# Detection and analysis of anthropogenic patterns in a phase-stabilized optical fiber network

Dominik Husmann*, Serge Zaugg and Jacques Morel
*Federal Institute of Metrology METAS*
Bern-Wabern, Switzerland
* Email: dominik.husmann@metas.ch

***Abstract*—Phase-stabilized fiber networks for state-of-the-art optical frequency dissemination promise versatile application as environmental sensors, being sensitive to seismic waves, temperature variations or anthropogenic acoustic noise. Here we analyze the noise picked up in such a fiber network connecting METAS in Bern to the University of Basel over a length of 123 km. In the frequency range attributed to anthropogenic effects, we observe broadband noise level fluctuations following day-night cycles, as well as sporadic signals from localized sources. We apply unsupervised clustering to group these singular events and identify regularity in the schedule of similar events, that is indicative of railway activity. Our work provides a descriptive analysis of the observed signals and identifies possible exploits for application in terms of optimization of metrology networks as well as remote fiber sensing.**

***Index Terms*—frequency metrology, phase noise, fiber sensing, frequency dissemination**

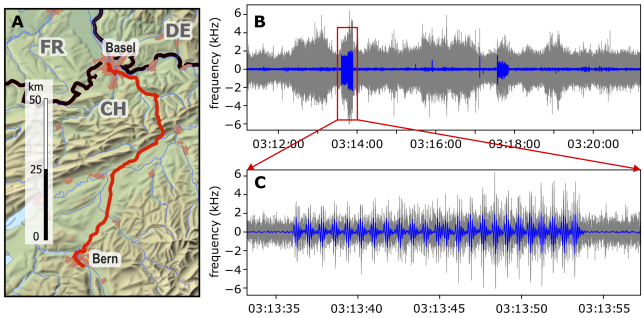


Fig. 1. A) Geographic layout of the fiber. B) Example of frequency correction (gray: raw data, blue: low-pass filtered). C) Zoom on frequency correction, revealing distinct pulse trains.

## I. Introduction

Optical frequency dissemination via fiber networks requires phase-noise read-out and compensation in order to minimize the disturbances from environmental noise sources. To that end, phase-noise compensation (PNC) schemes have been developed and deployed on frequency metrology networks spanning thousands of kilometers worldwide [1]–[6]. These networks enable state-of-the-art comparisons of atomic clocks at national and international levels, and are a necessary infrastructure for the planned redefinition of the SI second based on optical transitions [7].

As a side effect, the noise that is registered by the fibers and canceled in the PNC schemes can be exploited for fiber sensing, hence monitoring and analyzing the PNC frequency correction signal has gained interest in the past years in various fields. For example, it has been shown that signals from such transmission-based coherent detection schemes are exploitable in geophysics for the measurement of seismic events [8], [9].

The interpretation of the PNC signal is challenging due to the lack of spatial resolution: the net phase error is given by the integration along the full fiber length, hence all noise processes along the fiber contribute to the PNC signal. Here we single out and describe a specific family of noise phenomena observed in the PNC signal acquired in the Swiss frequency metrology network [1]. These signals are characterized by distinct spectral features and regularity in time. In order to gain insights into the source of these noise processes, we detect and categorize these events, and find characteristic features in frequency and schedule. Based on this observation and the fact that the fibers mainly follow road and railway paths, a plausible origin of these signals is given by railway traffic at a few distinct points of high mechanical coupling to the fiber.

## II. Experimental setup

The PNC data we analysed here was acquired in the 123 km long fiber segment of the Swiss frequency metrology network, connecting METAS in Bern to the University of Basel (see Fig. 1A and [1] for details on the network and PNC scheme). In short, an ultrastable telecommunication L-band laser (frequency 190.7 THz) is used to measure the phase noise on the 123 km long fiber network, in a coherent phase detection scheme. The laser is multiplexed into a standard telecommunication network, and passes two bidirectional erbium-doped amplifier stages to compensate optical attenuation. A feedback loop on the optical frequency via a frequency correction on an acousto-optic modulator is used to compensate the phase error. For processes slower than the loop bandwidth, this signal describes well the noise picked up by the fiber. The frequency correction signal $\Delta\nu(t)$ is recorded as a time series with 500 samples per seconds (sps) via an analog-to-digital converter, and stored for postprocessing as described in the following.

## III. Method

We analyzed $\Delta\nu(t)$ over a time period of 8 months (July 2022 to March 2023). An example of a time trace is shown in Fig.1B, with the zoom in Fig.1C revealing a typical recurring noise feature.

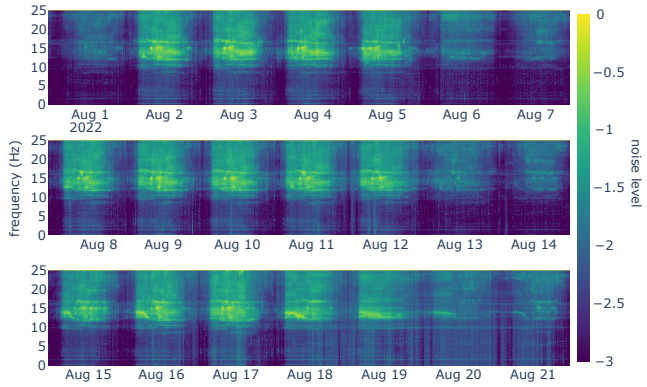


Fig. 2. Per-week spectrograms for August 2022. Each row is a calendar week starting at Monday. Work days show increased noise compared to weekends and holidays (e.g. August 1).

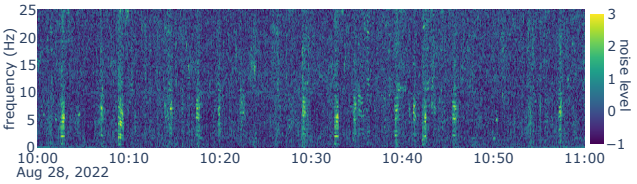


Fig. 3. Spectrogram over 1 hour, after equalizing with median spectrum.

Initial exploration showed that the events of interest had high signal-to-noise ratio (SNR) below 15 Hz. At higher frequencies, the noise spectrum is dominated by broadband noise. To reduce the data load, the signal was down-sampled to 50 sps, resulting in an effective bandwidth of 25 Hz. To visualize long-term trends in the noise spectrum, we transform the data using short time Fourier transform (STFT) with window length of 22 minutes. Fig. 2 shows the spectrograms of three selected weeks.

*a) Signal pre-processing:* In order to identify distinct spectral features, we group the time series into arrays of 1 hour, remove the mean value per array, and transform each into the time-frequency domain via STFT (Hann window of 5.12 and step of 1.28 seconds), resulting in arrays $\mathbf{X}$ of 129 by 2813 frequency-time bins. Further we normalize each frequency bin by its hourly median, in order to suppress spectral patterns that appear stationary over 1 hour duration. To attenuate extreme values, we take the logarithm of the absolute values and cap the spectrogram between -1 and +3. Fig. 3 shows an example of a spectrogram over 1 hour duration, where events with strong spectral features are revealed as bright vertical stripes.

*b) Detection of ROIs:* Exploration of the spectrograms revealed that the events, termed here regions of interest (ROIs), were characterized by well defined spectral lines, often with several harmonics, and showed a typical duration of 10 - 30 seconds. Start and end point of ROIs were identified using a detection score along the time axis, where a modified version of the Shannon entropy (1) was computed for each spectrum $x_i$ of $\mathbf{X}$.

$$e = \frac{1}{\log(n)} \sum_{i=1}^{n} y_i \log(y_i), \quad y_i = x_i / (\sum_{i=1}^{n} x_i) \tag{1}$$

Here the minus sign of the standard entropy was replaced by normalization by the number of bins, which casts the results into the interval [0,1]. A score of 0 corresponds to a flat spectrum, while a score of 1 describes a spectrum with all energy concentrated in one bin (see [10] for details). The results are time series $e_j$ of score values. To enforce the detection of sequences of consecutive high score values rather than single spurious peaks, a moving average was applied to $e_j$. Finally, events were detected by applying an empirically determined threshold of 0.1. For each detected ROI, we extract its starting time stamp and a short spectrogram from $\mathbf{X}$.

*c) Feature extraction:* ROIs shorter than 5 bins (∼10 s) were excluded to keep computation time feasible. In order to group similar events, we first define features, based on which the ROIs are categorized. Here, the ROI mean along time was taken as features, resulting in a fixed size vector of dimension 129, which is then fed into a clustering algorithm. Each feature was normalized to zero mean and unity standard deviation.

*d) Clustering of Detected ROIs:* The objective of the clustering procedure was to manually explore the data and find patterns of interest. Specifically we were aiming for short signals, with potentially simple interpretation in terms of antropogenic phenomena, such as railway signals from trains (see Fig. 1C for a representative pattern of interest). The overall procedure involved 3 steps: pre-clustering to select ROI candidates, dimension reduction, and final clustering. Pre-clustering of ROIs with similar feature values was achieved by sequential application of Agglomerative Hierarchical Clustering [11] with incremented distance threshold. Further we

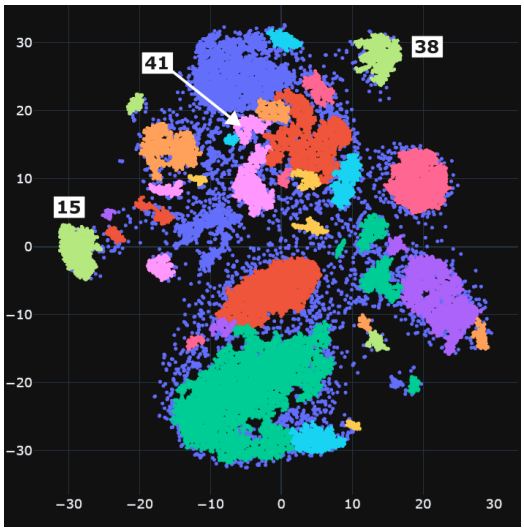


Fig. 4. Reduced set of 44 clusters, mapped to 2D feature space via t-SNE. Each color corresponds to a certain cluster, unclustered ROIs are shown in light blue. Cluster ID 15, 38 and 41 are discussed in more detail.

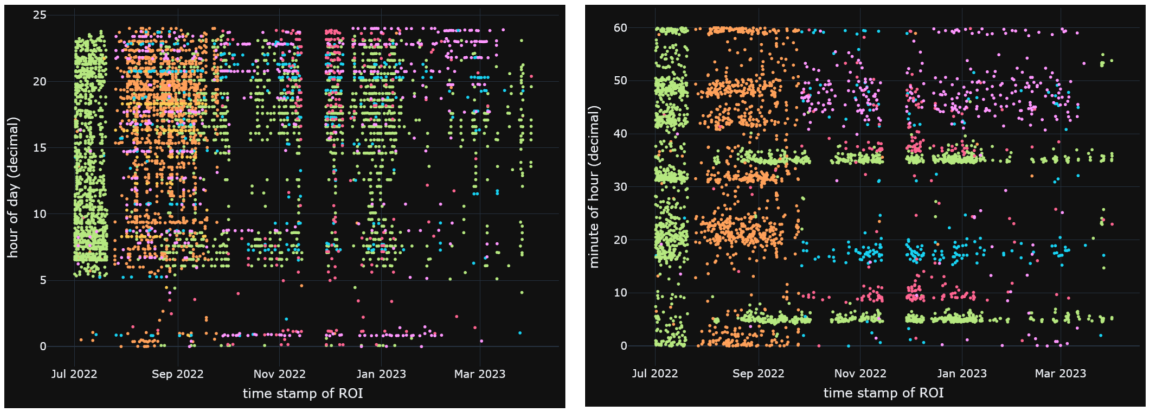


Fig. 5. Time stamps of six selected clusters in hour-of-day (left) and minute-of-hour (right) plot. Each color represents a single cluster. The horizontal patterns indicate regularity in the occurrence of events. ID 15: Green dots starting from September, ID 38: Green dots in July, ID 41: Magenta dots starting from October.

used average linkage with the Euclidean distance as a dissimilarity metric. In the first iteration with the lowest value of the threshold, clusters containing at least $N = 100$ ROIs where extracted. The remaining ROIs were passed to the next iteration with the threshold incremented by 0.20. The process was iterated for thresholds between 10.0 and 12.0. At this point, ROIs with similar spectral signatures are assigned to pre-clusters while unassigned ROIs are excluded. In the second step we perform dimensionality reduction of the pre-clustered ROIs. To that end, the 129 features were mapped to 2D feature space via t-SNE [12]. Pre-cluster IDs were used to fine-tune the t-SNE to ensure their proximity in the 2D mapping. Finally, the ROIs in 2D space were clustered via DBSCAN [13] into larger consistent clusters.

## IV. Results

### A. Trends in broadband noise

Anthropogenic features in the PNC signal is revealed in the weekly spectrograms in Fig.2: A diurnal pattern with high signal amplitude during daytime is manifested by a sharp rise around 6 AM and a smooth drop around 18 PM local time. This is commensurate with standard local work hours and day schedules, leading to increased traffic on roads and railways during the day. A weekly pattern was also observed with increased noise during workdays (Monday to Friday) and decreased noise on weekends or public holidays (e.g. Swiss national holiday August 1). Several narrow-band components are visible on the spectrograms, some have a near constant frequency (between 0 - 5 Hz) while others are frequency modulated (around 15 Hz). The origin of these signals has not been determined, and may originate either from the events on the fiber or from the detection system itself.

### B. Clustering and event regularity

A total of 53352 ROIs were extracted for the 8 months period of interest. Among those, 31607 ROIs were assigned to 223 pre-clusters, while the remaining unassigned ROIs were discarded. The assigned ROIs were grouped into the reduced set of 44 larger clusters (see Fig. 4). The 2D feature space of the reduced set obtained via t-SNE is shown in Fig.4. This reduced set shows how the clusters are grouped in dense regions, which are easily identified by DBSCAN. Six clusters ($N = 3895$ ROIs) were chosen for closer analysis, and are visualized on hour-of-day and minute-of-hour (Fig. 5) plots. These plots illustrate how events with similar spectral characteristics occur at near identical time-of-day, and in some cases are scheduled at certain minutes every hour. In a daily schedule, events start appearing at 5 AM and end just after midnight. The regularity and daily schedule of these events are well aligned with railway public transport, which often follow a 30 minute schedule. This hypothesis is also in agreement with the geographic fiber layout. Note that between 10 am and 3 pm we detected considerably fewer events, due to globally increased broadband noise during the daytime hours, which masks the events from the ROI detection algorithm.

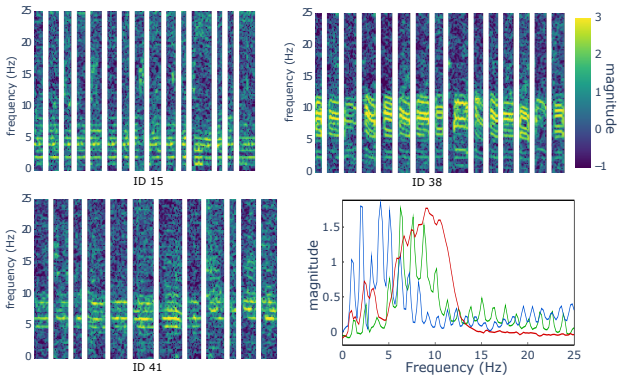


Fig. 6. Representative spectrograms of a few ROIs from three selected clusters. The bottom right shows the average spectrum of the corresponding clusters. Blue: ID 15, Red: ID 38, Green: ID 41

### C. Description of selected clusters

We select 3 specific cluster that show regularity in schedule for closer analysis: cluster 15, 38 and 41 (see Fig. 4). Exemplary spectrograms of ROIs from these clusters are shown in Fig. 6, as well as the mean spectrum over each cluster.

Cluster 15 containing a total ot 1174 ROIs showed the highest level of schedule, with ROIs occurring regularly around minute 05 and 35 every hour during daytime. The spectrum exhibits well defined frequency bands in the low Hertz range. A possible source of such a spectrum could be a railway train passing over a point of high coupling to the fiber. Whenever a wheel axis hits the fiber, it distorts the phase. The wheelbase and the speed of the train then give rise to a characteristic frequency pattern. E.g. a train with wheelbase 10 m and speed of 40 km/h would give rise to a fundamental frequency of around 1.1 Hz. Cluster 38 is an example of distinct frequency patterns with occasionally decreasing frequencies within a single ROI. This could be explained by braking of a train. We also observe increasing frequency patterns in other cluster, indicating accelerating trains. Cluster 41 is another example of well pronounced frequency bands similar to cluster 15, here with higher frequencies on the order of 5-10 Hz.

## V. Discussion

The results show on one hand the general correlation of broadband noise levels with human activity, as seen in the weekly spectrograms in Fig. 2. The fibers in this project follow mainly highways and railway tracks and pass through several kilometers of urban areas, so influences of acoustic vibrations from transportation are expected. The detection of singular noise ROIs, well defined in time, and the subsequent successful classification and clustering was surprising, as the noise is integrated along the whole fiber. The interpretation in terms of few points along the fiber strongly coupled to the environment, picking up railway traffic vibrations, could be verified by monitoring certain geographic points of interest locally with a camera, and comparing time stamps of passing trains with the detected ROIs. A deeper analysis of the observed ROIs in terms of duration and spectral components could be done to extract specific information of the model of the trains (number of wagons, speed, acceleration, wheelbase, etc.).

Our observations presented here are coincidental in a sense where the fiber network was not designed to specifically measure such signals, and the points of high coupling are neither controlled nor known a priori. However one could design fiber network layouts, where a set of well defined points with high coupling to the environment are introduced in a controlled way, e.g. along train tracks, or in low-noise environments such as oceanic fiber networks. This would then allow for long-range fiber sensing applications going well beyond the possibilities of well established techniques such as distributed acoustic fiber sensing.

For the metrological aspect of the network, the gained knowledge on noise sources could be used to specifically tailor the noise suppression mechanisms. Further it allows to specify network performances based on time of day and day of week, which again may be beneficial for planning of measurement campaigns.

## VI. Conclusion and Outlook

Our study presents a method for feature extraction of transient signals in coherent fiber optic networks. Applying this method to a specific fiber segment of the Swiss frequency metrology network, we were able to identify distinct clusters of signals. The regularity and spectral fingerprint of the signals strongly hint their origin to be in railway traffic. The method however is not limited to this regime, and possibly could be applied to other spectral ranges (e.g. sub-Hz) in order to detect and classify other noise sources. This has a two-fold application: First, the gained insight into the signal clusters can help identify critical noise sources detrimental for coherent fiber optic networks. Second, the method could be exploited for fiber sensing of remote places, as it provides a remote fiber sensing example where the fiber spans over 100 km - a range that is commonly not available to conventional sensing techniques such as distributed acoustic sensing.


## Acknowledgment

We thank Stefan Willitsch for hosting the PNC setup in his laboratory in Basel, Fabian Mauchle from SWITCH for providing the network infrastructure, Oscar de Feo for insightful discussions, and Antoine Jallageas for careful reading of the manuscript.


## References


[1] D. Husmann *et al.*, "Si-traceable frequency dissemination at 1572.06 nm in a stabilized fiber network with ring topology," *Optics Express*, vol. 29, pp. 24 592–24 605, 2021.

[2] C. Lisdat *et al.*, "A clock network for geodesy and fundamental science," *Nature Communications*, vol. 7, p. 12443, 2016.

[3] D. Calonico *et al.*, "High-accuracy coherent optical frequency transfer over a doubled 642-km fiber link," *Applied Physics B*, vol. 117, no. 3, pp. 979–986, 2014.

[4] T. Takano *et al.*, "Geopotential measurements with synchronously linked optical lattice clocks," *Nature Photonics*, vol. 10, no. 10, pp. 662–666, 2016.

[5] Boulder Atomic Clock Optical Network (BACON) Collaboration: K. Beloy *et al.*, "Frequency ratio measurements at 18-digit accuracy using an optical clock network," *Nature*, vol. 591, no. 7851, pp. 564–569, Mar. 2021.

[6] M. Cizek *et al.*, "Coherent fibre link for synchronization of delocalized atomic clocks," *Optics Express*, vol. 30, pp. 5450–5464, 2022.

[7] N. Dimarcq *et al.*, "Roadmap towards the redefinition of the second," *Metrologia*, vol. 61, 2 2024.

[8] S. Noe, D. Husmann, N. Müller, J. Morel, and A. Fichtner, "Long-range fiber-optic earthquake sensing by active phase noise cancellation," *Scientific Reports*, vol. 13, pp. 1–7, 8 2023.

[9] G. Marra *et al.*, "Ultrastable laser interferometry for earthquake detection with terrestrial and submarine cables," *Science*, vol. 361, no. 6401, pp. 486–490, 2018.

[10] S. Zaugg, M. van der Schaar, L. Houégnigan, and M. André, "A framework for the automated real-time detection of short tonal sounds from ocean observatories," *Applied Acoustics*, vol. 73, no. 3, pp. 281–290, 2012.

[11] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[13] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.