the distance $\mu(\Delta t)$ for that object at a later time $t = \Delta t$ is also known from this general series:

$$\mu(\Delta t) = \sum_{j=0}^{\infty} \frac{\mu^j(0)}{j!}\,\Delta t^j, \tag{10}$$

where $\mu^j(0) \leftrightarrow \partial^j/\partial t^j \mu(t)|_{t=0}$. The zeroth order term ($j = 0$) in eq. 10 is the starting distance $\mu(0)$ from the temporal origin. Order $j = 1$ corresponds to the distance travelled away from $\mu(0)$ during the interval Δt at a constant rate (speed), and $j = 2$ is distance travelled during Δt due to accelerations and so on for $j > 2$. Distance $\mu(\Delta t)$ is the *predicted* distance travelled by the object during the interval $t = \Delta t$ due to the physical-state variables μ^j evaluated at $t = 0$ (Kalman, 1960).

Of the distances $d_{DAS}(t)$ that we *pick* (measure directly) from the DAS soundfield (Fig. 7) in time t for a moving train, $d_{DAS}(\Delta t)$ is an observed distance at $t = \Delta t$ that we can compare to the predicted distance $\mu(\Delta t)$ in eq. 10.

Besides the GPS logs, only the speed logs $\alpha(0, \Delta t, t_2, \cdots, t_N)$ are available to us for each car so we predict the physical state during the interval Δt only to first order in eq. 10 as

$$\mu_\alpha(\Delta t) = \mu(0) + \Delta t\,\alpha(\Delta t). \tag{11}$$

The unknown 2*nd* order-and-above terms in eq. 10 plus measurement noise in α and Δt add noise to the μ_α prediction (eq. 11). Uncertainty in the DAS picks adds noise to d_{DAS} observations as we explore

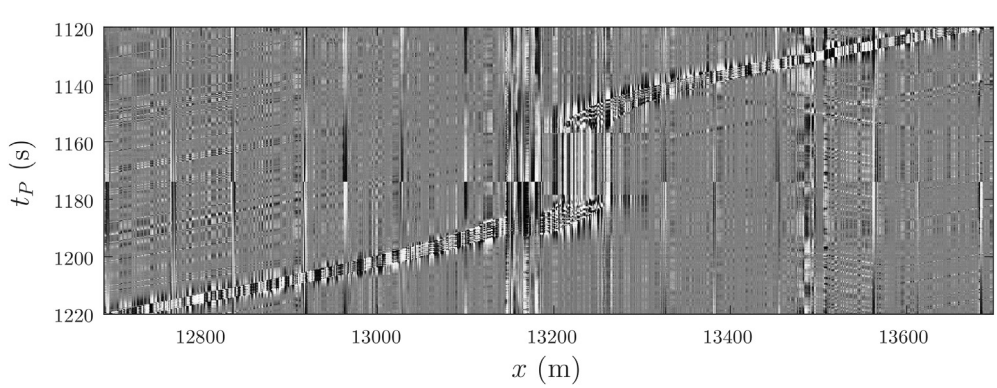

Fig. 6. Raw intensity $I(t_P,x)$ (eq. 6) for a train eastbound on the Red Line (region Dal* in Fig. 9). The train enters this region at ($t = 1120$ s, $x = 13500$ m) and stops at Dalhousie station ($x = 13280$ m) for approximately 30 s between $t = 1155$ s and $t = 11855$ s.

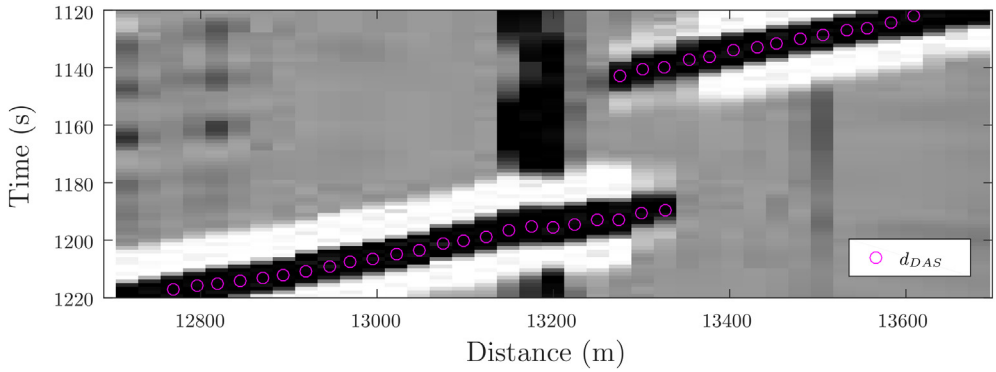

Fig. 7. Processed intensity data $I(t,d_{DAS})$ (eqs. 8 and 9) for the eastbound train of Fig. 6 (region Dal* in Fig. 9). Noise is reduced in this image and the train intensity is enhanced relative to Fig. 6. The DAS *picks* for d_{DAS} are indicated.

later on. Statistically, μ_α and d_{DAS} are different *expected values* for the true distance at Δt, and if the noise in μ_α and d_{DAS} and d_{GPS} is random, then variances σ_α^2 and σ_{DAS}^2 correspond to the expected values μ_α and d_{DAS} respectively (Kalman, 1960).

2.2.1. Effective expected-value k_{DAS}

Following Faragher (2012), if noise in the expected values μ_α (prediction) and d_{DAS} (observation) is random, then Gaussian functions $G(x; d_{DAS},\sigma_{DAS}^2)$ and $G(x;\mu_\alpha,\sigma_\alpha^2)$ centre on μ_α and d_{DAS} according to the general form

$$G(x;\mu,\sigma^2) = \frac{1}{\sigma^2\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{[x-\mu]^2}{\sigma^2}}, \tag{12}$$

where distance $x \equiv \Delta t\alpha$, and μ in eq. 12 takes on the expected value of μ_α or d_{DAS}. From statistics, the product of two Gaussians is also a Gaussian so the product

$$G(x;k_{DAS},\sigma_k^2) = G(x;d_{DAS},\sigma_{DAS}^2) \times G(x;\mu_\alpha,\sigma_\alpha^2), \tag{13}$$

centres on an effective expected-value k_{DAS} (with an effective variance σ_k^2). The maximum value of $G(x;k_{DAS},\sigma_k^2)$ (eq. 13) occurs at $x = k_{DAS}$ according to

$$0 = \frac{\partial}{\partial x} G(x;k_{DAS},\sigma_k^2) = x - k_{DAS}; \tag{14}$$

substitute $G(x;k_{DAS},\sigma_k^2)$ with the right-hand side of eq. 13 to get

$$\begin{aligned} 0 &= \frac{\partial}{\partial x} G(x;d_{DAS},\sigma_{DAS}) \times G(x;\mu_\alpha,\sigma_\alpha) \\ &+ G(x;d_{DAS},\sigma_{DAS}) \times \frac{\partial}{\partial x} G(x;\mu_\alpha,\sigma_\alpha) \\ &= \sigma_{DAS}^2[x - d_{DAS}] + \sigma_\alpha^2[x-\mu_\alpha]. \end{aligned} \tag{15}$$

Substitute k_{DAS} for x (eq. 14) in eq. 15 and then solve for k_{DAS} to get

$$k_{DAS} = \frac{\sigma_{DAS}^2\mu_\alpha + \sigma_\alpha^2 d_{DAS}}{\sigma_{DAS}^2 + \sigma_\alpha^2}. \tag{16}$$

The effective expected-value $k_{DAS} \to \mu_\alpha$ as $\sigma_{DAS}^2 \to \infty$ in eq. 16 (large uncertainty in the observation). In general, eq. 16 favours the prediction when σ_{DAS}^2 is large relative to σ_α^2. As $\sigma_\alpha^2 \to \infty$ (large uncertainty in the prediction), $k_{DAS} \to d_{DAS}$ so eq. 16 generally favours the observation d_{DAS} when σ_{DAS}^2 is small relative to σ_α^2. The expected value k_{DAS} from eq. 16 is equivalent to the optimal expectation value of a Kalman filter (Kalman, 1960) as is shown by Faragher (2012). Variances σ_α^2 and σ_{DAS}^2 are left to us to estimate empirically and/or analytically (Antoniou et al., 2007, for example).

The Kalman DAS-distance k_{DAS} in eq. 16 corresponds to the the first time-step Δt in the discrete α and d_{DAS} logs; the sample interval in the logs is not necessarily constant so for later times

$$(t_2 > \Delta t), (t_3 > t_2), \cdots, (t_N > t_{N-1}),$$

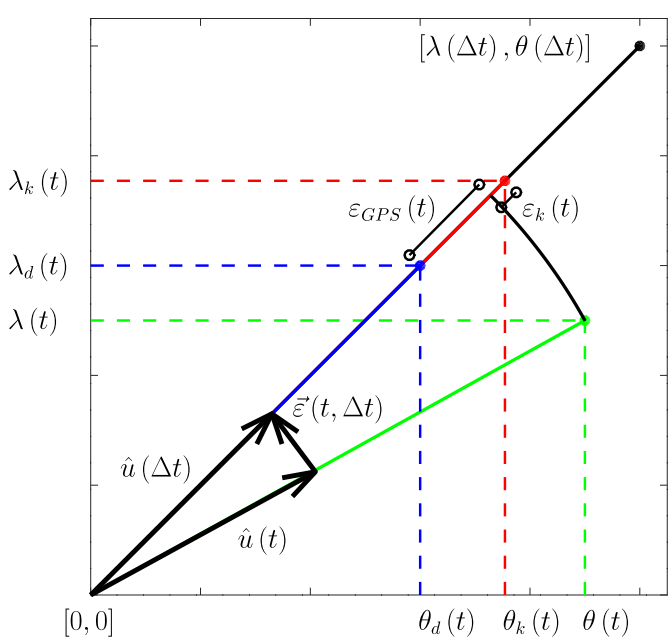

Fig. 8. Approximate positions $[\lambda_d(t),\theta_d(t)]$ (eq. 19) and $[\lambda_k(t),\theta_k(t)]$ (eqs. 23 and 24) compared to true position $[\lambda(t),\theta(t)]$. The origin [0,0] and position $[\lambda(\Delta t),\theta(\Delta t)]$ are observed positions (from GPS) and $0 < t < \Delta t$. Approximate heading-vector $\hat{u}(\Delta t)$ (eq. 25) and true heading-vector $\hat{u}(t)$ point from $[0,0] \to [\lambda(\Delta t),\theta(\Delta t)]$ and from $[0,0] \to [\lambda(t),\theta(t)]$ respectively. The difference between $\hat{u}(\Delta t)$ and true heading-vector $\hat{u}(t)$ is represented by the error vector $\vec{\varepsilon}(t,\Delta t)$. Distance errors $\varepsilon_{GPS}(t)$ and $\varepsilon_k(t)$ correspond to simple interpolation and interpolation by Kalman filter respectively with $\varepsilon_{GPS} \geq \varepsilon_k$.

Table 2
Departure and arrival times for our test train consisting of cars 2404 / 2411 / 2414 / 2437, December 142,018. Westbound departure from City Hall station arrives at Tuscany station approximately 30 min later.

Trip	Depart City Hall (UTC)	Arrive Tuscany (UTC)
1	16:03:33	16:32:09
2	18:06:26	18:34:53
3	20:06:13	20:34:16

Table 3
Table of Δt_{GPS} statistics (ms) for each car, all trips.

Car	min	max	mean	σ
2404	3	30,440	88	424
2411	4	28,900	58	219
2414	3	34,030	79	388
2437	2	28,140	76	354

where $t_0 = 0$, $t_1 = \Delta t$ and N is the total number of time-samples in the logs, eq. 16 becomes

$$k_{DAS}(t_n) = \frac{\sigma^2_{DAS}\mu_\alpha(t_n) + \sigma^2_\alpha d_{DAS}(t_n)}{\sigma^2_{DAS} + \sigma^2_\alpha} \tag{17}$$

for the nth log-sample ($0 \le n \le N-1$), where

$$\mu_\alpha(t_n) = k_{DAS}(t_{n-1}) + [t_n - t_{n-1}]\alpha(t_n). \tag{18}$$

Given a heading (λ and θ) log on the same time grid as k_{DAS}, then $k_{DAS}(t_n)$ and position $[\lambda(t_n), \theta(t_n)]$ correspond at time t_n. Due to the monotonic increase in k_{DAS}, and for as long as the fibre and the track remain stationary, position at future times $t > t_N$ is determined uniquely from k_{DAS}. In practice, however, d_{DAS} is acquired separately from position and speed so the time grids for d_{DAS} and $[\alpha, \lambda, \theta]$ are not the same. To reconcile the time grids, we interpolate d_{DAS} onto the typically finer $[\alpha, \lambda, \theta]$ time grid with linear interpolation and then compute k_{DAS} using eq. 17 on that finer grid.

As we discuss in detail later, we find in our logs that most of the $[\lambda, \theta]$ positions are invalid so the time-grid for α differs strongly from the time grid for $[\lambda, \theta]$. We must, therefore, interpolate $[\lambda, \theta]$ onto the same time-grid as d_{DAS} and α.

2.2.2. Unit heading-vector $\hat{u}_k$

We wish to interpolate positions $[\lambda, \theta]$ onto the same time-grid as the speedometer data α, but α are always positive valued, and so $[\lambda, \theta]$ cannot be Kalman filtered directly whilst preserving the direction of travel (the heading).

Instead, return to the first time-interval Δt in the logs and compute the approximate latitude at time t according to

$$\lambda_d(t) \approx \frac{\Delta\lambda(\Delta t)}{\Delta t} t + \lambda(0), \tag{19}$$

where $\Delta\lambda(\Delta t) = \lambda(\Delta t) - \lambda(0)$ and $0 < t < \Delta t$. Swap λ and θ in eq. 19 to get the corresponding longitude $\theta_d(t)$. The hypothetical example in Fig. 8 illustrates the accuracy of this approximation where, given the observed position $[\lambda(\Delta t), \theta(\Delta t)]$, the approximate position $[\lambda_d(t), \theta_d(t)]$ lies on the line defined by unit vector $\hat{u}(\Delta t)$ that joins $[\lambda(\Delta t), \theta(\Delta t)]$ to the origin. The actual position of the hypothetical car at time t is $\lambda(t) \neq \lambda_d(t)$ and $\theta(t) \neq \theta_d(t)$, and so observation noise plus the linear assumption in eq. 19 place this car some distance from its actual position as in Fig. 8.

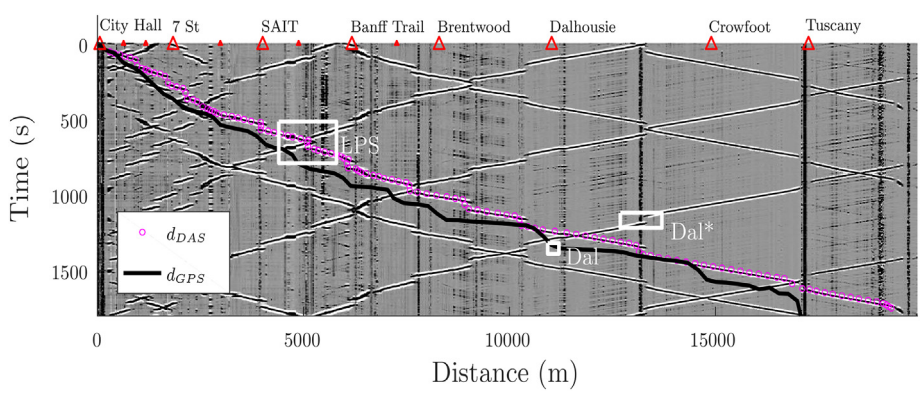

Fig. 9. A DAS soundfield. City Hall station provides the time and distance origin (0,0). Distances d_{DAS} are *picks* of the position of our test train during trip 2 (Table 2) and d_{GPS} is the corresponding distance from the GPS $[\lambda, \theta]$ coordinates (eq. A.1, A). Distance d_{DAS} becomes progressively larger than d_{GPS} due to the presence of slack fibre, fibre meanders (see Fig. 11) and other effects. The "Dal" region surrounds the GPS-derived distance (d_{GPS}) between City Hall and Dalhousie station (Fig. 17). The "LPS" region surrounds d_{GPS} to Lions Park station and d_{DAS} to the SAIT station (Fig. 10). The "Dal*" region surrounds d_{DAS} from City Hall to Dalhousie station (Figs. 6 and 7).

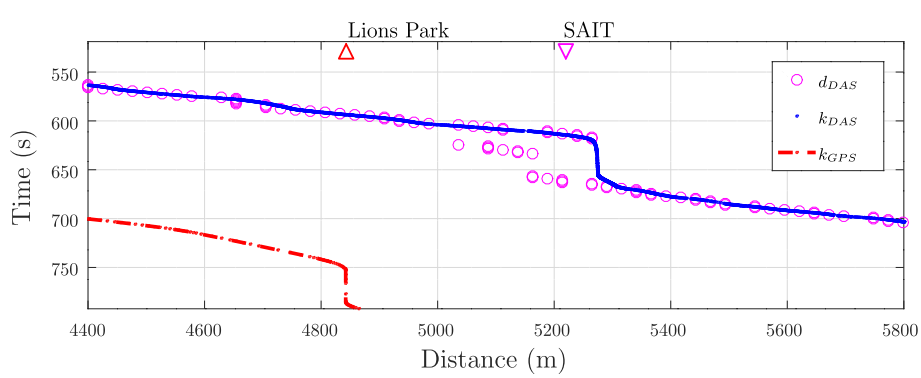

Fig. 10. Distance d_{DAS} and Kalman distances k_{DAS} (eq. 17) and k_{GPS} (eq. 21) near Lions Park station ("LPS" in Fig. 9) for the soundfield in Fig. 11. The inverted station marker ∇ indicates d_{DAS} for SAIT station. We give low weight ($\sigma^2_{DAS} \ll \sigma^2_\alpha$) in eq. 16 to d_{DAS} distances that don't increase with time so that $k_{DAS} \approx \mu_\alpha$ for example, for $660 < t < 700$ s and $5030 < d_{DAS} < 5270$ m in this Figure. Otherwise, $\sigma^2_{DAS} \approx \sigma^2_\alpha$ and k_{DAS} is a smooth interpolation of d_{DAS}.

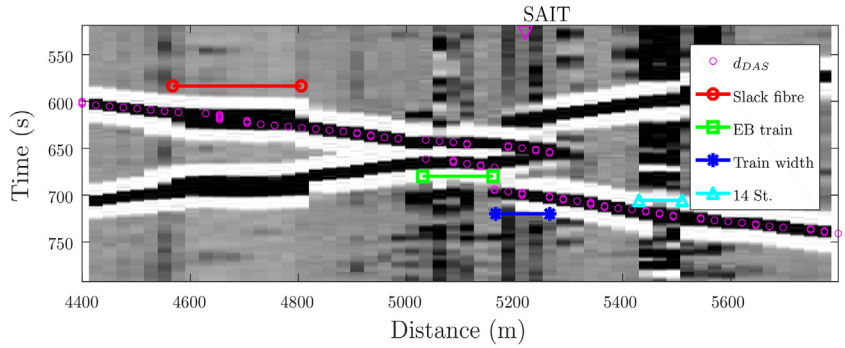

Fig. 11. DAS soundfield for region "LPS" in Fig. 9 annotated with noise sources. The DAS image of SAIT appears approximately 1180 m west of its true GPS position due to the cumulative effects of slack fibre, fibre meanders and other effects. A loop of slack fibre is indicated by the sudden increase in d_{DAS} at a constant time of 620 s. Interference from an eastbound train (EB) is seen as repeated d_{DAS} values as time increases. The width of the train (Train width) is approximately 100 m and road noise (14 St.) is apparent.

Define $\varepsilon_{GPS}(t)$ and $\vec{\varepsilon}(t,\Delta t)$ to be noise added to the true $[0,0]\rightarrow[\lambda(t),\theta(t)]$ distance $d(t)$ and to the true heading $\widehat{u}(t)$ respectively. If $d(t)$ is the true distance and $\widehat{u}(t)$ is the true heading, then the vector

$$d_{GPS}(t,\Delta t)\widehat{u}(\Delta t) = d(t)\widehat{u}(t) + \vec{\varepsilon}(t,\Delta t) + \varepsilon_{GPS}(t,\Delta t)\widehat{u}(\Delta t) \tag{20}$$

points from the origin [0,0] an approximate distance $d_{GPS}(t)$ (replace t_n with t in eq. A.1 and use $n = 2$) along the approximate heading $\widehat{u}(\Delta t)$ to the approximate position $[\lambda_d(t),\theta_d(t)]$ in Fig. 8.

The absence of an independent heading-log prevents our reducing $\vec{\varepsilon}(t,\Delta t)$, but we know that a computed Kalman distance $k_{GPS}(t)$ is closer to $d(t)$ than is $d_{GPS}(t)$. Substitute $t_n \leftarrow t$ in eq. 17 and swap subscript *GPS* for *DAS* to get

$$k_{GPS}(t) = \frac{\sigma^2_{GPS}\mu_\alpha(t) + \sigma^2_\alpha d_{GPS}(t)}{\sigma^2_{GPS} + \sigma^2_\alpha} \tag{21}$$

for $0 < t < \Delta t$, where

$$\mu_\alpha(t) = k_{GPS}(0) + t\alpha(t),$$

and then replace $d_{GPS}(t,\Delta t)$ with $k_{GPS}(t)$ in eq. 20 to get

$$k_{GPS}(t)\widehat{u}(\Delta t) = d(t)\widehat{u}(t) + \vec{\varepsilon}(t,\Delta t) + \varepsilon_k(t)\widehat{u}(\Delta t), \tag{22}$$

where $\varepsilon_k(t) < \varepsilon_{GPS}(t)$ for $0 < t < \Delta t$. Eq. 22 places the hypothetical car of Fig. 8 a distance $k_{GPS}(t)$ along the line defined by $\widehat{u}(\Delta t)$ to the position $[\lambda_k(t),\theta_k(t)]$. Eqs. 20 and 22 differ only in the error terms $\varepsilon_k(t)$ and $\varepsilon_{GPS}(t)$, and because $\varepsilon_k(t) < \varepsilon_{GPS}(t)$, position $[\lambda_k(t),\theta_k(t)]$ for the car is closer to true position as in Fig. 8.

Let $[0,\tau_1,\tau_2,\cdots,\tau_m,\cdots,\tau_{M-1}]$ be the coarse $[\lambda,\theta]$ time-axis, and let $[0, t_1,t_2,\cdots,t_n,\cdots,t_N]$ be the finer α time-axis ($M < N$), then $\Delta t \rightarrow \tau_{m+1} - \tau_m$ and the interpolated position $[\lambda,\theta]$ at the nth time-sample t_n is

$$\lambda_k(t_n) = \frac{\partial\lambda}{\partial x}k_{GPS}(t_n)\widehat{u}(\tau_{m+1}-\tau_m)\cdot\widehat{i}, \tag{23}$$

where $k_{GPS}(t_n) = k_{GPS}(t)|_{tn}$ (eq. 22) and

$$\theta_k(t_n) = \frac{\partial\theta}{\partial x}k_{GPS}(t_n)\widehat{u}(\tau_{m+1}-\tau_m)\cdot\widehat{j}, \tag{24}$$

where $\tau_{m-1} < t_n < \tau_m$. Operators $\partial\lambda/\partial x$ and $\partial\theta/\partial x$ in eqs. 23 and 24 are the reciprocals of eqs. A.4 and A.5 respectively evaluated at $j = m$ and $j = m + 1$ and averaged. We then compute the unit heading-vector according to

$$\widehat{u}(\tau_{m+1}-\tau_m) = \frac{\vec{u}(\tau_{m+1}-\tau_m)}{\left|\vec{u}(\tau_{m+1}-\tau_m)\right|}, \tag{25}$$

where

$$\vec{u}(\tau_{m+1}-\tau_m) = \frac{\partial x}{\partial\lambda}[\lambda(\tau_{m+1})-\lambda(\tau_m)]\widehat{i} + \frac{\partial x}{\partial\theta}[\theta(\tau_{m+1})-\theta(\tau_m)]\widehat{j}, \tag{26}$$

with $[\partial x/\partial\lambda, \partial x/\partial\theta]$ (Eqs. A.4 and A.5) evaluated at $j = m + 1$ and $j = m$ and then averaged. Given k_{GPS} (Eq. 21), then, position $[\lambda(t_n),\theta(t_n)]$ for

Table 4
Table of GPS coordinate totals and averages for all cars and trips. Of the 631,847 coordinate $[\lambda,\theta]$ pairs, 2.2% are zero and only 1.1% are non-zero and unique. Approximately 93.8% of the coordinates correspond a moving train so only P_{GPS}= 1.2% of the coordinates of the moving train are non-zero and unique. The average distance $\overline{\Delta d_{GPS}}$ between non-zero, unique coordinates is 55.7 m.

Trip	Car	N $[\lambda,\theta]$	N [0,0]	Unique	Moving (%)	P_{GPS} (%)	$\overline{\Delta d_{GPS}}(m)$
1	2404	44,569	902	561	92.5	1.3	55.5
	2411	72,635	1594	579	93.8	0.8	54.6
	2414	50,693	998	558	94.5	1.1	57.1
	2437	58,756	1628	572	93.2	1	54.6
2	2404	43,981	798	558	93	1.3	56.3
	2411	63,906	1315	583	94	0.9	54.4
	2414	46,052	795	567	94.1	1.3	56.3
	2437	51,590	1453	564	93.8	1.1	55.8
3	2404	43,157	934	561	93.5	1.4	56.2
	2411	63,614	1437	574	94.2	0.9	55.2
	2414	49,551	1106	563	94.9	1.2	56.6
	2437	43,343	1203	557	94.5	1.3	56.6
Total		631,847	14,163 (2.2%)	6797 (1.1%)			
Mean					93.8	1.2	55.7

Table 5
Table of k_{DAS} coordinate totals and averages for all Cars and Trips. DAS picks are available for the Red Line outbound journey only so the total number of $[\lambda,\theta]$ pairs ($N^\dagger$) is about half the number of d_{DAS} pairs (Table 4). Of the 302,634 coordinate (λ,θ) pairs, none are zero valued and 91.1% are unique values. Approximately 93.8% of the coordinates correspond a moving train so ε_{DAS}= 97.2% of the coordinates of the moving train are non-zero and unique. The average distance $\overline{\Delta k_{DAS}}$ between non-zero, unique coordinates is 0.7 m.

Trip	Car	$N^\dagger[\lambda,\theta]$	N [0,0]	Unique	Moving (%)	ε_{DAS} (%)	$\overline{\Delta k_{DAS}}(m)$
1	2404	22,141	0	19,706	92.5	96.2	0.9
	2411	35,761	0	31,878	93.8	94.9	0.5
	2414	25,226	0	23,109	94.5	96.8	0.7
	2437	27,338	0	24,890	93.2	97.6	0.7
2	2404	21,147	0	19,457	93	98.8	0.9
	2411	29,680	0	26,887	94	96.3	0.6
	2414	21,968	0	20,215	94.1	97.7	0.9
	2437	27,242	0	24,783	93.8	96.9	0.7
3	2404	20,547	0	18,864	93.5	98.1	0.9
	2411	30,098	0	27,510	94.2	96.9	0.6
	2414	21,714	0	20,156	94.9	97.7	0.9
	2437	19,772	0	18,353	94.5	98.1	1
Total		302,634	0 (0%)	275,808 (91.1%)			
Mean					93.8	97.2	0.7

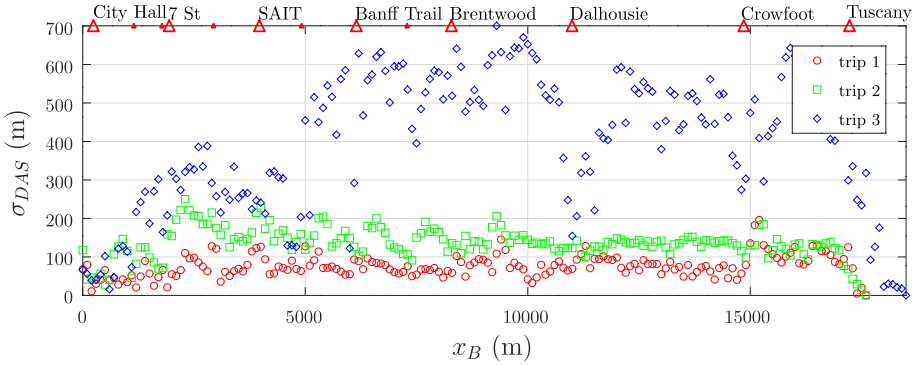

Fig. 12. Standard deviation (eq. 27, $\Delta B = 100$ m) of k_{DAS} for trips 1, 2 and 3. $\sigma_{DAS} \approx 100$ m for trip 1 is the lowest of the 3 trips overall and suggests that k_{DAS} for trip 1 has the least noise. Departure times in local time correspond to 10:00, 12:00 and 14:00 respectively, and perhaps building ridership and traffic volume causes the noise increase through the day.

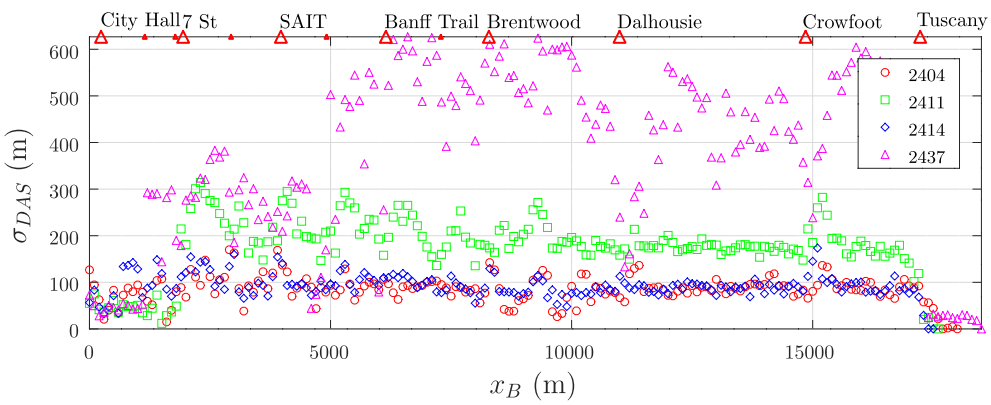

Fig. 13. Standard deviation (σ) of k_{DAS} within 100 m bins: common car. The σ values of approximately the bin-width for cars 2404 and 2414 suggest that the DAS picks correspond most closely to the front of the train.

time t_n is computed with Eqs. 23 and 24 where $\widehat{u}(\tau_m - \tau_{M-1})$ is computed with Eqs. 25 and 26.

3. Data

Our test train consists of four cars (2414, 2404, 2411 and 2437) coupled together for three trips (1, 2, and 3) from City Hall station to Tuscany station along the westbound journey of the Red Line (Fig. 1). Car 2414 leads, cars 2404 and 2411 make up the middle of the train, and car 2437 trails during the westbound leg of each trip. As is summarized in Table 2, our train leaves City Hall (CH) station at 16:03:33 on December 14, 2018 during trip 1 arriving approximately 29 min later at Tuscany (TUS) station (16:32:09). Elapsed time for trip 2 is also about 29 min, and trip 3 is shorter by approximately 1 min.

Data logs from each car consist of latitude and longitude $[\lambda, \theta]$ plus speedometer α. We find that for each trip and for each car, the numerical values for the time intervals vary, and only about 1.2% of the $[\lambda, \theta]$ positions (see Table 3 are valid.

Fig. 9 shows the thresholded ratio overlaid on top of the original band-limited energy.

3.1. Optimal Kalman distance k_{DAS}

Fig. 9 plots the observed-distance d_{DAS} from the DAS soundfield near Lions Park station (Fig. 11 and "LPS" in Fig. 9) acquired during trip 1 along with the Kalman-distances k_{DAS} (eq. 17) and k_{GPS} (eq. 21). The d_{DAS} picks between $660 < t < 700$ s and $5030 < d_{DAS} < 5270$ m increase in time without a corresponding increase in distance, and this is due to interference from a passing train. During the k_{DAS} computation we give low weight ($\sigma^2_{DAS} \ll \sigma^2_\alpha$) to d_{DAS} wherever they appear to reverse so that $k_{DAS} \approx \mu_\alpha$ (eq. 11) for these erroneous d_{DAS} values. For forward-moving d_{DAS} values, we set $\sigma^2_{DAS} \approx \sigma^2_\alpha$ so that k_{DAS} intersects only those d_{DAS} picks that increase in both time and distance and gives smooth interpolation between as is evident in Fig. 10.

3.1.1. k_{DAS} noise: common trip

Travel time for trips 1, 2 and 3 is approximately 30 min (Table 2) so we expect minimal variation in the noise of the physical state (eq. 11) between trips due to unknown accelerations. Our trips occur at 10:00, 12:00 and 14:00 respectively (UTC - 6 h, Table 2) in local time, however, and perhaps an increase in ridership and road traffic near the Red Line

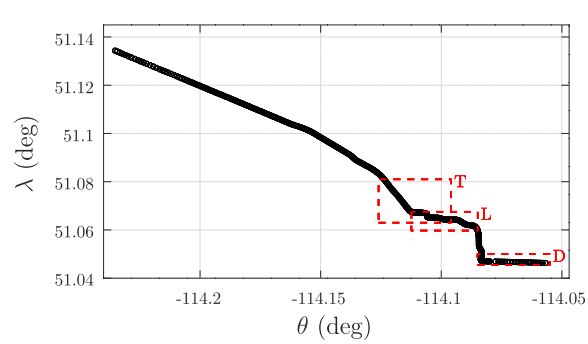

Fig. 14. GPS coordinates (cars 2404, 2411, 2414, and 2437) for trip 1. The regions marked "T" and "D" correspond to, respectively, tunnels (Fig. 15) and the city centre (Fig. 16). Region "L" corresponds to Fig. 3.

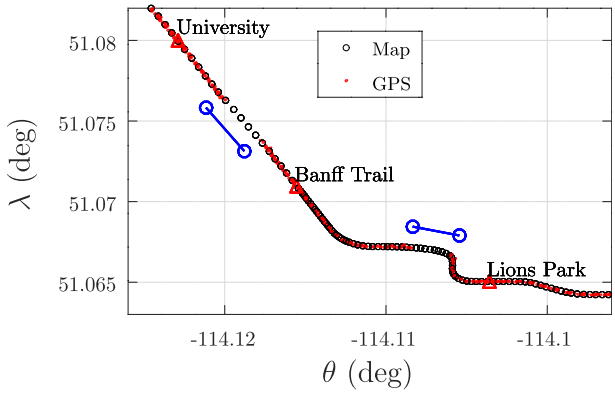

Fig. 15. GPS coordinates (cars 2404, 2411, 2414, and 2437) for trip 1 between University station and Lions Park station. Coordinates (black dots) hand picked from satellite imagery align with the GPS coordinates and connect the tunnel openings (indicated by circles). The GPS coordinates (red dots) drop out in the tunnels.

between mid-morning and mid-afternoon is responsible for the increase in noise through the day.

So that we may estimate the relative noise experienced by our test train during each trip, we merge the 12 k_{DAS} distances for cars 2404, 2411, 2414, and 2437 into three sets of $k_{DAS;\,T}$ $(1 \le T \le 3)$ distances and compute the standard deviation for each trip T according to

$$\sigma_{DAS,T}(x_B) = \sqrt{\frac{1}{N_B - 1}\sum_{j=1}^{N_B}\left[k_{DAS;\,T,j} - \bar{k}_{DAS;\,T}\Big|_{x_B}\right]^2}, \tag{27}$$

where x_B is the distance to the centre of a bin that is ΔB wide according to

$$x_B - \frac{\Delta B}{2} < k_{DAS;\,T,j} \le x_B + \frac{\Delta B}{2},$$

with N_B and $\bar{k}_{DAS;\,T}$ as the number and mean value respectively of the d_{DAS} values from cars 2404, 2411, 2414 and 2437 that fall withing the ΔB wide-bin centred on x_B.

Fig. 12 plots $\sigma_{DAS,\,T}(x_B)$ for $\Delta B = 100$ m. Note, $\sigma_{DAS,\,T}$ plots beyond Tuscany station because the fibre is longer than the Red Line due to meanders, loops of slack fibre and other effects. Trip 1 has the lowest $\sigma_{DAS;\,1}$ overall ($\sigma_{DAS;\,1} \approx 100$m). Trip 2 $\sigma_{DAS;\,2} \approx 150$m is noisier than trip 1 and trip 3 [$100 < \sigma_{DAS;\,3} < 700$ m] is noisiest. Trips 1, 2 and 3 correspond to local departure times of 10:00, 12:00 and 14:00 (Table 2) respectively so the $\sigma_{DAS,\,T}$ increase between mid-morning ($T = 1$), mid-day ($T = 2$) and mid-afternoon ($T = 3$) probably reveals a pattern of increased ridership and/or traffic noise.

3.1.2. k_{DAS} noise: common car

The peak intensity (the DAS pick) in the DAS soundfield corresponds to a position along the length of the train that is somewhat ambiguous. This ambiguity is especially evident at stations with trains decelerating, remaining stationary while passengers embark and disembark, and then accelerating towards the next station. The DAS picks near SAIT station in Fig. 10, for example, follow the lead car 2414 until the train stops at approximately 650 s. At 700 s the train sets off and the DAS picks return from following (erroneously) an eastbound train (EB train, Fig. 11). Rather than resume their correspondence with lead car 2414, we infer from the Train width (~ 100 m) annotation in Fig. 11 that the picks follow the trailing car 2437 for some time after 700 s. The ongoing correspondence between the DAS picks and the cars in Fig. 10 becomes unclear as the train accelerates (a similar source of error is identified in (Papp et al., 2016)).

So that we may estimate the relative noise experienced by each car, we merge the 12 k_{DAS} estimates into four sets of $k_{AS;\,C}$, were $1 \le C \le 4$, and

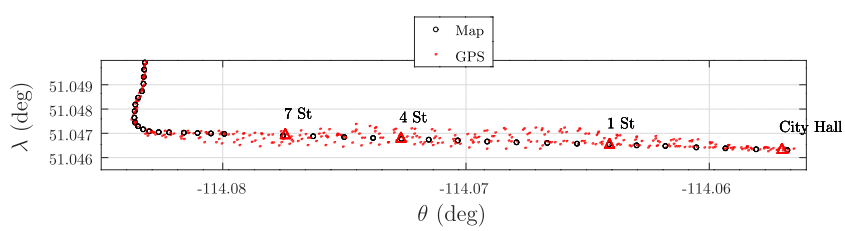

Fig. 16. All GPS coordinates (red dots) in the city centre (region "D", Fig. 14), trip 1. Hand-picked coordinates (black dots) from satellite imagery follow the northbound LRT track. The GPS coordinates and the picked coordinates align on the bridge (heading north) west of 7 St. station. Alignment becomes more vague between the bridge and City Hall station.

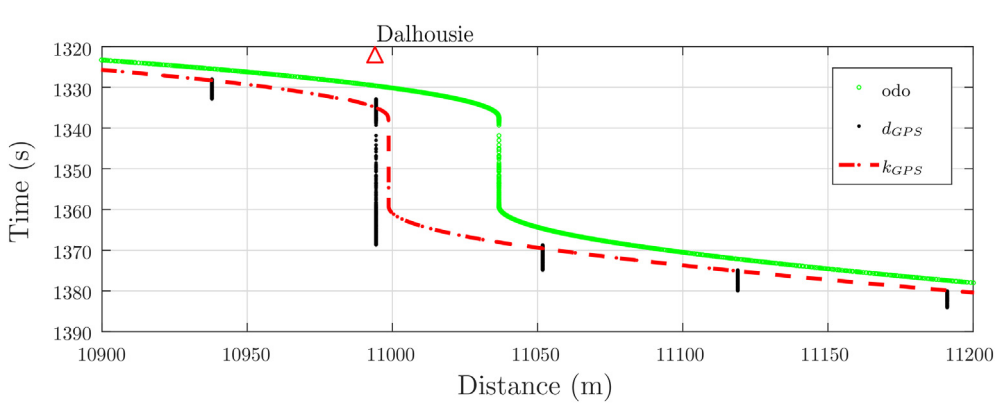

Fig. 17. Observed-distance d_{GPS} (eq. A.1) and optimal Kalman-distance k_{GPS} (eq. 21) compared to odometer (odo) near Dalhousie station ("Dal" in Fig. 9) for car 2411, trip 1. Of the hundreds of d_{GPS} values in this image only 5 are unique and non-zero. Equal weight $\sigma^2_{GPS} \approx \sigma^2_\alpha$ in eq. 21 for unique, non-zero d_{GPS} gives $k_{GPS} \approx d_{GPS}$ at the corresponding times. For times between the unique, non-zero d_{GPS} we set $\sigma^2_{GPS} \ll \sigma^2_\alpha$ in eq. 21 so that $k_{DAS} \approx \mu_\alpha$.

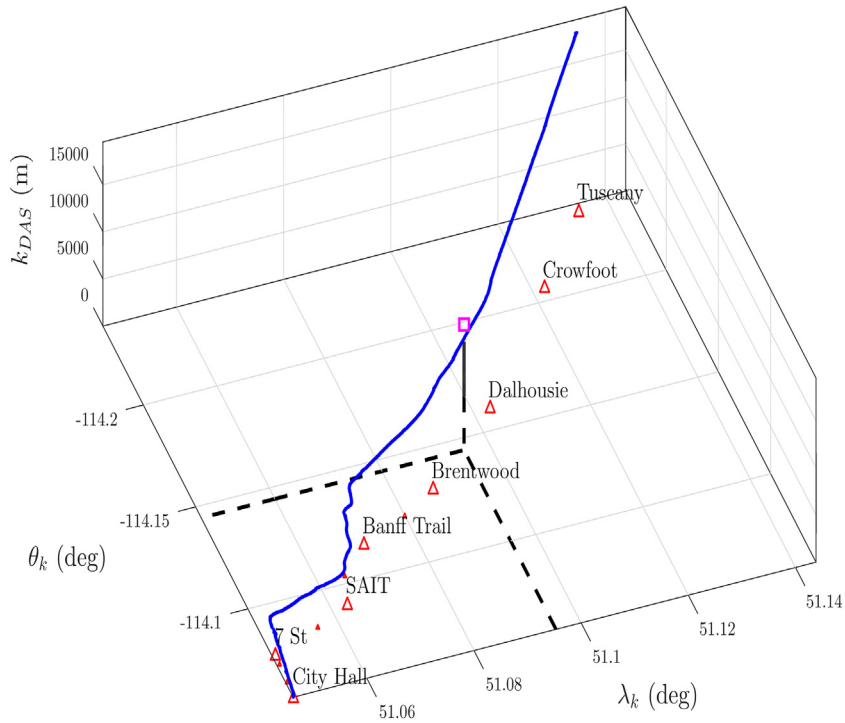

Fig. 18. Distance k_{DAS} verses position $[\lambda_k, \theta_k]$. A hypothetical train is indicated by the small box annotated midway between Brentwood and Dalhousie stations a DAS distance of $k_{DAS} =$ 13316 m. Its position is [51.09531, −114.1446] degrees.

substitute $k_{DAS;\,C,\,j}$ for $k_{DAS;\,T,\,j}$ and $\overline{k}_{DAS;\,C}$ for $\overline{k}_{DAS;\,T}$ in eq. 27 to compute $\sigma_{DAS,\,C}(x_B)$ for each car. Bin distance and width x_B and ΔB respectively are the same as in eq. 27 with

$$x_B - \Delta\frac{x_B}{2} < k_{DAS;\,C,j} \le x_B + \frac{\Delta B}{2},$$

where $1 \le j \le 3$. Using $\Delta B = 100$ m we find that $\sigma_{DAS} \approx 100$ m for car 2414 (the lead car) and car 2404 (the second car) in Fig. 13. Car 2411 is noisier ($\sigma_{DAS} \approx 200$ m), and the trailing car 2437 is the noisiest ($0 < \sigma_{DAS} < 600$ m).

The merged k_{DAS} distances for cars 2404, 2411, 2414 and 2437 are functions of the same prediction variances σ_α^2 and observation variances σ_{GPS}^2 so observation noise in the DAS picks probably causes the σ_{DAS} variation between cars that we see in Fig. 13. We conclude that the maximum intensity in the DAS soundfield corresponds to the lead cars 2414 and 2404 for most of the Red Line.

3.2. Optimal Kalman distance k_{GPS}

Fig. 14 maps the GPS coordinates from all cars during trip 1, and regions "D" and "T" correspond to zones where they are either inaccurate or dropout completely. (Region "L" corresponds to Fig. 3 in which we see that the fibre follows the LRT track with the exception of a meander.) Fig. 15 is a zoom in on region "T" and the spans of two tunnels are indicated by small circles; dark dots indicate handpicked coordinates from satellite imagery for reference. The GPS coordinates (lighter coloured dots) in this Figure track the handpicked coordinates until tunnel entrances and exits whereupon they scatter slightly, and subsequently dropout entirely inside the tunnels. Fig. 16 corresponds to region "D" on Fig. 14 in the city centre. The height and high density of office towers in the Downtown core scatter GPS coordinates.

Besides these GPS dropouts and scattering, we find overall, that most of the GPS coordinates are repeats of previous values. Fig. 17, for example, plots the observed-distance d_{GPS} (eq. A.1) and the optimal Kalman-distance k_{GPS} (eq. 21) against the odometer distance (odo) for the region

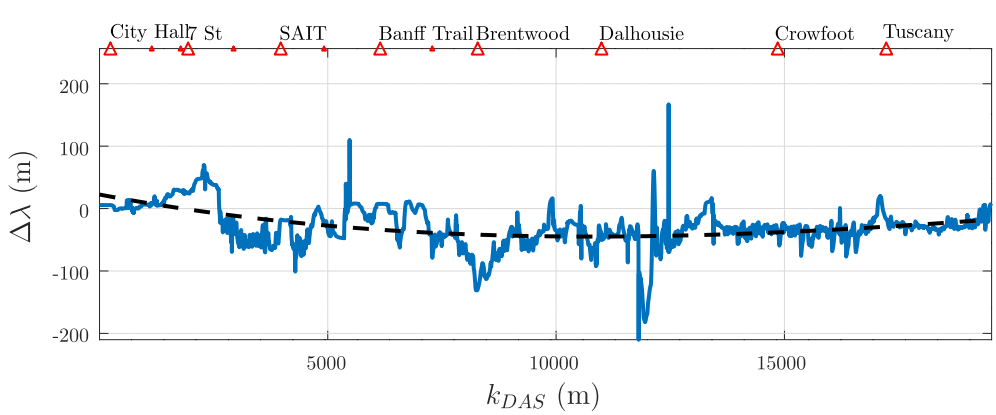

Fig. 19. The difference in λ (in m) estimated from k_{DAS} (eq. 29) between cars 2404 and 2414 during trip 1.

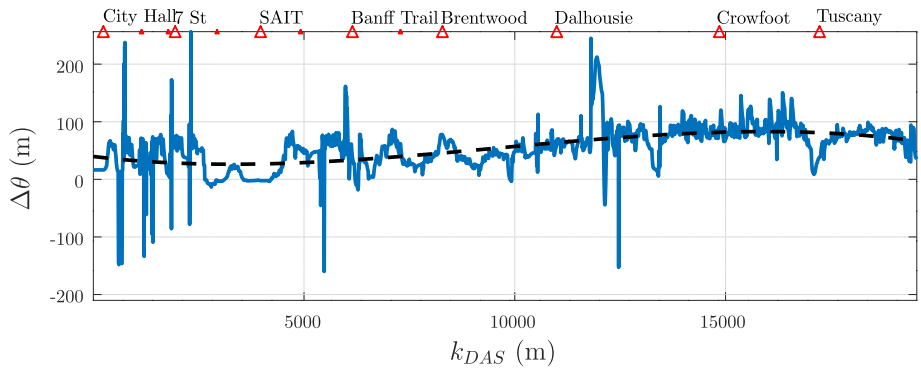

Fig. 20. The difference in θ (in m) estimated from k_{DAS} (eq. 30) between cars 2404 and 2414 during trip 1.

near Dalhousie station ("Dal" in Fig. 9). Of the hundreds of d_{GPS} in this image, only five are non-zero and unique and we find that this is typical as we discuss in detail later on.

For our k_{GPS} calculation, we set $\sigma^2_{GPS} \approx \sigma^2_\alpha$ in eq. 21 for unique, non-zero d_{GPS} values; this gives $k_{GPS} \approx d_{GPS}$ at the corresponding times. Where d_{GPS} is zero or a repeated value, we set $\sigma^2_{GPS} \ll \sigma^2_\alpha$ so that $k_{GPS} \approx \mu_\alpha$ at times that lie between unique, non-zero d_{GPS} values. As Fig. 16 shows, k_{GPS} intersects the unique d_{GPS} distances and it has a similar smooth-shape as the odometer log between. Note also that k_{GPS} remains constant while the train is at the station as expected.

3.2.1. k_{GPS} noise: Zeros and repeats

We find significant variation in the sample rate Δt_{GPS} of the time log for all cars in the test train for all trips. As a demonstration, statistics for Δt_{GPS} for all cars in trip 1 are given in Table 3. The maximum values correspond to GPS dropouts in the tunnels (see Fig. 15) and the minimum values correspond to open sky regions where numerous satellites are in view. The mean values and standard deviations indicate that Δt_{GPS} is never consistent. The average speed of our train is approximately 80 kph so the mean values of Δt_{GPS} suggest a spatial interval Δd_{GPS} of less than 2 m between $[\lambda, \theta]$ coordinates.

Though the Kalman filter accommodates the variable Δt_{GPS}, we find that more than 98% of the GPS positions $[\lambda, \theta]$ are either zero-valued or non-unique (Fig. 17 and Table 4) - the effective average-spacing in d_{GPS} is actually much greater than 2 m. At Dalhousie station in Fig. 17 for example, $d_{GPS}(t)$ is stationary as expected while the train is stopped beginning at approximately 1332 s. According to the odometer (odo), however, the train stops slightly later at approximately 1336 s and then departs at 1360 s - 9 s sooner than is indicated by d_{GPS}. The large gaps between unique d_{GPS} values in Fig. 17 suggests that the $[\lambda, \theta]$ sampling is much coarser than is indicated in Table 3. Inspection of Fig. 17 suggests that about 50 m spans each unique GPS distance and this is supported by the data given in Table 4.

Table 4 gives totals and averages for the $[\lambda, \theta]$ logs for each car and trip. Within each trip, cars 2404, 2414, and 2437 acquire similar numbers of $[\lambda, \theta]$ values ($N[\lambda, \theta]$~ 50 k), and 2411 acquires some tens of thousands more (60 k <$N[\lambda, \theta]$< 75 k). On average, slightly more than 2% of the $[\lambda, \theta]$ values equal zero {$[0, \theta]$, $[\lambda, 0]$ or $[0, 0]$} owing mostly to dropouts in the city centre and the tunnels. Of the non-zero values that remain, slightly more than 1% are unique - the remainder appear to be repeats of previous positions. We do expect repeated $[\lambda, \theta]$ values at stations - stationary trains log the same $[\lambda, \theta]$ until they set of again. The speedometer logs suggest, however, that our train travels for slightly less than 94% of its time in service. We estimate, then, that only about 1.2% (1.2 ≈1.1 ÷ 93.8 × 100, Table 4) of our unique non-zero $[\lambda, \theta]$ values correspond to moving trains; the effective average-spacing $\overline{\Delta d}_{GPS}$ between valid coordinates then is

$$\overline{\Delta d}_{GPS} \approx \frac{1}{12} \sum_{j=1,k=1}^{N_{j,k}} \frac{P_{GPS;j,k}}{100} \frac{d_{j,k}}{N_{j,k}}. \tag{28}$$

So from the numerical values in Table 4, we find that valid $[\lambda, \theta]$ coordinates are logged at a rather coarse, average spacing of $\overline{\Delta d}_{GPS} \approx$ 55.7 m (eq. 28). This value agrees well with the 50 m spacing seen in Fig. 17, and it is much greater than the 2 m spacing suggested by the time-variable sample rate from the discussion above.

4. Results

We use the *kalman* and *ss* functions from the Octave Eaton et al. (2019) *control* package to interpolate distances d_{DAS} (the DAS picks, Figs. 9, 10 and 11) and d_{GPS} (Eq. A.1) into distances k_{DAS} (eq. 17) and

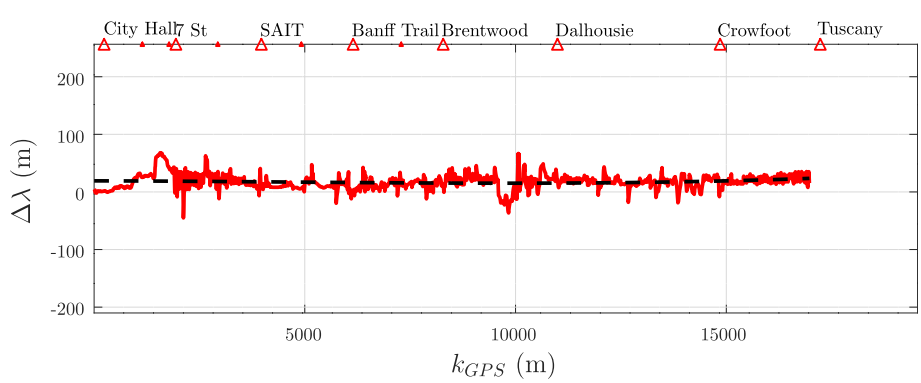

Fig. 21. The difference in λ (in m) estimated from d_{DAS} (replace k_{DAS} with k_{GPS} in eq. 29) between cars 2404 and 2414 during trip 1.

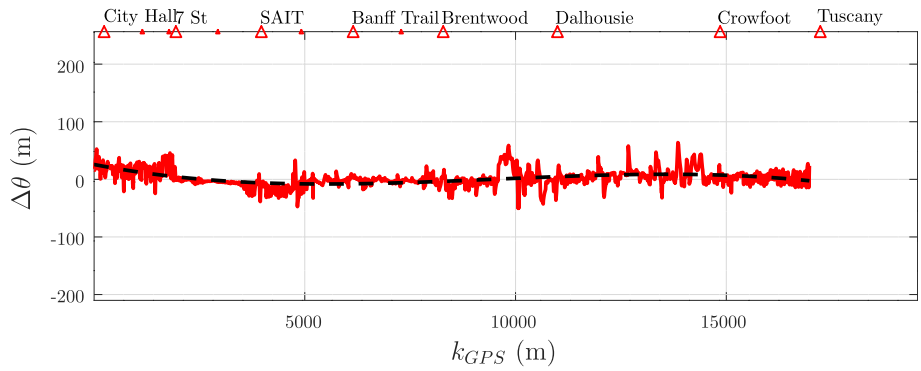

Fig. 22. The difference in θ (in m) estimated from d_{DAS} (replace k_{DAS} with k_{GPS} in eq. 30) between cars 2404 and 2414 during trip 1.

k_{GPS} (Eq. 21), respectively. We then compute the unit-heading vector $\widehat{u}$ (Eqs. 25 and 26) to provide the map between k_{GPS} and position $[\lambda_k, \theta_k]$ (Eqs. 23 and 24). The resulting distances k_{DAS} and k_{GPS} plus position $[\lambda_k, \theta_k]$ all reside now on the same time grid, and a distance $k_{DAS}(\tau)$ recorded at time τ from within our test data corresponds directly to position $[\lambda_k(\tau), \theta_k(\tau)]$.

DAS picks are available only for the westbound journey of the Red Line so the four cars and three trips of our test train contribute slightly less than 1/2 the number of k_{DAS} values ($N^{\dagger} = 302634$, Table 5) when compared to the original round-trip coordinates ($N = 631847$, Table 4). As is shown in Table 5, valid k_{DAS} values make up nearly 100% of those positions that correspond to the moving trains, and the average distance between values is approximately 1 m.

Each distance k_{DAS} along the Red Line has a unique position - for as long as the fibre and the Red Line do not move, we drop the temporal relationship between k_{DAS} and $[\lambda_k, \theta_k]$ and Fig. 18 illustrates a plot of their time-independent relationship. As is discussed previously, trip 1 and cars 2414 and 2404 (the lead car and second car respectively) have the least noise and Fig. 17 plots their combined k_{DAS} and $[\lambda_k, \theta_k]$ values. As Fig. 18 shows, the Red Line heads east from City Hall station ($k_{DAS} = 0$ m) until -114.07 degrees longitude where it heads north to SAIT station. From SAIT to Banff Trail station, the Red Line meanders north-west and then heads steadily north-west until Tuscany station at the end of the line ($k_{DAS} = 19534$ m).

The small box annotated in Fig. 18 midway between Brentwood and Dalhousie stations represents a hypothetical train detected at a DAS distance of 13,316 m in the present day, for example. The nearest k_{DAS} values in the curve in Fig. 18 are [13314, 13325] m; only 11 m separates these points so simple interpolation gives us the position of the hypothetical train as [51.09531, −114.1446] degrees.

4.1. Accuracy

Figs. 12 and 13 suggest that cars 2404 and 2414 correspond most closely to the DAS pick-time, and that they correspond best during trip 1. The accuracy of our $k_{DAS}(\lambda, \theta)$ relations, then, has meaning when we compute the difference of the estimated λ and θ between cars 2404 and 2414 from trip 1 for a set of the same distances k_{DAS}. The differences given in Figs. 19 and 20 are computed in units of metres according to

Table 6
Table $\Delta\lambda$ and $\Delta\theta$ statistics (m) from k_{DAS} for cars 2404 and 2414, trip 1.

	min	max	mean	σ
$\Delta\lambda$	−223.2	166.3	−28.7	32.8
$\Delta\theta$	−168.1	266.3	54.6	36.4

Table 7
Table of $\Delta\lambda$ and $\Delta\theta$ statistics (m) from k_{GPS} for cars 2404 and 2414, trip 1.

	min	max	mean	σ
$\Delta\lambda$	−42.6	68.1	17.2	12.7
$\Delta\theta$	−49.7	70.7	1.9	14.3

$$\Delta\lambda(k_{DAS}) = \frac{\partial x}{\partial \lambda}[\lambda_{2404}(k_{DAS}) - \lambda_{2414}(k_{DAS})], \tag{29}$$

and

$$\Delta\theta(k_{DAS}) = \frac{\partial x}{\partial \theta}[\theta_{2404}(k_{DAS}) - \theta_{2414}(k_{DAS})], \tag{30}$$

with $\partial x/\partial\lambda$ (Eq. A.4) and $\partial x/\partial\theta$ (Eq. A.5) evaluated at the midway points between $\lambda_{2404}(k_{DAS})$ and $\lambda_{2414}(k_{DAS})$, and between $\theta_{2404}(k_{DAS})$ and $\theta_{2414}(k_{DAS})$ respectively.

Table 6 summarizes some statistics for $\Delta\lambda$ and $\Delta\theta$. The range between the minimum and maximum values is 100's of metres, and the average values and standard deviations suggest that most values (one standard deviation from the mean) lie between approximately $-60 < \Delta\lambda(k_{DAS}) < 0$ m and $20 < \Delta\theta(k_{DAS}) < 90$ m. If we centre these ranges, then we estimate that most $\lambda(k_{DAS})$ lie within $\pm$ 30 m of the true value, and that $\theta(k_{DAS})$ lie within $\pm$ 60 m.

For comparison, we make similar calculations for k_{GPS} from cars 2404 and 2414 for trip 1 (Figs. 21 and 22 and Table 7). Overall, $\Delta\lambda(k_{GPS})$ and $\Delta\theta(k_{GPS})$ are closer to zero than the corresponding $\Delta\lambda(k_{DAS})$ and $\Delta\theta(k_{DAS})$ (Figs. 19 and 20).

Table 7 shows ranges of approximately 100 m, with standard deviations closer to means that are smaller than those of $\Delta\lambda(k_{DAS})$ and $\Delta\theta(k_{DAS})$ (Table 6) with most values falling between approximately $0 < \Delta\lambda(k_{GPS}) < 30$ m and $-10 < \Delta\theta(k_{GPS}) < 20$. If we centre these ranges, then we estimate that most $\lambda(k_{GPS})$ and $\theta(k_{GPS})$ lie within $\pm$ 15 m of the true values. If we treat $\Delta\lambda$ and $\Delta\theta$ as error estimates, then the accuracy of $\lambda(k_{DAS})$ is about half that of $\lambda(k_{GPS})$, and the accuracy of $\theta(k_{DAS})$ is less than one quarter that of $\theta(k_{GPS})$.

5. Conclusions

We track four-car test train for three trips using a Distributed Acoustic Sensing (DAS) system along with GPS coordinates and speedometer data from onboard the same train. We find that DAS data collected the telecom fibre running along the Red Line of the City of Calgary, Light Rail Transit (LRT) system provides latitude and longitude coordinates of trains on this system in real time. Though GPS coordinates are available from the cars that make up the test train, the majority of cars in

service on this system are not GPS-equipped, and no system yet exists to consume GPS coordinates from those that are GPS-equipped.

The DAS data and the GPS data are acquired on different time grids so we use Kalman filters to map the DAS distances and the GPS distances from the test train onto the same temporal grid. We find that slack fibre, fibre meanders, and interference from trains passing in the opposite direction, contribute noise to our estimated DAS distances.

To convert DAS distance to latitude and longitude coordinates, we interpolate the GPS coordinates onto the Kalman grid and compute unit heading vector for all points in time for each car and for each trip. This vector lies on the same time grid as the Kalman-filtered DAS distances so given a DAS distance, position in latitude and longitude is known.

We find that trip 1 has the most accurate DAS picks and that cars 2404 and 2414 (the lead and second cars of the test train) correspond most closely to that point along the test train closest to the DAS picks. Our final distance / latitude / longitude values correspond to cars 2404 and 2414 and trip 1, and this relationship depends only on the relative positions of the fibre and the track of the Red Line, and is independent of acquisition time. So long as the fibre and the track of the Red Line remain stationary, we estimate that DAS distances for any train on the Red Line, at any time, give the position of the first two cars with ± 30 m north-south and ± 60 m east-west.

Declaration of Competing Interest

None.

Acknowledgements

We thank Robert Palmer, George Palmer, and Kent Wardley of Fotech Solutions Inc. for providing a DAS system plus technical assistance. We thank Dr. Barry Phipps from the University of Calgary Office of the Urban Alliance and Nichole Wyatt from the City of Calgary Intellectual Property Analyst, Innovation, Data & External Access office for data-sharing coordination between the City, University and our industrial partner Fotech. For help with proofreading this manuscript we thank Natalie S. Robertson of the Municipality of Rocky View County, Alberta, and Mostafa Salari and David Henley of the University of Calgary. We thank also an anonymous reviewer for their suggestions regarding our DAS gauge length. Finally, we thank the City of Calgary Information Technology Infrastructure Services for their financial support of this project through Urban Alliance Research Agreement 1046981.

Appendix A. Convert position $[\lambda, \theta]$ to distance d_{GPS}

For a time sample t_n from the GPS logs where $1 \le n \le N$ and N is the number of samples in the log, we convert λ and θ to distance d_{GPS} from the origin using

$$d_{GPS}(t_n) = \sum_{j=1}^{n-1} \sqrt{d_{\lambda;j}^2 + d_{\theta;j}^2}, \tag{A.1}$$

where distance travelled north-south is

$$d_{\lambda;j} = [\lambda_{j+1} - \lambda_j] \left.\frac{\partial x}{\partial \lambda}\right|_j, \tag{A.2}$$

and distance travelled east-west is

$$d_{\theta;j} = [\theta_{j+1} - \theta_j] \left.\frac{\partial x}{\partial \theta}\right|_j. \tag{A.3}$$

Operators $\partial x/\partial \lambda$ and $\partial x/\partial \theta$ in eqs. A.2 and A.3 convert compass-degrees travelled to distance travelled according to

$$\left.\frac{\partial x}{\partial \lambda}\right|_j = \frac{\pi}{180} a \left(1-\varepsilon^2\right) \left(1-\varepsilon^2 \sin^2 \lambda_j\right)^{-\frac{3}{2}} \tag{A.4}$$

for latitude and

$$\left.\frac{\partial x}{\partial \theta}\right|_j = \frac{\pi}{180} a \cos \lambda_j \left(1-\varepsilon^2 \sin^2 \lambda_j\right)^{-\frac{1}{2}} \tag{A.5}$$

for longitude, where $a = 6378137.0$ m and $\varepsilon^2 = 0.00669437999014$ (Rapp, 1991, Eq. 3.105).

References

Aktas, M., Akgun, T., Demircin, M.U., Buyukaydin, D., 2017. Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications. *SPIE Commercial + Scientific Sensing and Imaging*, Pages 102080G–102080G. International Society for Optics and Photonics.

Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2007. Nonlinear kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Trans. Intell. Transp. Syst. 8 (4), 661–670.

Barnoski, M.K., Personick, S.D., 1978. Measurements in fiber optics. Proc. IEEE 66 (4), 429–441.

Bento, L.C., Bonnifait, P., Nunes, U.J., 2017. Cooperative GNSS positioning aided by road-features measurements. Transportation Research Part C: Emerging Technologies 79, 42–57.

Biondi, B., Martin, E., Cole, S., Karrenbach, M., Lindsey, N., 2017. Earthquakes analysis using data recorded by the Stanford DAS array. *SEG Technical Program Expanded Abstracts* 2017. Society of Exploration Geophysicists, pp. 2752–2756.

Dean, T., Papp, B., Hartog, A., 2015. Wavenumber response of data recorded using distributed fibre-optic systems.

Dean, T., Cuny, T., Hartog, A.H., 2017. The effect of gauge length on axially incident p-waves measured using fibre optic distributed vibration sensing. Geophys. Prospect. 65 (1), 184–193.

Dou, S., Lindsey, N., Wagner, A.M., Daley, T.M., Freifeld, B., Robertson, M., Peterson, J., Ulrich, C., Martin, E.R., Ajo-Franklin, J.B., 2017. Distributed acoustic sensing for seismic monitoring of the near surface: a traffic-noise interferometry case study. Sci. Rep. 7 (1), 11620.

Eaton, J.W., Bateman, D., Hauberg, S., Wehbring, R., 2019. GNU Octave version 5.1.0 manual: a high-level interactive language for numerical computations.

Faragher, R., 2012. Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]. IEEE Signal Process. Mag. 29 (5), 128–132.

Ferguson, R.J., 2015. Rayleigh wave dispersion function calculations with frequency dependent penetration depths. 77th EAGE Conference and Exhibition 2015.

Gadziński, J., 2018. Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. Transportation Research Part C: Emerging Technologies 88, 74–86.

Ganin, A.A., Mersky, A.C., Jin, A.S., Kitsak, M., Keisler, J.M., Linkov, I., 2019. Resilience in intelligent transportation systems (ITS). Transportation Research Part C: Emerging Technologies 100, 318–329.

He, M., Feng, L., Zhao, D., 2019. Application of distributed acoustic sensor technology in train running condition monitoring of the heavy-haul railway. Optik 181, 343–350.

Henningsen, S., Dietzel, S., Scheuermann, B., 2018. Misbehavior detection in industrial wireless networks: challenges and directions. Mobile Networks and Applications 23 (5), 1330–1336.

Hornman, J.C., 2017. Field trial of seismic recording using distributed acoustic sensing with broadside sensitive fibre-optic cables. Geophys. Prospect. 65 (1), 35–46.

Iida, D., Toge, K., Manabe, T., 2017. Distributed measurement of acoustic vibration location with frequency multiplexed phase-OTDR. Opt. Fiber Technol. 36, 19–25.

Jafarnia-Jahromi, A., Broumandan, A., Nielsen, J., Lachapelle, G., 2012. GPS vulnerability to spoofing threats and a review of antispoofing techniques. International Journal of Navigation and Observation 2012.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45.

Khakiev, Z., Shapovalov, V., Kruglikov, A., Morozov, A., Yavna, V., 2014a. Investigation of long term moisture changes in trackbeds using [gpr]. J. Appl. Geophys. 110, 1–4.

Khakiev, Z., Shapovalov, V., Kruglikov, A., Yavna, V., 2014b. [gpr] determination of physical parameters of railway structural layers. J. Appl. Geophys. 106, 139–145.

Krylov, V.V., 2017. Focusing of ground vibrations generated by high-speed trains travelling at trans-rayleigh speeds. Soil Dyn. Earthq. Eng. 100, 389–395.

Ku, E.M., Duckworth, G.L., 2013. Tracking a human walker with a fiber optic distributed acoustic sensor. Proceedings of Meetings on Acoustics ICA2013. vol. 19. ASA, p. 070053.

Martin, E.R., Castillo, C.M., Cole, S., Sawasdee, P.S., Yuan, S., Clapp, R., Karrenbach, M., Biondi, B.L., 2017. Seismic monitoring leveraging existing telecom infrastructure at the SDASA: active, passive, and ambient-noise analysis. Lead. Edge 36 (12), 1025–1031.

Mateeva, A., Lopez, J., Potters, H., Mestayer, J., Cox, B., Kiyashchenko, D., Wills, P., Grandi, S., Hornman, K., Kuvshinov, B., et al., 2014. Distributed acoustic sensing for reservoir monitoring with vertical seismic profiling. Geophys. Prospect. 62 (4), 679–692.

Mohammadi, R., He, Q., Ghofrani, F., Pathak, A., Aref, A., 2019. Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. Transp. Res. C 102, 153–172.

Moroz, T., McGhee, G., 2016. City of Calgary Engineering Standards 16–047.

Munn, J.D., Coleman, T.I., Parker, B.L., Mondanos, M.J., Chalari, A., 2017. Novel cable coupling technique for improved shallow distributed acoustic sensor vsps. J. Appl. Geophys. 138, 72–79.

Nakazawa, M., Horiguchi, T., Tokuda, M., Uchida, N., 1981. Measurement and analysis on polarization properties of backward rayleigh scattering for single-mode optical fibers. IEEE J. Quantum Electron. 17 (12), 2326–2334.

Papp, A., Wiesmeyr, C., Litzenberger, M., Garn, H., Kropatsch, W., 2016. Train Detection and Tracking in Optical Time Domain Reflectometry (Otdr) Signals. vol. 9796. Springer Verlag, pp. 320–331.

Pérez-Santisteban, I., Garca-Mayordomo, J., Martn, A.M., Carbó, A., 2011. Comparison among sasw, remi and ps-logging techniques: Application to a railway embankment. J. Appl. Geophys. 73 (1), 59–64.

Rapp, R.H., 1991. Geometric Geodesy Part I. Ohio State University Department of Geodetic Science and Surveying.

Rogers, A.J., Handerek, V.A., 1992. Frequency-derived distributed optical-fiber sensing: Rayleigh backscatter analysis. Appl. Opt. 31 (21), 4091–4095.

Sasani, S., Asgari, J., Amiri-Simkooei, A., 2016. Improving MEMS-IMU/GPS integrated systems for land vehicle navigation applications. GPS Solutions 20 (1), 89–100.

Tejedor, J., Macias-Guarasa, J., Martins, H.F., Piote, D., Pastor-Graells, J., Martin-Lopez, S., Corredera, P., Gonzalez-Herraez, M., 2017. A novel fiber optic based surveillance system for prevention of pipeline integrity threats. Sensors 17 (2), 355.

Wiesmeyr, C., Litzenberger, M., Waser, M., Papp, A., Garn, H., Neunteufel, G., Dller, H., 2020. Real-time train tracking from distributed acoustic sensing data. Appl. Sci. 10 (2), 448.

Wu, H., Wang, Z., Peng, F., Peng, Z., Li, X., Wu, Y., Rao, Y., 2014. Field test of a fully distributed fiber optic intrusion detection system for long-distance security monitoring of national borderline. In *OFS2014 23rd International Conference on Optical Fiber Sensors*. International Society for Optics and Photonics, p. 915790.

Zeng, X., Lancelle, C., Thurber, C., Fratta, D., Wang, H., Lord, N., Chalari, A., Clarke, A., 2017. Properties of noise cross-correlation functions obtained from a distributed acoustic sensing array at Garner Valley, California. Bulletin of the Seismological Society of America 107 (2), 603–610.