Research Internship Report
École polytechnique, Department of Applied Mathematics

Correlating Wind and Distributed Acoustic Sensing Data

Théophane Pincemin

April - August 2024

Supervisor: **Jo Eidsvik**[1,2]
Coordinator: **Aymeric Dieuleveut**[3]
Examiner: **Jean-Marc Bardet**[4]

[1] *Centre for Geophysical Forecasting, Norwegian University of Science and Technology, Trondheim, Norway*
[2] *Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway*
[3] *CMAP, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France*
[4] *SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris, France*

Statement of integrity regarding plagiarism

I hereby declare that I am the author of this report, that the results described are the product of my own work and have not used any third-party sources or results without clearly citing and referencing them according to the recommended bibliographic rules.

Abstract

Distributed Acoustic Sensing (DAS) is a technology that enables high-resolution strain measurements along optical fibers, proving valuable in various fields. This study aims to move beyond the qualitative correlation between strain in overhead cables and wind speed observed in DAS measurements by deriving quantitative relationships.

Firstly, we outline a methodology for data sourcing and processing. We then develop models based on power laws to predict strain using wind measurements from nearby weather stations and detail a parameter estimation methodology. Utilizing experimental data, we estimate the model parameters and demonstrate the importance of using spatially varying coefficients. Finally, we present a potential use case for our study, discuss the limitations of our approach, particularly regarding predictive power and numerical efficiency, and suggest incorporating more expert knowledge and adding structure to the error in future research.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Jo Eidsvik, for his invaluable guidance and support throughout this project. His expertise, advice, and encouragement have been instrumental in the successful completion of this work. I am also grateful to his Ph.D. student, Øyvind H. Singsaas, for his meticulous proofreading and insightful comments, which greatly enhanced the quality of this report.

Special thanks go to Robin A. Rørstadbotnen, who designed and performed the processing of the data, and shared his expertise in Distributed Acoustic Sensing. I am also indebted to Martin Landrø, who helped me understand the physics of overhead lines and their interaction with wind.

I would like to extend my appreciation to the remaining staff of the Center for Geophysical Forecasting, the Department of Mathematical Science, and the Norwegian *Meteorologisk institutt*, whom I had the pleasure of meeting during my internship and who significantly contributed to its success.

Finally, I would like to acknowledge my internship coordinator, Aymeric Dieuleveut, and my examiner, Jean-Marc Bardet, for their time and effort invested in this internship.

Contents

Chapter 1

Introduction

Distributed Acoustic Sensing (DAS) is an innovative technology that transforms a standard optical fiber into a series of thousands of virtual microphones by connecting it to a device called a *DAS interrogator*. The interrogator sends laser pulses through the fiber and measures the backscattered light. By analyzing this backscattered light, it is possible to infer the strain of the fiber at each point, which is related to the acoustic waves that propagate along the fiber. The Center for Geophysical Forecasting[1] (CGF) has developed a special expertise in this technology with its partners and fosters research and innovation around it (CGF, 2024).

DAS offers several advantages, including high spatial resolution (up to 1 meter) over long distances (more than 100 km), continuous real-time monitoring, and the ability to operate with power input and data output only at one end of the cable. Additionally, DAS has a flat response, a large bandwidth (roughly 10 mHz to 50 kHz), and is relatively low cost, especially when reusing existing fibers (Zhan, 2020). DAS data has proven valuable for a wide range of applications, such as seismologic surveys, monitoring oil and gas fields, wells and pipelines, listening to marine mammals, tracking avalanches, and monitoring seas, borders, roads, bridges, and railways (Bouffaut et al., 2022; Dou et al., 2017; Lindsey et al., 2019; Lindsey & Martin, 2021; Zhu et al., 2022)

While DAS technology offers numerous advantages and applications, leveraging its full potential presents certain challenges. DAS interrogators generate massive amounts of raw data, which can be overwhelming and difficult to interpret. To extract valuable information, it is crucial to develop appropriate models and algorithms that transform this raw data into actionable insights.

In this research internship, we focus on the correlation between wind and DAS data from a fiber hung in the air. It has been shown through data visualization that optical sensing measurements are *qualitatively* connected with wind speed, both with DAS or State of Polarization sensing (CGF, 2024). However, to our knowledge, no *quantitative* model has been proposed yet. Such a model would not only be of scientific interest but could also contribute to developing new applications of DAS technology and improving existing ones. For instance, given sufficient reliability and precision, it could, according to the Norwegian *Meteorologisk institutt*, help in better understanding local weather phenomena, especially orographic effects, since DAS can provide large scale, high resolution, and *in situ* data at reasonable cost.

The goal of this work is to develop models explaining the correlation between weather and DAS data, relying on tools developed in the fields of spatial statistics and Bayesian modeling. We will at first introduce the data used in Chapter 2. Then we propose models for the DAS response to wind in Chapter 3 and method for estimating their parameters, before we analyze the performances of the models on real data in Chapter 5. We finally discuss the results and propose some perspectives in Chapter 6, and conclude in Chapter 7.

[1] Their presentation webpage at www.ntnu.edu/cgf/about reads "The Centre for Geophysical Forecasting (CGF) is a joint enterprise funded by the Norwegian Research Council (NRC), Industrial partners and NTNU. It is a 'Research-based Innovation' (SFI) Centre, selected by the NRC for support in June 2020 and opening in December 2020."

Chapter 2

Problem description and Data

2.1 Experimental setup and geography

The DAS data used in this work comes from a fiber running along the 132 kV electric power line operated by Elmea between Fygle and Solbjørn, in the Lofoten islands, in Norway (OpenStreetMap contributors, 2024). Maps for location are provided in Figure 2.1 and visualization of the setup is provided in Figure 2.2.

2.2 DAS data

2.2.1 Data acquisition and processing

Properties of the data are summarized in Table 2.1. The strain is measured with a DAS Interrogator from Alcatel Submarine Networks (ASN) (product code: OptoDAS C01). The interrogator measures the time differentiated phase rate per distance on 82000 channels, each separated by about 1.021 m, and aggregate them with a gauge length of 8 channels. For more information on how DAS works, see Appendix 8.2. The interrogator performs one measurement every 0.0016 s and output one HDF5 file every 10 s. The following processing steps are then applied to the raw data:

1. Reduce the number of channels by selecting channel indices $3i$, $i \in [\![0, 1393]\!]$
2. Convert from time-differentiated phase change per distance to strain
3. Apply a low-pass Butterworth filter of order 5 with cutting frequency 21.25 Hz to avoid aliasing in the resampling
4. Apply a high-pass Butterworth filter of order 4 with cutting frequency 0.2 Hz
5. Resample the data to $f_s = 62.5$ Hz (1 every 10 points)
6. Compute the Root Mean Square (RMS) of the strain over 60 s windows.

We use the notation $strain(\mathbf{s}, t)$ to refer to processed signal at channel $\mathbf{s}$ and time period t, or rather the RMS of the strain on this space-time region. The $\mathbf{s}, t$ notation for location and time is consistent with the convention used in (Banerjee et al., 2014). Due to discretization in space and time with channels and minutes, $\mathbf{s}$ and t are actually integers.

2.2.2 Locating DAS channels

In order to be able to compare the DAS data with the wind data, it is necessary to know the geographical location of the DAS channels.

The DAS data is indexed with channels, which are evenly spaced along the fiber with a distance of about 24.5 m between two consecutive channels in the processed data. This defines a first curvilinear abscissa, which we name *fiber abscissa*. In addition to this, the power line is referenced by OpenStreetMap (OSM) (OpenStreetMap contributors, 2024), as

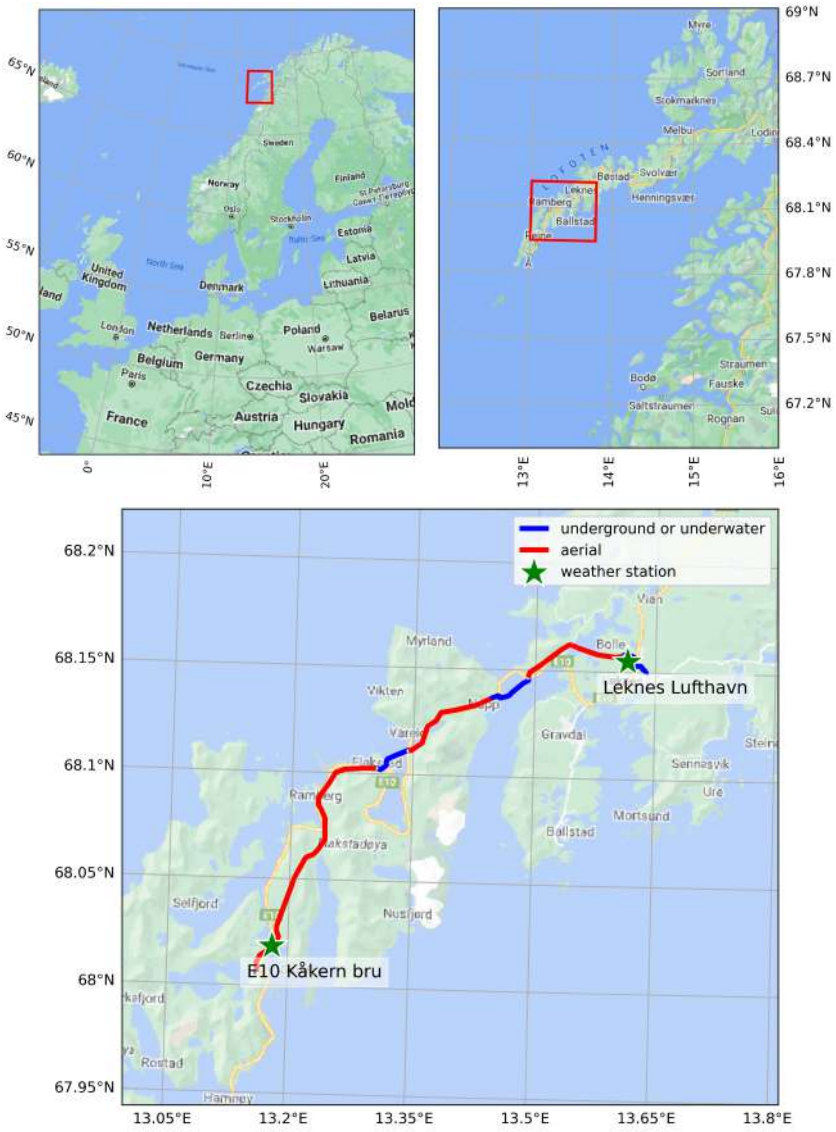

Figure 2.1: Maps locating the fiber at different scales, map background from Google Maps. The channel numbers are increasing from Leknes to Kåkern.

Property	Interrogator output	Processed data
Measurement	time differentiated phase rate per distance	RMS of the strain
Unit	rad/s/m	(dimensionless)
Sampling period	0.0016 s	60 s
Channel spacing	$\approx$ 8.16 m	$\approx$ 24.5 m
Number of channels	10250	1394
File format	HDF5	Matlab v5 mat-file
Duration of a file	10 s	$\approx$ 30 min

Table 2.1: DAS data used in this work. Both raw and processed data are contiguously spaced in time and space. The data used in this study corresponds to the following days, bounds included: 8th to 9th and 25th to 29th of January 2024, and 3rd to 4th of February 2024.

shown in Figure 2.1. It enables us to define a curvilinear abscissa along the projection of the power line route on the earth surface, the *geographical abscissa*. We finally need to map

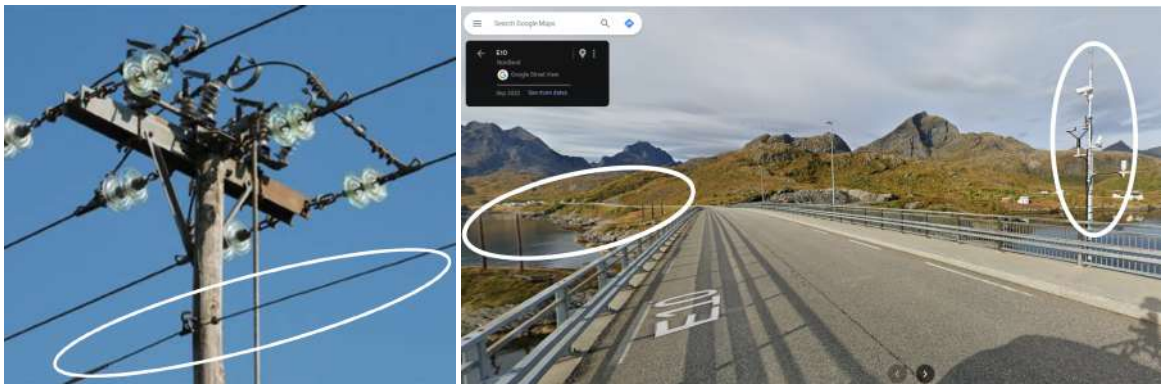

Figure 2.2: Photograph showing the location of the fiber with respect to the power line and Google Street View screenshot from Kåkern bridge. The Fygle-Solbjørn power line has three conductors and the fiber is in a cable slightly below, as the circled cable on the first picture. On the second picture, the Fygle-Solbjørn line is circled on the left and the E10 Kåkern bru weather station from the Norwegian *Meteorologisk institutt* is circled on the right.

Metric	Value
Leave-one-out mean absolute error	65 m
$1 - R^2$	$9.0 \cdot 10^{-5}$
Standard deviation on the slope	$5.2 \cdot 10^{-3}$
Standard deviation on the intercept	69 m

Table 2.2: Metrics of the linear regression mapping fiber abscissa to geographical abscissa. The linear regression is performed using `scipy.stats.linregress`, which automatically computes the 3 last metrics. R is the Pearson correlation coefficient, so that $1-R^2 = SS_{res}/SS_{tot}$ where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. If one notes $(x_i)_{i\in[\![1,5]\!]}$ the fiber abscissa of our five landmark channels, $(y_i)_{i\in[\![1,5]\!]}$ the corresponding geographical abscissa, $(\hat{y}_i)_{i\in[\![1,5]\!]}$ the estimations of the $(y_i)_{i\in[\![1,5]\!]}$ by linear regression and $\bar{y} = \sum_{i=1}^{5} y_i/5$, then $SS_{res} = \sum_{i=1}^{5}(y_i - \hat{y}_i)^2/5$ and $SS_{res} = \sum_{i=1}^{5}(y_i - \bar{y})^2/5$.

the fiber abscissa to the geographical abscissa. We do this learning an affine transformation between the two abscissas through linear regression using a few known points. Those known points correspond to transition of the fiber from the ground or sea to the air, which are referenced in the OSM data and can be identified in the DAS data with change in order of magnitude of the signal, as shown in Figure 2.3. Some metrics on this linear regression are summarized in Table 2.2.

2.3 Wind data

The goal of this internship is to correlate wind and DAS data, it is therefore crucial to have access to meteorological data. We naturally relied on the Norwegian *Meteorologisk institutt* for this, with two main sources:

- Weather stations data, which are the most accurate and reliable data, but also the most spatially sparse.
- Analysis data, which are more continuous in space and cover a larger area, but are less accurate and sparser in time.

2.3.1 Weather stations data

The records of weather stations used by the Norwegian *Meteorologisk institutt* are available at https://seklima.met.no/. In this study, we focus on the two weather stations closest to

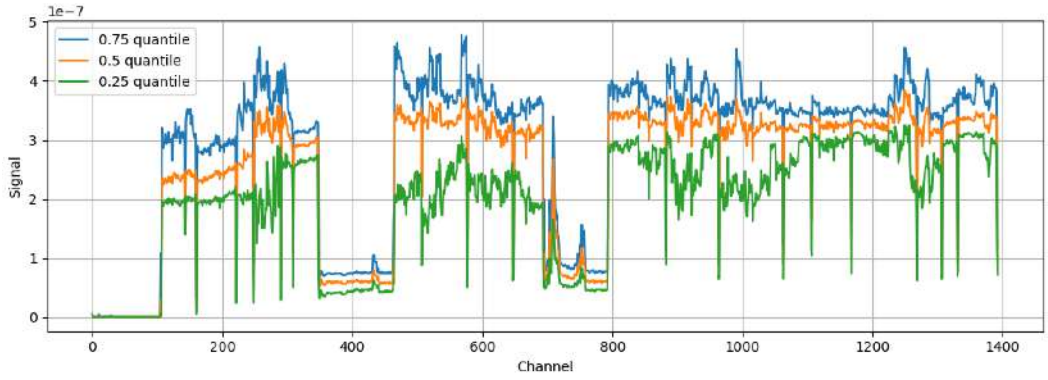

Figure 2.3: Identification of landmark channels in the DAS data corresponding to transition between aerial sections and underground and undersea sections. This graph features plots of quantile the RMS of strain aggregated over time for each channel. Identified landmark channels are identified at indices 106.5, 348, 464.5, 693.5 and 793. Leknes airport weather station is close to channel 90 and Kåkern bridge weather station is close to channel 1330.

the DAS cable, which are located at Leknes airport and E10 Kåkern bridge, displayed in Figure 2.1.

The data variables we download are: average temperature, cumulative precipitation amount, average wind direction, average wind speed and wind gust, which is defined as the maximum wind speed exceeded over a 3 seconds period. This data has the best available accuracy and a quite high temporal resolution with a 10 minutes sampling period, but it is spatially very sparse and full of missing data. The measurements were retrieved as CSV files, and then processed with the `pandas` library (McKinney, 2010; "Pandas", 2024) in python.

In addition to that, seklima.met.no provides aggregated data such as wind roses, which are useful to visualize the wind direction and speed distribution.

2.3.2 Analysis data

Analyses are based on the results of numerical weather prediction models fed with data assimilation, such that the analysis at time t uses measurements up to time t. In order to make smaller files, analyses do not include the whole output of the numerical weather prediction model, but only the most relevant variables: for instance, the analyses we use in this study are not designed to be used in aeronautics and therefore focus on the surface.

Analysis data from the Norwegian *Meteorologisk institutt* is available at https://thredds.met.no/thredds/catalog/metpparchive/catalog.html. The data come as NetCDF files ("NetCDF", n.d.) with 11 geophysical and meteorological data variables on a 1 km $\times$ 1 km $\times$ 1 hour grid (no vertical dimension) covering Scandinavia with a Lambert conformal projection. In order to process the NetCDF files in python, we relide on the `xarray` library (Hoyer & Hamman, 2017; Hoyer et al., 2024).

2.3.3 Discrepancy between analysis and measurements

We realized that analysis data can be quite different from the measurements, such as presented in Figure 2.4. We considered that the weather stations data was more reliable and therefore did not use analysis data but only weather stations data for quantitative studies. However, we used analysis data for visualization, such as presented in Figure 2.5.

2.4 Association between DAS and weather data

The models we use predict the RMS strain from the wind speed. In order to estimate their parameters, we need data points that are pairs of wind speed and RMS strain. In this section, we discuss how we pair weather data with the DAS signal.

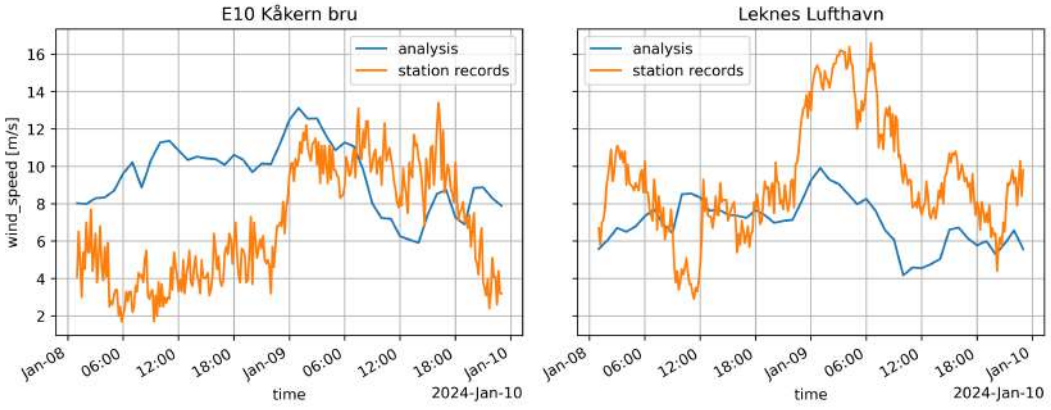

Figure 2.4: Comparison of weather data from the Leknes airport and E10 Kåkern bridge weather stations and the analysis data. The analysis data is interpolated to the location of the weather stations with the `xarray.DataArray.interp` method, which computes a triangulation and then a linear barycentric interpolation. This comparison shows that the analysis data can be quite different from the measurements, even though weather station data was taken into account in order to produce the analysis.

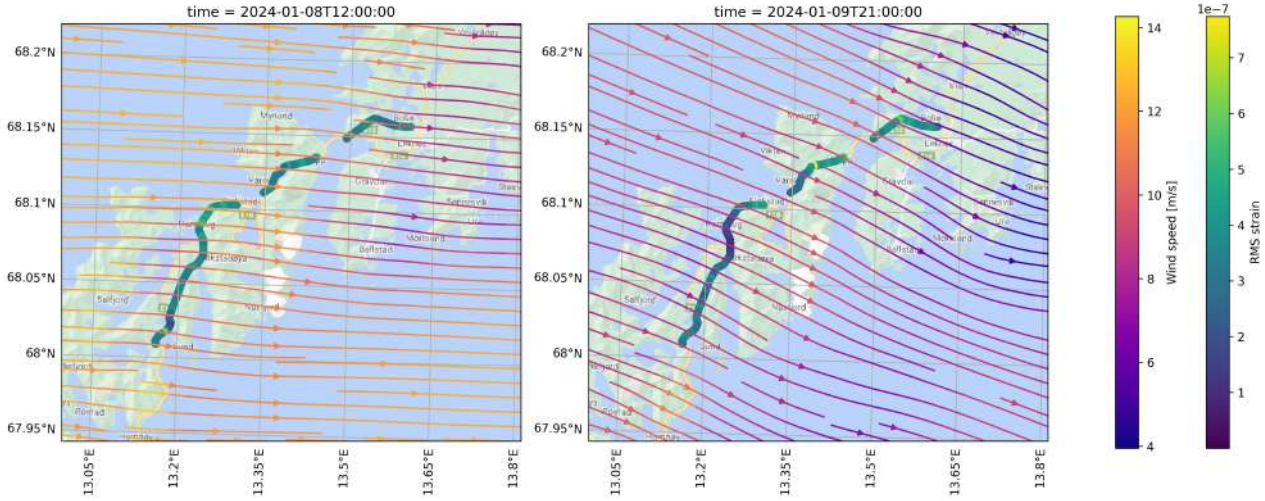

Figure 2.5: DAS data and weather analysis data from the Norwegian *Meteorologisk institutt*, Google Maps background. We can see that on January 8th at 12 pm, both the wind and the strain measured by DAS are stronger than on January 9th at 9 pm. It is this kind of correlation that motivated this internship.

As mentioned in 2.3.3, we only use weather stations data for parameter estimation. We pair these weather stations measurements with DAS signal at the same time, with a time tolerance of 3 minutes, and only from aerial sections of the fiber *close enough* from the weather station. We consider that a channel is *close enough* from a weather station if it is one of the 40 aerial channel closest to the station, based on the channel locations computed in Chapter 2. This results in channels 107 to 146 for the Leknes airport station and channels 1309 to 1348 for the Kåkern bridge station. The maximal distances between the stations and their related channels are estimated to be 445 m between Kåkern bridge and channel 1309, and 1233 m between Leknes airport and channel 146, which is significantly more, but the landscape is also flatter there. Figure 2.6 shows the situation in a mapview.

In order to avoid errors when computing the logarithm of the wind speed, we change the records with a wind speed of 0 m/s to 5 cm/s, which makes sense since the measurements have a precision of 0.1 m/s. We finally save the data in a CSV file, with the following columns: channel, RMS strain, timestamp of the DAS signal, wind speed, wind direction,

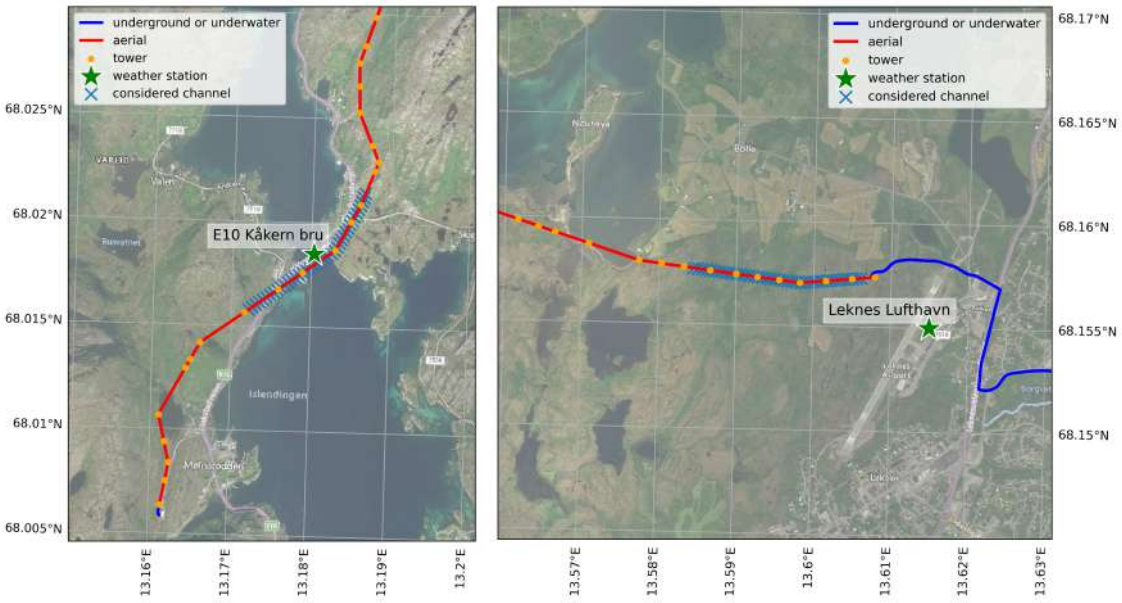

Figure 2.6: Map of the surroundings of the Leknes airport and E10 Kåkern bridge weather stations from the Norwegian *Meteorologisk institutt* and locating the channels paired with each station. Map background from Google Maps.

gust, temperature, precipitation, and weather station. This dataset is named `dataset_a` in the code. For the models with spatially varying parameters, we consider separately the data from the two stations, while we consider both stations together otherwise. The models we use only depend on the wind speed, and on the channel in the case of spatially varying parameters, but we keep the other variables in the dataset in order to identify confounding factors.

Chapter 3

Modelling the DAS response to wind

In this section we define models that can be used to predict the DAS response to wind. We first present a model based on a power law and a quotient of power laws model, then a power law model with spatially varying coefficients. From a Bayesian perspective, having a range of models allows us to relax the prior set on the model structure. We can then compare the models.

3.1 Power law model

3.1.1 Model description

A first simple model that can be used to describe the DAS response to wind is a power law model, where the signal power is proportional to the wind speed raised to some power. In order to account for background noise, we add a constant term to the model. The model is defined as follows:

$$y_{\mathbf{s},t} = \beta_0 + \beta_1 \cdot x_{\mathbf{s},t}^{\alpha} + \epsilon_{\mathbf{s},t}, \quad \epsilon_{\mathbf{s},t} \overset{i.i.d}{\sim} \mathcal{N}(0, \tau^2), \tag{3.1}$$

where $y_{\mathbf{s},t}$ and $x_{\mathbf{s},t}$ are the RMS strain measured by DAS and the wind speed at location $\mathbf{s}$ and time t, $\mathbf{s} \in [\![1, N]\!]$ and $t \in [\![1, T]\!]$ with discretization in channels and minutes, while α, β_0, and β_1 are parameters to be estimated. The noise term $\epsilon_{\mathbf{s},t}$ is assumed to be Gaussian white noise with zero mean and variance τ^2, the mean being already captured by β_0. All parameters can be grouped in a vector $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1, \tau)$.

3.1.2 Motivation

Power laws are commonly used in physics, and especially in fluid dynamics, to express relationships between quantities without deriving a comprehensive set of equations, often performing surprisingly well when used in the frame of dimensional analysis (Jacquin, 2022). For instance, the power of the wind is proportional to the square or the cube of the wind speed, depending on the flow type characterized by the Reynolds number, and the elastic energy is proportional to the square of the strain, assuming that the fiber obeys to Hooke's law.

We can recover the power law relation between the wind speed and the RMS strain with the following reasoning. We assume that $\partial_t E = -a \cdot E^p + b \cdot ws^q$ where E is the elastic energy, ws the wind speed, and a, b, p, q are constants. This means that the power provided by the wind to the fiber (and converted in elastic energy) is proportional to ws^q, and E is dissipated at a rate proportional to E^p. Assuming further that the stationary regime is reached much faster than the timescale of the variations of the wind speed, we can ignore the time derivative and obtain $E = (ws^q \cdot b/a)^{1/p}$. We then assume a relation between the elastic energy and the RMS strain, $E = k \cdot strain^r$, where k and r are constants. We finally obtain $strain = (ws^q \cdot b/a)^{1/(rp)}/k$.

Whereas rational exponents are the norm in physics, due to dimensional analysis, we allow here for any power relation, but the model is still simple and interpretable.

3.1.3 Posterior computation

In Bayesian modelling, we are interested in the probability distribution of the parameters given the data. Since our prior on the parameters will be continuous distributions and the only source of randomness in the model is the i.i.d Gaussian noise term, the posterior distribution has a density with respect to (w.r.t) the Lebesgue measure, and the posterior distribution is entirely described by its density function. We are therefore interested in the posterior density function of the parameters given the data, which we denote by $p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are the wind speed and signal power data, respectively.

Using conditional probabilities, we can write

$$p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}, \mathbf{x})} \tag{3.2}$$

Where,

- $p(\mathbf{y}, \mathbf{x})$ is the probability of the data, which is a constant w.r.t $\boldsymbol{\theta}$, and can therefore be ignored
- $p(\boldsymbol{\theta})$ is the prior on the parameters
- $p(\mathbf{x} \mid \boldsymbol{\theta})$ is a constant w.r.t $\boldsymbol{\theta}$ and can be ignored as well, since we assume that the wind speed and the model parameters are independent. This assumption actually means that the wind speed is not affected by the cable, which is a reasonable approximation
- $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$ is the likelihood of the signal power given the wind speed and the parameters.

We therefore have:

$$p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{3.3}$$

and since the stochastic part in $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$ is only due to $\boldsymbol{\epsilon}$, $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = p(\boldsymbol{\epsilon} \mid \boldsymbol{\theta})$, ie:

$$-\ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \frac{n}{2} \ln(2\pi\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^{n} (y(t_i, \mathbf{s}_i) - \beta_0 - \beta_1 \cdot x(t_i, \mathbf{s}_i)^{\alpha})^2 \tag{3.4}$$

where n is the number of data points and ln the natural logarithm, using the probability density function of the normal distribution.

3.1.4 Prior choice

We consider independent priors on the parameters, which we detail below:

- $\alpha \sim \Gamma(\alpha = 2, \beta = 1)$. α is positive and should not be excessively high or low for physical significance.
- $\beta_0 \sim \text{Exp}(\lambda = \text{mean}(\mathbf{y}))$. β_0 is positive, with a preference for small values relatively to the mean of the RMS strain.
- $\beta_1 \sim \text{Lognormal}\left(\mu = \ln\left(\frac{\text{mean}(\mathbf{y})}{\text{mean}(\mathbf{x})}\right), \sigma^2 = 3.\right)$. β_1 is positive, with a preference for values close to the ratio of the means of the RMS strain and the wind speed, which would be the value of β_1 assuming that $\alpha = 1$ and that $x = 0 \Rightarrow y = 0$ and $x = \text{mean}(x) \Rightarrow y = \text{mean}(y)$. Yet, it allows for a wide range of variation in orders of magnitude since we do not have a strong prior belief on the value of β_1.
- $\tau \sim \text{InvGamma}(\alpha = 1, \beta = \text{std}(y))$. τ is positive and should be around the standard deviation of the RMS strain since the data is very noisy, without being too informative with a shape parameter of 1. The choice of an inverse gamma prior for the noise variance is common in Bayesian statistics since it is the conjugate prior, which means that the posterior distribution is also an inverse gamma distribution if the likelihood is Gaussian.

3.2 Fractional model

3.2.1 Model description

Another model we can consider is a quotient of power laws:

$$y_{\mathbf{s},t} = \beta_0 + \beta_1 \cdot \frac{(x_{\mathbf{s},t}/\alpha_1)^{\alpha_0}}{1 + (x_{\mathbf{s},t}/\alpha_1)^{\alpha_0}} + \epsilon_{\mathbf{s},t}, \quad \epsilon_{\mathbf{s},t} \overset{i.i.d}{\sim} \mathcal{N}(0, \tau^2), \tag{3.5}$$

where α_0, α_1, β_0, and β_1 are parameters to be estimated and $\epsilon_{\mathbf{s},t}$ is a noise term, here assumed to be Gaussian white noise with zero mean and variance τ^2, since the mean is already captured by β_0. All parameters can be grouped in a vector $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_0, \beta_1, \tau)$, just as we did with the power-law model with noise.

3.2.2 Motivation

This fractional model is similar to the power-law model, yet allows a change of regime in the response of the DAS to the wind. This is physically sound since the fiber can react differently to the wind depending on the wind speed. For instance, in fluid mechanics, depending on the Reynolds number, the drag force is proportional to the square of the wind speed or to the wind speed itself. This formulation as a quotient is also quite common in frequency response functions, where the response of a system to a forcing is expressed as a ratio of polynomials in the forcing frequency, and we could imagine that the fiber has a similar response to the wind, oscillating with a frequency that depends on the wind speed.

3.2.3 Posterior computation

Following the same reasoning as for the power-law model in Section 3.1, we are interested in $p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y})$, and we have:

$$p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{3.6}$$

$$-\ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \frac{n}{2} \ln(2\pi\tau^2) - \frac{1}{2\tau^2} \sum_{i=1}^{n} \left(x(t_i, \mathbf{s}_i) - \beta_0 + \beta_1 \cdot \frac{(x_{\mathbf{s},t}/\alpha_1)^{\alpha_0}}{1 + (x_{\mathbf{s},t}/\alpha_1)^{\alpha_0}} \right)^2 \tag{3.7}$$

where n is the number of data points.

3.2.4 Prior choice

We use a prior very similar to the one used for the power-law model in Section 3.1, with independent priors on the parameters:

- $\alpha_0 \sim \Gamma(\alpha = 2, \beta = 0.5)$. The rate parameter is lower than for the power-law model, which means that we allow higher values of α_0 in the fractional model. This is because the fractional model will not explode for high wind speed with high values of the exponent parameters as the power-law model would.
- $\alpha_1 \sim \text{Lognormal}(\mu = \text{mean}(\ln(\mathbf{x})), \sigma^2 = 2\text{std}(\ln(\mathbf{x})))$. α_1 is positive with a preference for values close to the geometric mean of the wind speed, and in the range of measured wind speed values. We use a logarithmic scale since it is a scaling factor.
- $\beta_0 \sim \text{Exp}(\lambda = \text{mean}(\mathbf{y}))$, $\beta_1 \sim \text{Lognormal}\left(\mu = \ln\left(\frac{\text{mean}(\mathbf{y})}{\text{mean}(\mathbf{x})}\right), \sigma^2 = 3.\right)$, and $\tau \sim \text{InvGamma}(\alpha = 1, \beta = \text{std}(y))$, same as the power-law model, Section 3.1

3.3 Power law with spatially varying parameters

3.3.1 Model description

The model presented in Section 3.1 can be extended to allow for spatially varying parameters on β_0 and β_1. The model becomes:

$$y_{\mathbf{s},t} = \beta_{0,\mathbf{s}} + \beta_{1,\mathbf{s}} \cdot x_{\mathbf{s},t}^{\alpha} + \epsilon_{\mathbf{s},t}, \quad \epsilon_{\mathbf{s},t} \overset{i.i.d}{\sim} \mathcal{N}(0, \tau^2), \tag{3.8}$$

and rewriting in matrix form, vectorizing on channels:

$$\mathbf{y}_t = \mathbf{F}_t\boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{i.i.d}{\sim} \mathcal{N}(\mathbf{0}, \tau^2\mathbf{I}_N), \tag{3.9}$$

where

$$\begin{aligned}
\mathbf{y}_t &= (y_{1,t}, \dots, y_{N,t})^\top \in \mathbb{R}^N, \\
\mathbf{F}_t &= [I_N, \mathrm{diag}(x^\alpha_{1,t}, \dots, x^\alpha_{N,t})] \in \mathbb{R}^{N\times 2N}, \\
\boldsymbol{\beta} &= (\beta_{0,1}, \dots, \beta_{0,N}, \beta_{1,1}, \dots, \beta_{1,N})^\top \in \mathbb{R}^{2N}, \\
\boldsymbol{\epsilon}_t &= (\epsilon_{1,t}, \dots, \epsilon_{N,t})^\top \in \mathbb{R}^N.
\end{aligned}$$

We also assume a prior on $\boldsymbol{\beta}$ given $\boldsymbol{\theta}$:

$$\boldsymbol{\beta} \mid \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}\right) := \mathcal{N}\left(\begin{pmatrix}\mu_0\mathbf{1}_N \\ \mu_1\mathbf{1}_N\end{pmatrix}, \begin{pmatrix}\sigma_0^2\mathbf{R}(\phi) & \rho\sigma_0\sigma_1\mathbf{R}(\phi) \\ \rho\sigma_0\sigma_1\mathbf{R}(\phi) & \sigma_1^2\mathbf{R}(\phi)\end{pmatrix}\right), \tag{3.10}$$

where the spatial correlation $\mathbf{R}_{ij}$ is given by Matérn-3/2 type covariance function:

$$R_{ij} = \left(1 + \sqrt{3}\phi|s_i - s_j|\right)\exp\left(-\sqrt{3}\phi|s_i - s_j|\right), \quad i,j = 1,\dots,N, \tag{3.11}$$

and $\boldsymbol{\theta}$ regroups all parameters:

$$\boldsymbol{\theta} = (\alpha, \mu_0, \mu_1, \rho, \sigma_0, \sigma_1, \phi, \tau). \tag{3.12}$$

3.3.2 Motivation

This form of model with spatially varying parameters is inspired by the fact that the fiber can react differently to the wind depending on its location. For instance, the fiber can be more or less exposed to the wind, or the wind can be more or less turbulent at different locations. This model could better capture this kind of phenomena than a model with constant parameters. However, we do not want α to vary spatially, since it is a physical parameter that should be the same for all the fiber. The noise variance τ^2 could vary, but we assume it to be constant over space for simplicity.

Having spatial correlation between the parameters is beneficial since it can help in estimation when the data is scarce since the parameters can borrow information from neighboring locations. It is also a way to regularize the estimation of the parameters, and is physically sound since the parameters should not vary too much from one location to a neighboring one.

The choice of a Gaussian process prior on the parameters $\boldsymbol{\beta}$ knowing $\boldsymbol{\theta}$ was inspired by Banerjee et al., 2014. Gaussian distributions also have the advantage of being relatively easy to work with. Finally, the Matérn-3/2 covariance function is a common choice for the spatial correlation, and it is a good compromise between smoothness and flexibility.

3.3.3 Posterior computation

Once again, $p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y})$ is of interest and the relation $p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})$ holds. However, the likelihood $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$ is more complex than in the previous models, since it involves the latent coefficients $\boldsymbol{\beta}$:

$$p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \int p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\theta})p(\boldsymbol{\beta} \mid \boldsymbol{\theta})d\boldsymbol{\beta}. \tag{3.13}$$

This requires to integrate over $\mathbb{R}^{2N}$, which is computationally intractable. We can however use another expression of the likelihood, that stem from the definition of conditional probabilities and holds for any $\boldsymbol{\beta} \in \mathbb{R}^{2N}$:

$$p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\beta} \mid \mathbf{x}, \boldsymbol{\theta})}{p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta})}. \tag{3.14}$$

Letting details of the computation of the likelihood in the Appendix 8.3, we finally have:

$$\begin{aligned}\ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = &- \frac{NT}{2} \ln \tau^2 - \frac{1}{2\tau^2} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}})^\top (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}) \\ &- \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}| - \frac{1}{2} (\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}}) \\ &+ \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}| + C,\end{aligned}$$

where C is constant w.r.t $\boldsymbol{\theta}$ and therefore plays no role in the estimation of $\boldsymbol{\theta}$, and:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} = \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{F}_t \right]^{-1}, \quad \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{y}_t \right].$$

3.3.4 Prior choice

We again consider independent priors on the parameters, with detail below:

- $\mu_0 \sim \text{Exp}(\lambda = 1/\text{mean}(\mathbf{y}))$, $\mu_1 \sim \text{Lognormal}\left(\mu = \ln\left(\frac{\text{mean}(\mathbf{y})}{\text{mean}(\mathbf{x})}\right), \sigma^2 = 3\right)$, $\alpha \sim \Gamma(\alpha = 2, \beta = 1)$, $\tau \sim \text{InvGamma}(\alpha = 1, \beta = \text{std}(y))$, same as β_0, β_1, α and τ in the power-law model, Section 3.1
- $\rho \sim \mathcal{U}([-1, 1])$. The correlation coefficient should be between -1 and 1, but we do not add more information.
- $\sigma_0 \sim \text{Exp}(\lambda = 1/\text{mean}(\mathbf{y}))$. σ_0 is positive, with a preference for small values relatively to the mean of the RMS strain, just as for μ_0.
- $\sigma_1 \sim \text{Exp}\left(\lambda = \frac{\text{mean}(\mathbf{x})}{\text{mean}(\mathbf{y})}\right)$. σ_1 is positive, with a preference for values small relatively to our initial guess for mu_1.
- $\phi \sim \text{Lognormal}(\mu = \ln(4), \sigma^2 = 3.)$. ϕ is positive, with a preference for values around 4 channels, which corresponds to approximately 100 meters. The logarithmic scale 3 allows for a wide range of variation in orders of magnitude.

Chapter 4

Parameter estimation methodology

4.1 Data rescaling

Before any estimation, we systematically rescale the data. This step is widely recommended in order to avoid numerical issues. The rescaling step is usually done by subtracting the mean and then dividing by the standard deviation of the data. However, in our case, the wind speed and the RMS strain are always positive, since our power-law and fractional models would not make sense otherwise: null wind speed should give almost null RMS strain, and a negative number cannot be raised to a non-integer power unambiguously. Therefore, we rescale the data by dividing by the mean value of the data. This rescaling is done for both the wind speed and the RMS strain.

When we estimate parameters, we use the following process: 1. rescaling the data, 2. estimating rescaled parameters, 3. recover the parameters in the original scale.

4.2 Maximum a posteriori estimation

4.2.1 Estimator choice

A first approach to parameter estimation is to use a point estimate. Under good regularity assumptions the maximum likelihood estimate exists and is unique, asymptotically unbiased, convergent, asymptotically normal, asymptotically efficient (Garnier, 2024). Therefore, it appears as a good choice for parameter estimation.

However, it does not include prior information. In a Bayesian context, one would rather use the maximum a posteriori (MAP) estimate, which is the mode of the posterior distribution. It is therefore very similar to the maximum likelihood estimate, but with the addition of prior information. The likelihood estimate is actually a special case of the MAP estimate, where the prior is uniform. We will therefore focus on the MAP estimate in the following, but the likelihood estimate can be obtained by setting the prior to be uniform.

4.2.2 Numerical optimization

The MAP estimate is obtained by solving an optimization problem, where the objective function the posterior probability or posterior probability density. For numerical stability, it is often better to minimize the negative log-posterior, which is equivalent to maximizing the posterior.

We solve this optimization problem with the `scipy.optimize` python library (Virtanen et al., 2020), which provides a wide range of optimization algorithms and shows good performances due to its use of compiled code and `numpy` arrays (Harris et al., 2020). We will now describe the optimization algorithms we used for the different models.

Least-squares estimation

When assuming that the prediction error is Gaussian white noise and using normal or uniform priors on the parameters, the negative log-posterior is a quadratic form. It is actually equivalent to least-squares regression, with the addition of a quadratic regularization term if a normal prior is used. The optimization problem is then a least-squares optimization problem, which is very convenient since efficient algorithms specifically tailored for this optimization problem exist. In this case, we therefore use the `scipy.optimize.least_squares` functionwith the Trust Region Reflective algorithm (default choice for optimization with bounds).

General case

When the optimization problem is not obviously equivalent to a least-square problem, for instance if we use priors that are neither Gaussian nor uniform, we minimize the negative log-posterior with the `scipy.optimize.minimize` function with the Nelder-Mead algorithm, which empirically performed best. We were quite surprised that a derivative-free algorithm was faster and / or more stable than the other methods available, even when computing derivatives with automatic differentiation with the JAX python library (Bradbury et al., 2018).

4.3 Posterior sampling

Point estimates can be useful, but they discard a lot of information. In a Bayesian context, one would rather use the full posterior distribution, which contains all the information about the parameters given the data and the prior. This is especially useful when the data is noisy, as it allows to quantify the uncertainty on the parameters. The posterior distribution is a complex object, and it is often intractable to obtain marginal distributions or point estimates analytically, yet this fully Bayesian approach has gained in popularity with the so-called "Markov Chain Monte Carlo (MCMC) revolution" around 1990 (Banerjee et al., 2014). MCMC methods allow sampling from the posterior distribution without having to compute it explicitly and proved to be a powerful tool for Bayesian inference.

Here, we rely on the `numpyro` python library (Bingham et al., 2018; Phan et al., 2019) to perform MCMC sampling with the No-U-Turn Sampler (NUTS) (Hoffman & Gelman, 2011)

`numpyro` is a probabilistic programming library built on top of `jax` (Bradbury et al., 2018), which provides high-performance linear algebra and automatic differentiation tools, originally developed with deep learning in mind. `numpyro` makes it easier to handle more complex models and leverage advanced techniques such as the NUTS sampler with a high level of abstraction.

The NUTS sampler is a variant of Hamiltonian Monte Carlo which is supposed to allow for more efficient exploration of the parameter space than many other MCMC techniques. It has therefore become a popular choice for Bayesian inference in the last years.

Chapter 5

Estimation on Data

In this chapter, we estimate parameters for the models presented in Chapter 3 using the real-world data presented in Chapter 2 and the methodology presented in Chapter 4.

5.1 Least Squares Regression

We have explained in Chapter 4 that without informative priors, the maximum a posteriori estimate is equivalent to the maximum likelihood estimate. We have also explained that when we assume that the model prediction error is Gaussian white noise, the maximum likelihood estimate is equivalent to the least squares estimate, and that the least squares estimate has the advantage of being computationally efficient.

In this section, we will therefore present the least squares estimate for the parameters of the power law and fractional models presented in Chapter 3. The initial choice and estimation results for the power law model are presented in Table 5.1, and the fractional model in Table 5.2.

	α	β_0	β_1	τ
init.	1	0	$\frac{\text{mean}(\mathbf{y})}{\text{mean}(\mathbf{x})}$	$\text{std}(y)$
	1.00×10^{0}	0	3.67×10^{-8}	1.18×10^{-7}
LS	1.94×10^{-1}	1.09×10^{-19}	1.96×10^{-7}	NA
MAP	1.94×10^{-1}	1.69×10^{-16}	1.96×10^{-7}	1.11×10^{-7}

Table 5.1: Initialization and resulting power law model parameters estimates with least squares (LS) and maximum a posteriori (MAP). The two estimators lead to the same values, except for β_0, but the estimated values for β_0 are negligible compared to τ in both cases.

	α_0	α_1	β_0	β_1	τ
init.	1	$e^{\text{mean}(\ln(\mathbf{x}))}$	0	$\frac{\text{mean}(\mathbf{y})}{\text{mean}(\mathbf{x})}$	$\text{std}(y)$
	1.00×10^{0}	6.10×10^{0}	0	3.67×10^{-8}	1.18×10^{-7}
LS	3.77×10^{0}	2.57×10^{0}	1.40×10^{-7}	1.67×10^{-7}	NA
MAP	3.77×10^{0}	2.57×10^{0}	1.40×10^{-7}	1.67×10^{-7}	1.08×10^{-7}

Table 5.2: Initialization and resulting fractional model parameters estimates with least squares (LS) and maximum a posteriori (MAP).

5.2 Maximum a posteriori estimation

In this section we use the maximum a posteriori estimate for the parameters of the power law and fractional models presented in Chapter 3. The initial choice are the same as for least squares regression (Section 5.1), the estimation results for the power law model are

presented in Table 5.1, and the fractional model in Table 5.2. We also plot the predicted curves and compare them to the observed data in Figure 5.1. It seems that the prior we chose has little influence on the estimation results. This can be explained by the fact the prior was not very informative and that we have a lot of data. Therefore, the likelihood dominates the prior in the posterior distribution.

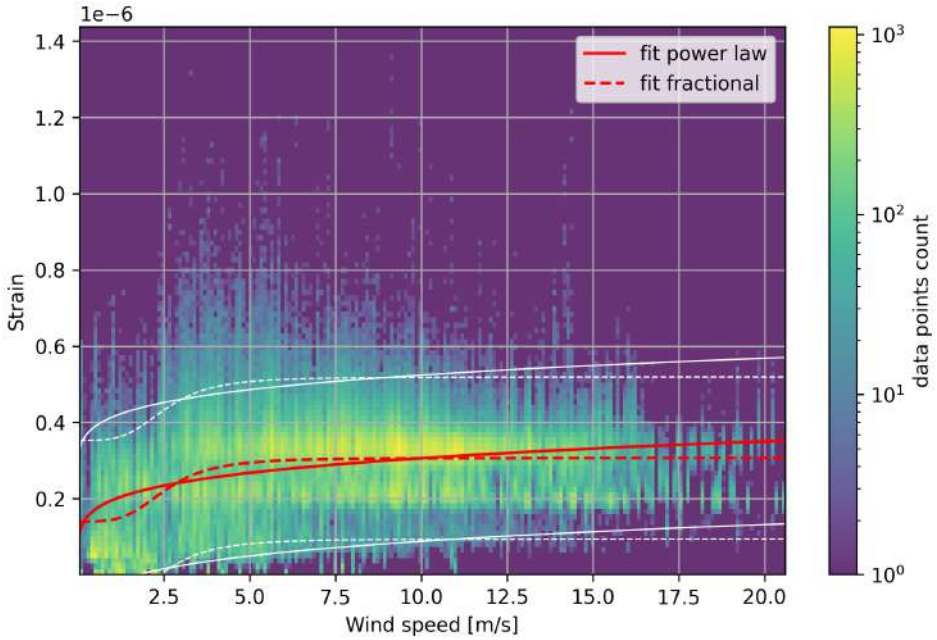

Figure 5.1: Comparison of the observed data and the curves of the power law and fractional models with parameters estimated with posterior maximization (see Tables 5.1 and 5.2). White lines show 0.95 estimated confidence intervals. No visible difference compared to least squares estimation.

5.3 Posterior sampling

As explained in Chapter 4, the maximum a posteriori estimate is a point estimate of the parameters. It is useful for obtaining a single best guess of the parameters, but it does not provide information about the uncertainty of the estimate. In this section, we present the results of the posterior sampling for the power law and fractional models presented in Chapter 3 with the methodology presented in Chapter 4. We run the NUTS sampler with 1 chain with 2000 warmup iterations and 2000 sampling iterations, starting from the parameters maximum a posteriori estimate. We initially wanted to run several chains, sampling the initial location from the prior, but samples of the exponent parameter of the fractional model would perform worse than the maximum a posteriori estimate. This is a sign that the sampler is not exploring the parameter space well and show the difficulty of MCMC, even with advanced samplers like NUTS.

We present summary statistics in Table 5.3 for the power law model and in Table 5.4 for the fractional model. We also plot the posterior distributions pair plots of the parameters for the power law and fractional models in Figures 5.2 and 5.3. Effective sample sizes are computed with `arviz` (Kumar et al., 2019), which is a Python library for exploratory analysis of Bayesian models, with a formula from Vehtari et al., 2019. For both models and every parameter, the effective sample size is above 500, which is a sign that there is a good mixing.

For both models, most parameters samples seem to follow a Gaussian distribution centered around the MAP estimate and have very narrow credible intervals. We interpret this as a sign that the amount of data makes the posterior distribution very peaked, with its

Variable	Median	5%	95%	ESS
α	1.94×10^{-1}	1.93×10^{-1}	1.95×10^{-1}	9.68×10^{2}
β_0	5.48×10^{-11}	4.43×10^{-12}	2.43×10^{-10}	6.82×10^{2}
β_1	1.96×10^{-7}	1.95×10^{-7}	1.97×10^{-7}	1.02×10^{3}
τ	1.11×10^{-7}	1.11×10^{-7}	1.11×10^{-7}	2.11×10^{3}

Table 5.3: Posterior sampling summary statistics of the power law model parameters sampling. 'ESS' stands for effective sample size, and '5%' and '95%' are the 0.05 and 0.95 quantiles of the posterior distribution.

Variable	Median	5%	95%	ESS
α_0	3.77×10^{0}	3.70×10^{0}	3.84×10^{0}	6.68×10^{2}
α_1	2.57×10^{0}	2.55×10^{0}	2.58×10^{0}	7.35×10^{2}
β_0	1.40×10^{-7}	1.39×10^{-7}	1.41×10^{-7}	5.58×10^{2}
β_1	1.67×10^{-7}	1.65×10^{-7}	1.68×10^{-7}	5.45×10^{2}
τ	1.08×10^{-7}	1.08×10^{-7}	1.09×10^{-7}	2.56×10^{3}

Table 5.4: Posterior sampling summary statistics of the fractional model parameters sampling. 'ESS' stands for effective sample size, and '5%' and '95%' are the 0.05 and 0.95 quantiles of the posterior distribution.

weight concentrated around the maximum a posteriori estimate and confined in a region where a second order Taylor expansion of the log-posterior, the Laplace approximation, is a good approximation.

5.3.1 Deviance Information Criterion comparison

The Deviance Information Criterion (DIC) is a measure of the goodness of fit of a model (Banerjee et al., 2014). It is defined as:

$$\begin{aligned} & DIC = 2 \cdot \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y},\mathbf{x}}[D(\boldsymbol{\theta})] - D(\mathbb{E}_{\boldsymbol{\theta}|\mathbf{y},\mathbf{x}}[\boldsymbol{\theta}]) \\ \text{with} \quad & D(\boldsymbol{\theta}) = -2 \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{x}) + 2 \ln h(\mathbf{y}, \mathbf{x}) \end{aligned}$$

where $h(\mathbf{y}, \mathbf{x})$ is "some standardizing function of the data alone" (Banerjee et al., 2014) and therefore plays no role in the comparison of models. The DIC is a trade-off between the goodness of fit of the model and its complexity, and the model with the lowest DIC is favored by the data.

While the fractional model performed better than the power law model in terms of the log-likelihood due to its flexibility, the power law model has fewer parameters and is therefore less complex. When we compute the DIC values with the posterior samples and $h = 0$, we find DIC values of -9.00×10^6 for the power law model and -9.02×10^6 for the fractional model, which means that the fractional model is favored by the data.

5.4 Relevance of spatially varying coefficients

We noticed that the predictive error of the power law model is correlated with spatial location, as shown in Figure 5.4. We therefore designed a power law model with spatially varying coefficients (SVC) presented in Section 3.3. We estimated the parameters of the SVC model using the methodology presented in Section 4. Figure 5.4 show a comparison of the predictive error of the SVC model with the predictive error of the 'simple' power law model, with maximum a posteriori (MAP) estimates of the parameters. We can see that the SVC model has a lower predictive error than the power law model, with a mean absolute residual of 7.15×10^{-8} for the SVC model, and 8.35×10^{-8} for the power law model. The SVC model is clearly more accurate than the simple power law model, yet in terms of computational resources and data, the SVC model is more demanding. The choice

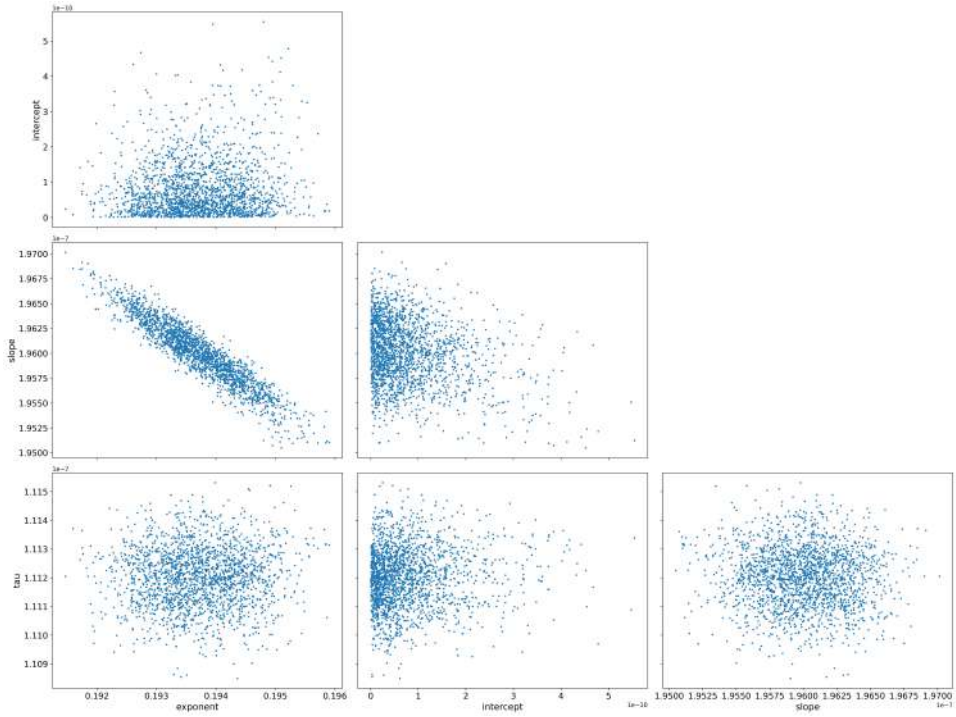

Figure 5.2: Pair plots of the parameters of the power law model, from top to bottom: β_0, β_1, τ, from left to right: α, β_0, β_1. There is strong negative correlation between β_0 and α and dimmer negative correlation between β_0 and β_1. Correlation between these parameters is not surprising, and there is no strong correlation for other pairs of parameters. Samples seem to have an almost Gaussian distribution for each parameter, except for β_0 due to the positivity constraint.

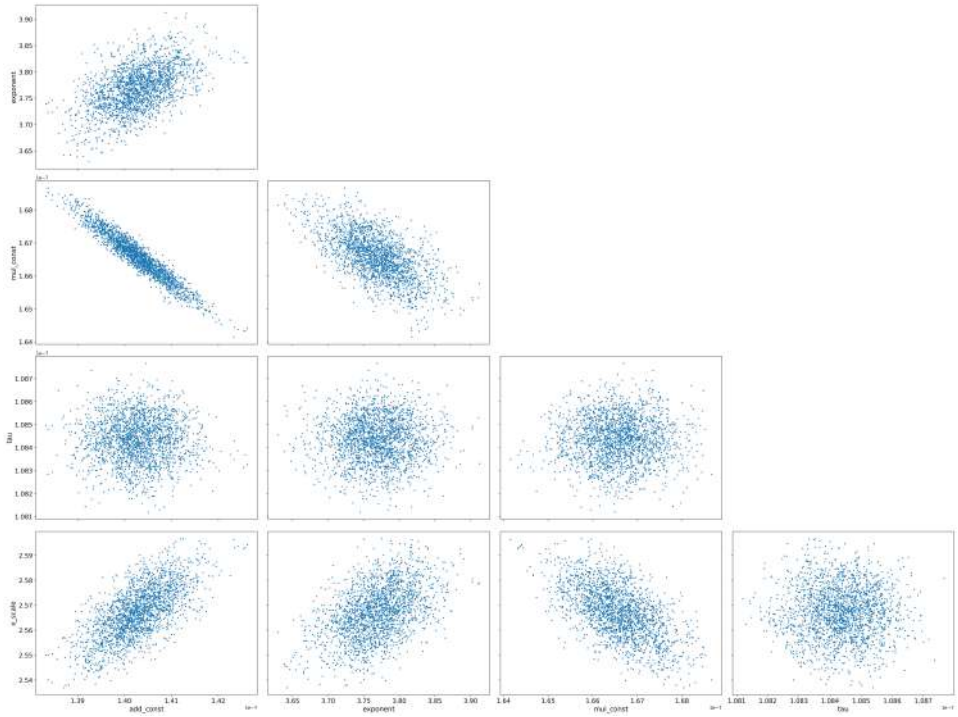

Figure 5.3: Pair plots of the parameters of the fractional model, from top to bottom: α_0, β_1, τ, α_1, from left to right: β_0, α_0, β_1, τ. β_0 and β_1 show a strong negative correlation, τ shows no visible correlation with other parameters, and other parameter pairs show weak correlation. Samples seem to follow an almost Gaussian distribution.

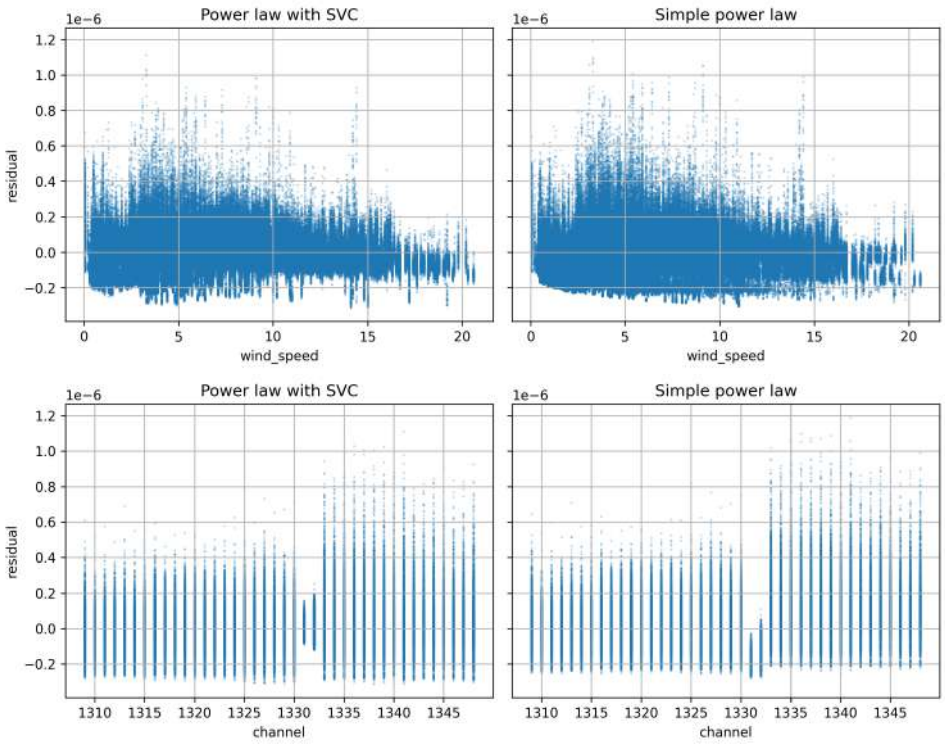

Figure 5.4: Predictive error of the power law models with or without spatially varying coefficients (SVC) against wind speed and channels. The SVC model accounts for the spatial variability of the parameters, and has a lower predictive error than the 'simple' power law model.

of the model should therefore be made according to the desired accuracy and computational resources available.

Chapter 6

Discussion and further work

6.1 Improving accuracy

6.1.1 Lack of predictive power

As shown in Figure 5.1, there is no clear relation between wind speed and RMS strain measured with DAS. Consequently, both the power law model and the fractional model fail to accurately predict the strain given the wind speed. Even when accounting for spatial variations with spatially varying coefficients, the square root of mean squared residuals is 9.74×10^{-8}, while the standard deviation of the strain is 1.18×10^{-7}. This indicates that our models perform only marginally better than a naive constant model that predicts the mean strain for all wind speeds and locations. This striking metric should be weighed against the fact that the concentration of the wind speed helps the constant model, yet the power law model and the fractional model are not very accurate.

6.1.2 Confounding variables

We could likely improve our models. The predictive error of the SVC model is correlated with time, as shown in Figure 6.1. This suggests that incorporating time into our model could reduce the prediction error, similar to how spatially varying coefficients account for spatial correlation in the residuals of the power law model. While we may not want to include time to avoid updating the model's parameters in the future, there might be other confounding variables besides time that could help improve our model.

One variable of interest is wind direction. It is reasonable to believe that wind blowing in the direction of the cable would have a different effect than wind blowing perpendicularly to the cable. However, this requires computing the direction of the wind relative to the cable, adding complexity to our model, especially since we do not have access to the exact location of each channel in the cable. We could imagine learning the geographical orientation of the fiber with an objective function such as

$$\sum_{\mathbf{s},t} (model(x_{\mathbf{s},t}, \hat{\varphi}_{fiber}(\mathbf{s}) - \varphi_{wind}(\mathbf{s}), \boldsymbol{\theta}) - y_{\mathbf{s},t})^2 + L(\hat{\varphi}_{fiber}(\mathbf{s}), \tilde{\varphi}_{fiber}(\mathbf{s}), T)$$

with

$$L(\hat{\varphi}, \tilde{\varphi}, T) = \int_{\mathbf{s}} \alpha |\hat{\varphi}(\mathbf{s} + T(\mathbf{s})) - \tilde{\varphi}(\mathbf{s})| + \beta |\partial_{\mathbf{s}} \hat{\varphi}(\mathbf{s} + T(\mathbf{s})) - \partial_{\mathbf{s}} \tilde{\varphi}(\mathbf{s})| + \gamma |T(\mathbf{s})| + \delta |\partial_{\mathbf{s}} T(\mathbf{s})|$$

where $\varphi_{fiber}(\mathbf{s})$ and $\varphi_{wind}(\mathbf{s})$ are the orientation of the fiber and the wind at channel $\mathbf{s}$. We assume that φ_{wind} is known, and estimate $\varphi_{fiber}(\mathbf{s})$ with $\hat{\varphi}_{fiber}(\mathbf{s})$, based on $\tilde{\varphi}_{fiber}(\mathbf{s})$, the orientation of the fiber computed with linear regression on the extremities of underwater sections, see Chapter 2. T ('T' for 'transport') corresponds to the estimated error of linear mapping from channels to locations described in Chapter 2. L ('L' for 'loss') can be seen as a mixture of Sobolev and Wasserstein distances between $\tilde{\varphi}$ and $\tilde{\varphi}$. $\boldsymbol{\beta}$, $\hat{\varphi}$, and T are learned together while α, β, γ, and δ are hyperparameters to tune. $\hat{\varphi}$ and T must be learned for

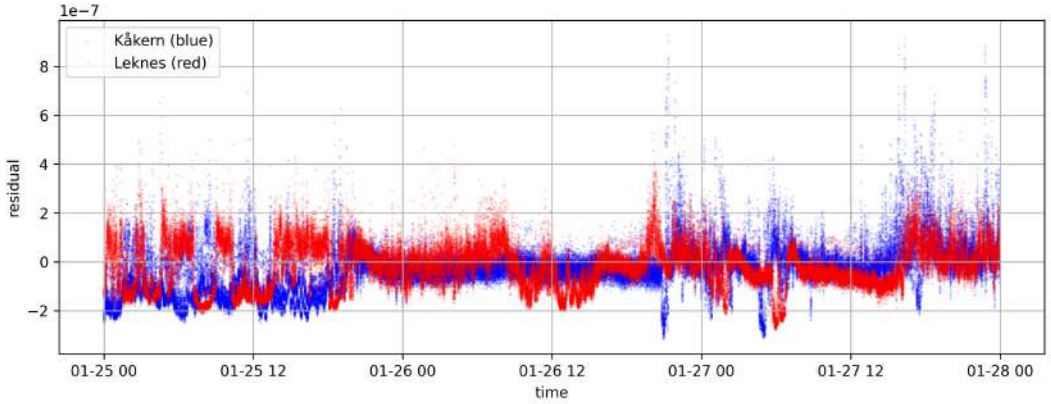

Figure 6.1: Predictive error of the power law model with spatially varying coefficients (SVC) against time. The SVC model accounts for the spatial variability of the parameters, but the predictive error is still correlated with time.

every channel, making this model expensive to train, all the more that α, β, γ, and δ must be tuned as well

We did not try this approach, but we plotted the residuals of our models against the absolute wind direction and could not detect any correlation.

Other conditions, such as the presence of snow or ice on the cable, could influence its response to the wind. For instance, Daniotti et al., 2021 found that the presence or absence of rain was a key factor in oscillations of a bridge in Stavanger. We tried to detect correlations visually between the residuals of our models and the meteorological conditions we had access to: temperature, gusts, and precipitation, but we could not detect any correlation.

6.1.3 Using different data

We only used data from a few days. Using data from a longer period could help us improve our model by encountering a wider range of conditions. We could also use different sources of data, which could account for confounding factors we encountered.

We used one specific processing method for the DAS data. This processing has been designed by experienced researchers and is likely a good trade-off between memory usage and relevant information, but it discards a lot of information that could be valuable. It might be interesting to try to improve this feature extraction from the data. Deep learning in computer vision has shown that learned features can outperform handcrafted ones when enough data is available.

We could also use more information on atmospheric dynamics than just weather station data. We could refine weather data locally using numerical models such as those presented in Benestad et al., 2015; Skamarock et al., 2021. Refined simulations are already produced near Leknes airport for aeronautic purposes and could be reused. The ultimate improvement would be to conduct ad hoc surveys, but this comes with high costs. However, line operators such as Stattnett could be interested in joining.

6.1.4 Improving error structure

A strong assumption made so far is that the wind speed measured by the station corresponds to the true wind speed on the fiber. A better model would account more precisely for causality, assuming that there exists a true wind field in space and time, and that this wind field influences both the DAS signal and the measurements, but that both DAS and weather measurements are also influenced by other factors. This way of causal thinking is particularly emphasized by McElreath, 2020.

Another strong assumption we have made is to assume that the prediction error is Gaussian, which could be false. The Gaussian assumption is often a consequence of the central

limit theorem, but this theorem only applies to distributions with finite variance, while many physical phenomena such as turbulence feature heavy-tailed distributions.

Finally, we assumed that the residuals are independent, which is false, as shown in Figure 6.1. Even if we do not explain this correlation, we could account for it with a surrogate model, as suggested by Gramacy, 2020 and Brynjarsdóttir and O'Hagan, 2014.

6.1.5 Expert knowledge

Expert knowledge is a very useful prior from a Bayesian perspective, as it relies on massive amounts of data observed by humans over centuries. There is certainly more expert knowledge that could be included in the models we have used so far. The mechanics of overhead lines and fluid-structure interactions have been studied extensively (Doaré, n.d.; Ibrahim, 2005; "Stockbridge damper", 2023), as have the dynamics of the atmosphere and meteorological measurements.

The importance of expert knowledge is emphasized in the literature: Brynjarsdóttir and O'Hagan, 2014 insists on the importance of including expert knowledge in the priors, and McElreath, 2020 explains that causal reasoning is based on expert knowledge.

6.1.6 Further mathematical theory

There certainly exist many mathematical tools that could be of interest for this study. For instance, sensitivity analysis could help in identifying critical variables (Garnier, 2024). Bayesian theory could also help in comparing models, with many more tools than the Deviance Information Criterion (DIC) such as the Continuous Rank Probability Score (CRPS) described in Banerjee et al., 2014.

The recent boom in Machine Learning has also led to a flourishing of both theory and numerical methods, and there are certainly many tools coming from there which could be interesting for this study.

6.2 Improving numerical efficiency

Besides improving accuracy, we could also enhance the numerical efficiency of our models.

In the estimation of parameters, we used all available data, which is numerically expensive and not always necessary, especially given the high redundancy and spatial and temporal correlation in the data. We could have used subsets of the data to estimate the parameters, similar to how many deep learning models are trained using stochastic gradient descent, which uses a subset of the data at each iteration instead of the entire dataset. This approach would also allow us to use more data without increasing the computational cost.

Another area for improving numerical efficiency lies in our implementation of spatially varying coefficients. This currently requires solving many linear systems, the size of which is proportional to the number of spatial locations. We could potentially improve this by using the precision matrix of the Gaussian Process instead, which is the inverse of the covariance matrix. This would avoid solving linear systems at each iteration and would benefit from sparsity.

6.3 Use case: learning on local wind patterns

Having a model that compute the likelihood of DAS response given the weather conditions, we can now use it to gain knowledge on the wind given measurements. This is an usual inverse problem, and many methods have been developed for this, among which Markov Chain Monte Carlo (MCMC) methods are prominent (Garnier, 2024). Typically, one think of the Metropolis-Hastings algorithm, where the likelihood given by the model is injected in the acceptance ratio, while the proposal sampling is the critical step. Several techniques have been designed to improve the proposal sampling, among which Hamiltonian Monte Carlo (HMC) is a popular one, and especially its variant No-U-Turn Sampler (NUTS) (Hoffman & Gelman, 2011).

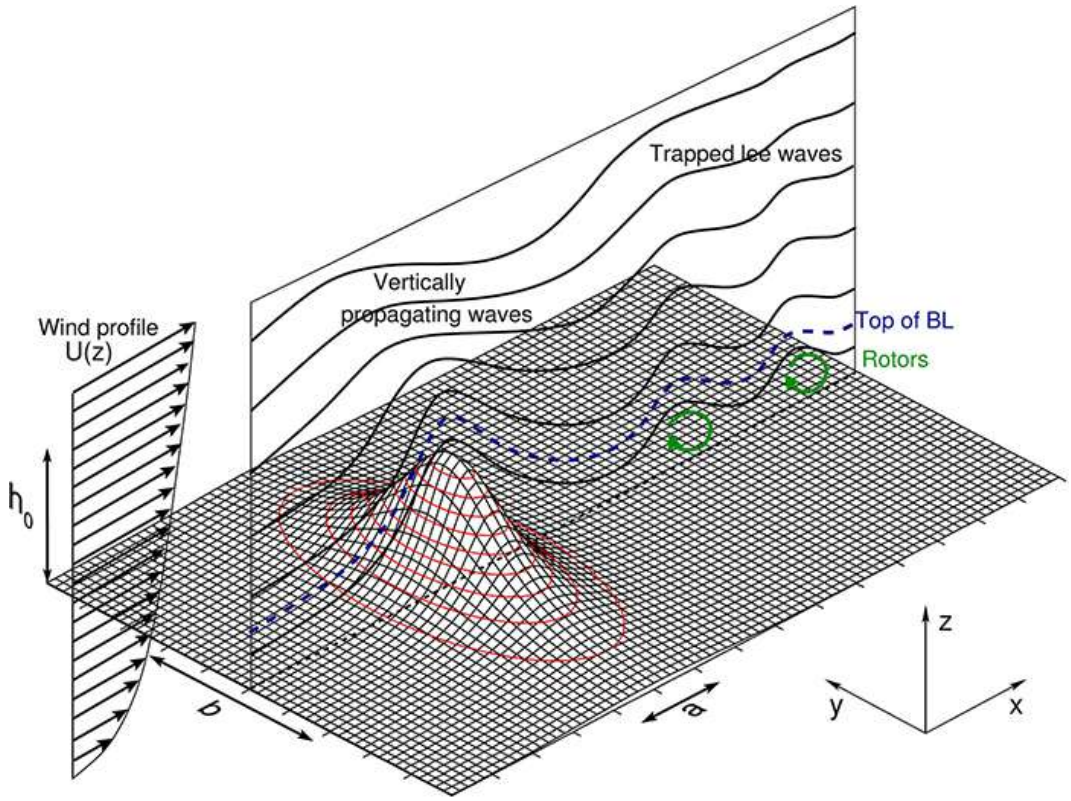

Figure 6.2: Orographic wave, from Teixeira, 2014.

Since we also have physical models for the dynamic of the atmosphere, it is more than welcome to include them. Kalman filters and variational methods are exactly designed for this, and they are widely used in meteorology.

Orographic waves are a kind of meteorological phenomena the study of which could benefit of dense in situ sensors. They can be observed in the lee of mountains, as presented in Figure 6.2, and understanding these phenomena are therefore especially important in mountainous countries such as Norway (**drobinski_meteorologie_nodate**). Numerical weather models in use in Norway struggle to take them into account due to the high resolution it would require. Those predictions are however crucial for the safety of aviation, for instance, so specific local models are used for the surroundings of airports, such as SIMRA (Midjiyawa et al., 2023), now being replaced by models based on convolutional neural network (CNN). The use of dense in situ sensors could help to better take these phenomena into account at a larger scale and improve models that describe them.

Chapter 7

Conclusion

In this study, we explored the correlation between the strain measured by Distributed Acoustic Sensing (DAS) in overhead fiber optic cables and the wind blowing around them. We went beyond the state of the art by deriving quantitative relations based on mathematically simple and physically sound models using power laws. Our focus was on models that compute the likelihood of the strain response given the wind speed, an approach that can be easily incorporated into data assimilation schemes widely used in meteorology.

Through numerical implementation and estimation on experimental data, we demonstrated the relevance of using spatially varying coefficients. However, we also found that the wind speed measured at weather stations, combined with our models, is not sufficient to accurately predict the strain response. We discussed these limitations and other issues, and provided recommendations for future research, such as improving the structuring of noise and incorporating more comprehensive expert knowledge in priors and models.

This work was an enriching experience at the intersection of many fields. The problem we tackled was challenging and largely unexplored, and we hope to have contributed to its understanding.

Bibliography

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/b17115

Benestad, R., Mezghani, A., & Parding, K. (2015). 'esd' - The Empirical-Statistical Downscaling tool & its visualisation capabilities. https://doi.org/10.6084/M9.FIGSHARE.1454425

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., & Goodman, N. D. (2018). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*.

Bouffaut, L., Taweesintananon, K., Kriesell, H. J., Rørstadbotnen, R. A., Potter, J. R., Landrø, M., Johansen, S. E., Brenne, J. K., Haukanes, A., Schjelderup, O., & Storvik, F. (2022). Eavesdropping at the Speed of Light: Distributed Acoustic Sensing of Baleen Whales in the Arctic [Publisher: Frontiers]. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.901348

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., & Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs. http://github.com/google/jax

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy [Publisher: IOP Publishing]. *Inverse Problems*, *30*(11), 114007. https://doi.org/10.1088/0266-5611/30/11/114007

CGF. (2024). Annual Report 2023. Retrieved May 14, 2024, from https://www.ntnu.edu/cgf/publications

Daniotti, N., Jakobsen, J. B., Snæbjörnsson, J., Cheynet, E., & Wang, J. (2021). Observations of bridge stay cable vibrations in dry and wet conditions: A case study. *Journal of Sound and Vibration*, *503*, 116106. https://doi.org/10.1016/j.jsv.2021.116106

Doaré, O. (n.d.). Interaction fluide-structure. Retrieved April 19, 2024, from https://sites.google.com/view/interaction-fluide-structure/accueil

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 [Publisher: Nature Publishing Group]. *Scientific Reports*, *7*(1), 11620. https://doi.org/10.1038/s41598-017-11986-4

Garnier, J. (2024). *Gestion des incertitudes et analyse de risque, cours MAP568 de l'école polytechnique*.

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman; Hall/CRC. https://doi.org/10.1201/9780367815493

Harris, C. R., Millman, K. J., Walt, S. J. v. d., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. v., Brett, M., Haldane, A., Río, J. F. d., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy [Publisher: Springer Science and Business Media LLC]. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo [_eprint: 1111.4246]. https://arxiv.org/abs/1111.4246

Hoyer, S., & Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python [Publisher: Ubiquity Press]. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Hoyer, S., Roos, M., Joseph, H., Magin, J., Cherian, D., Fitzgerald, C., Hauser, M., Fujii, K., Maussion, F., Imperiale, G., Clark, S., Kleeman, A., Nicholas, T., Kluyver, T., Westling, J., Munroe, J., Amici, A., Barghini, A., Banihirwe, A., … Wolfram, P. J. (2024). Xarray. https://doi.org/10.5281/zenodo.11183201

Ibrahim, R. A. (2005). Nonlinear vibrations of suspended cables—Part III: Random excitation and interaction with fluid flow. *Applied Mechanics Reviews*, *57*(6), 515–549. https://doi.org/10.1115/1.1804541

Jacquin, L. (2022). *Mécanique des fluides, cours MEC432 de l'école polytechnique*.

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python [Publisher: The Open Journal]. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Lindsey, N. J., Dawe, T. C., & Ajo-Franklin, J. B. (2019). Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing [_eprint: https://www.science.org/doi/pdf/10.1 *Science*, *366*(6469), 1103–1107. https://doi.org/10.1126/science.aay5881

Lindsey, N. J., & Martin, E. R. (2021). Fiber-Optic Seismology [ISSN: 0084-6597, 1545-4495 Issue: 1 Journal Abbreviation: Annu. Rev. Earth Planet. Sci.]. *Annual Review of Earth and Planetary Sciences*, *49*, 309–336. https://doi.org/10.1146/annurev-earth-072420-065213

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9780429029608

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. v. d. Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Midjiyawa, Z., Venås, J. V., Kvamsdal, T., Kvarving, A. M., Midtbø, K. H., & Rasheed, A. (2023). Nested computational fluid dynamic modeling of mean turbulent quantities estimation in complex topography using AROME-SIMRA [Accepted: 2024-05-28T12:45:18Z Publisher: Elsevier]. *1-21*. https://doi.org/10.1016/j.jweia.2023.105497

NetCDF. (n.d.). https://doi.org/10.5065/D6H70CW6

OpenStreetMap contributors. (2024). Planet dump retrieved from https://planet.osm.org [Published: https://www.openstreetmap.org].

Pandas. (2024). https://doi.org/10.5281/zenodo.10957263

Phan, D., Pradhan, N., & Jankowiak, M. (2019). Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *arXiv preprint arXiv:1912.11554*.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., & Huang, X.-y. (2021). A Description of the Advanced Research WRF Model Version 4.3. https://doi.org/10.5065/1dfh-6p97

Stockbridge damper [Page Version ID: 1180728961]. (2023). Retrieved April 19, 2024, from https://en.wikipedia.org/w/index.php?title=Stockbridge_damper&oldid=1180728961

Teixeira, M. A. C. (2014). The physics of orographic gravity wave drag [Publisher: Frontiers]. *Frontiers in Physics*, *2*. https://doi.org/10.3389/fphy.2014.00043

Vehtari, A., Gelman, A., Simpson, D. P., Carpenter, B., & Burkner, P.-C. (2019). Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*. https://api.semanticscholar.org/CorpusID:88522683

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Zhan, Z. (2020). Distributed Acoustic Sensing Turns Fiber-Optic Cables into Sensitive Seismic Antennas. *Seismological Research Letters*, *91*(1), 1–15. https://doi.org/10.1785/0220190112

Zhu, H.-H., Liu, W., Wang, T., Su, J.-W., & Shi, B. (2022). Distributed Acoustic Sensing for Monitoring Linear Infrastructures: Current Status and Trends [Number: 19 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *22*(19), 7550. https://doi.org/10.3390/s22197550

Chapter 8

Appendices

8.1 Code

The code is publicly available on GitHub at https://github.com/theophane-p/stage3a-pub.

8.2 DAS technology

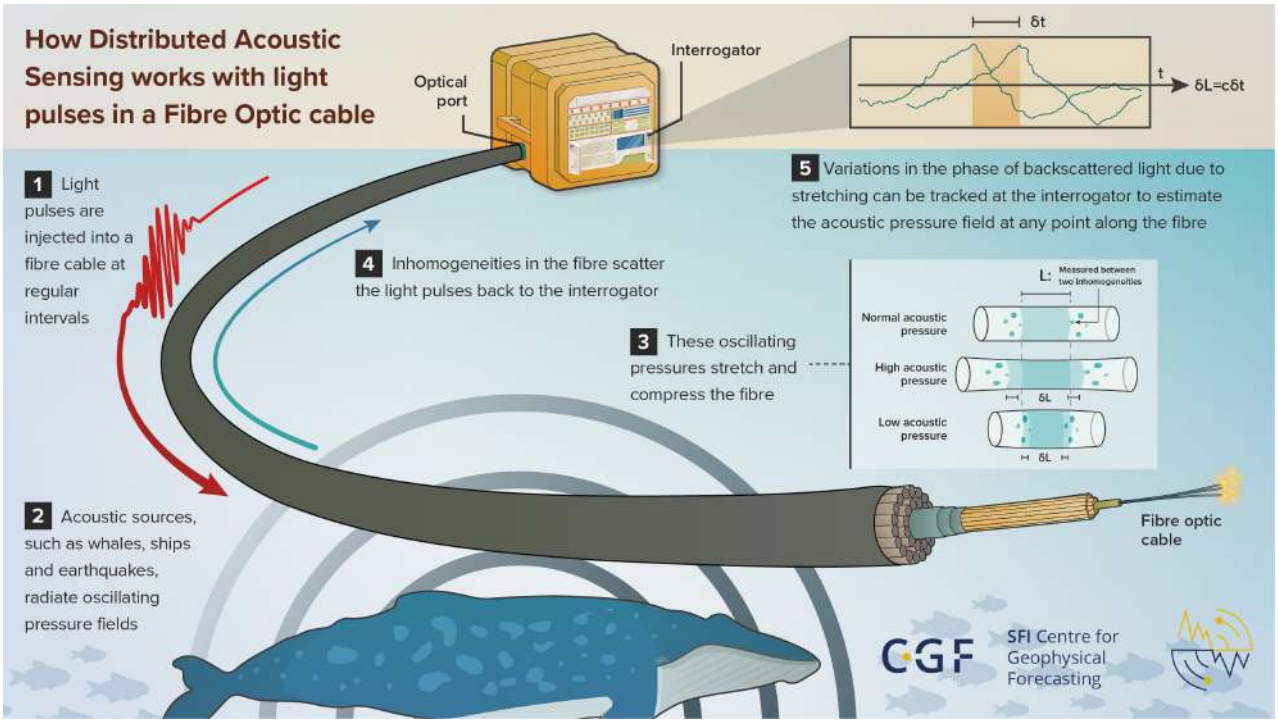

Figure 8.1: Drawing explaining the Distributed Acoustic Sensing (DAS) technology.

8.3 Likelihood computation details for the SVC model

In this appendix, we derive a closed form expression for the probability density function of the likelihood data given parameters $p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$ for the power law model with spatially varying coefficients (SVC). Recalling equation 3.14, that stem from the definition of conditional probabilities and holds for any $\boldsymbol{\beta} \in \mathbb{R}^{2N}$:

$$p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\beta} \mid \mathbf{x}, \boldsymbol{\theta})}{p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta})} \tag{8.1}$$

We now need to compute the three terms in this expression for a given $\boldsymbol{\beta}$.

$p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}, \boldsymbol{\theta})$: This term is the likelihood of the data given the parameters $\boldsymbol{\beta}$. Since $\epsilon_{\mathbf{s},t} \overset{i.i.d}{\sim} \mathcal{N}(0, \tau^2)$ (see Section 3.3):

$$\ln p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}, \boldsymbol{\theta}) = -\frac{NT}{2} \ln \tau^2 - \frac{1}{2\tau^2} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\beta})^\top (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\beta}) \tag{8.2}$$

$p(\boldsymbol{\beta} \mid \mathbf{x}, \boldsymbol{\theta})$: This term is the prior distribution of the parameters $\boldsymbol{\beta}$. Since $\boldsymbol{\beta} \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}})$ (see Section 3.3):

$$\ln p(\boldsymbol{\beta} \mid \mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}| - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}}) \tag{8.3}$$

$p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$: This remaining term in equation 8.1is harder to compute, yet it is possible. Using Bayes' theorem, we have:

$$\ln p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = \ln p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{x}, \boldsymbol{\theta}) + \ln p(\boldsymbol{\beta} \mid \mathbf{x}, \boldsymbol{\theta}) - \ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})$$

and using equations 8.2 and 8.3, we obtain:

$$\ln p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2} \boldsymbol{\beta}^\top \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{F}_t \right] \boldsymbol{\beta} + \boldsymbol{\beta}^\top \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{y}_t \right] + C_1$$

where C_1 is a constant w.r.t $\boldsymbol{\beta}$. From this, we deduce the posterior distribution of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}})$$

and therefore:

$$\ln p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}| - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}) \tag{8.4}$$

where:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} := \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{F}_t \right]^{-1}$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} := \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} \left[\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}} + \frac{1}{\tau^2} \sum_{t=1}^{T} \mathbf{F}_t^\top \mathbf{y}_t \right]$$

We can finally use equations 8.2, 8.3, and 8.4 in the equation 8.1, with $\boldsymbol{\beta} = \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}$ to compute the likelihood of the data given the parameters $\boldsymbol{\theta}$:

$$\begin{aligned}
\ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = & \ln p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}) + \ln p(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} \mid \mathbf{x}, \boldsymbol{\theta}) \\
& - \ln p(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) + C_2 \\
= & -\frac{NT}{2} \ln \tau^2 - \frac{1}{2\tau^2} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}})^\top (\mathbf{y}_t - \mathbf{F}_t \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}) \\
& - \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}| - \frac{1}{2} (\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{\theta}}) \\
& + \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{x},\mathbf{y}}| + C_3
\end{aligned}$$

where C_2 and C_3 are constants w.r.t $\boldsymbol{\theta}$.