14th International Symposium on Turbulence and Shear Flow Phenomena (TSFP14)
Heidelberg, Germany, July 28–31, 2026

SCATTERING MODELS FOR TURBULENCE SPECTRA

Enrico Foglia[1,2]
[1]DAEP, ISAE-Supaero, Toulouse, France
[2]Department of Mechanical Engineering, University of Sherbrooke, Sherbrooke, Canada
enrico.foglia@isae-supaero.fr

Stéphane Moreau[2]
[2]Department of Mechanical Engineering, University of Sherbrooke, Sherbrooke, Canada

Michael Bauerheïm[1]
[1]DAEP, ISAE-Supaero, Toulouse, France

Thierry Jardin[1]
[1]DAEP, ISAE-Supaero, Toulouse, France

ABSTRACT

We present a new methodology for predicting turbulent wall pressure fluctuations statistics by using a generative model based on the scattering transform, as an extension to the one presented recently by Lempereur & Mallat (2026). In particular, we introduce a new estimation algorithm based on the Hutchinson approximation, and an improved sampling strategy. The model is trained on data from DNS of the flow over a CD airfoil and is able to predict the autocorrelation and the coherence length well, with possible applications to trailing-edge noise prediction.

Introduction

The training of deep neural networks on turbulence problems is difficult because of the scarce availability of large, high-quality datasets. The non-linear, multi-scale nature of turbulent fields makes simulations computationally expensive, since to accurately predict the dynamics of large scale structures it is often necessary to account for the presence of the most minute eddies of the flow. And since large datasets, covering a wide range of flow conditions, are needed for the development of accurate and reliable machine learning models, one rightly wonders if such approaches will ever be effective in our field as they currently are in other, more data-intensive, areas of science and engineering. It is surely reassuring that, given enough data, models like GenCast (Lam *et al.* (2023)) are able to out-compete even state-of-the-art numerical solvers, but in most situations collecting datasets of the size of ERA5 would be prohibitively expensive. It is thus necessary to find ways to make machine learning models of fluid flows more data efficient.

Data efficiency can be achieved by imposing a well chosen *inductive bias*—a known structure of the data that the model can exploit. In the case of turbulence, such a structure is given by its multi-scale nature, as it is known that the rich dynamics of turbulent flows is a result of the interaction between small and large eddies. Data of this nature can be efficiently analyzed in terms of the wavelet-scattering transform, which yields a sparse, but highly informative set of statistics constructed from a cascade of linear filtering and non-linear operations (Cheng & Ménard (2021)). These statistics are particularly interesting in this role because they convergence quickly to their expected value, their number grows logarithmically with the field size, and they can capture non-Gaussian behavior. In practice, however, other kinds of statistics are needed, as most engineering problems require statistics in the Fourier domain: as a motivating example, take the phenomenon of noise generation at the trailing-edge of an airfoil. Following Amiet (1976), trailing-edge noise is caused by the scattering of the pressure fluctuations p, which are generated when turbulent structures in the boundary layer cross the trailing edge. This gives rise to a radiating sound field, whose far-field power spectral density at the location $\mathbf{x}$ can be given as:

$$\overbrace{S_{pp}(\mathbf{x},\omega)}^{\text{radiated sound}} \propto \underbrace{\ell_z(\omega)\Phi(\omega)}_{\text{source terms}}\underbrace{|\mathscr{L}(\omega/U_c,x_1,x_2)|^2}_{\text{radiation integral}} \tag{1}$$

To compute the radiated sound it is then necessary to know the correlation length ℓ_z and the autocorrelation Φ of turbulent pressure fluctuations near the trailing edge. The objective of the present paper is to explore a methodology, based on Jaynes' maximum-entropy principle, that allows to transfer the knowledge gathered on the scattering transform values to knowledge about other statistics, in particular the spectrum and the span-wise correlation length.

We compare two methods: the first, which we call *microcanonical*, works by directly matching the known statistics using an optimization process. While effective, microcanonical models are hard to generalize, because they require the knowledge of the target statistics. The second, which we call *macrocanonical*, indirectly matches the moments by using a set of parameters. Its form make it simpler to generalize to unseen flow conditions, which makes it the best candidate for the long term goal of the project.

METHODOLOGY: MAXIMUM-ENTROPY MODELS

Notation

We will use $\pi(p)$ to represent the probability density function of the random variable p. The terms probability density function and probability distribution will be used interchangeably in the rest of the paper. The probability $\pi(p|\mathbf{y})$ indicates the probability of p conditioned on the value $\mathbf{y}$ of another random variable. The expected value of a function $f(p)$ with respect to the distribution π is denoted $\langle f(p)\rangle_\pi = \int \mathrm{d}p\, \pi(p) f(p)$. If no distribution is indicated, we consider the average taken over the dataset.

Maximum-entropy models

We concentrate on the wall-pressure fluctuations $p(t,z;x)$ on an airfoil, at fixed locations along the chord x, where t and z are the time and span-wise location, respectively. Suppose we have access to a set of generalized moments $\mathbf{m}(x) = \langle \mathbf{U}(p)\rangle$, $\mathbf{m} \in \mathbb{R}^d$, computed on a training dataset of fields $p \in \mathbb{R}^n$. The energy functions $\mathbf{U}(p)$ should be chosen to faithfully capture the most important features of the dataset: for the rest of the paper, they will be the scattering covariance statistics, as in (Cheng *et al.* (2024)), which achieve a good tradeoff between sparsity and expressivity. Given this information, we want to estimate the value of other statistics $\langle f(p)\rangle$. For trailing-edge noise modeling, we are interested in the quantities (Roger & Moreau (2005)):

$$\Phi(\omega;x) = \Gamma(\omega,0;x) \tag{2}$$

$$\ell_z(\omega;x) = \frac{1}{2}\int \mathrm{d}\zeta\, \sqrt{\frac{|\Gamma(\omega,\zeta;x)|^2}{\Phi^2(\omega;x)}} \tag{3}$$

called the spectrum and the (span-wise) coherence length, respectively, where Γ is the cross-spectral density:

$$\Gamma(\omega,\zeta;x) = \left\langle \int_{\mathbb{R}} \mathrm{d}\tau\, p(t,z;x)p(t+\tau,z+\zeta;x)e^{-\imath\omega\tau} \right\rangle \tag{4}$$

Following Jaynes (1957), to compute Φ and ℓ_z we proceed by estimating the maximum-entropy probability distribution π, constrained to respect the information we already have, that is, imposing $\langle \mathbf{U}(p)\rangle_\pi = \mathbf{m}$. The entropy of π is defined as (Shannon (1948)):

$$\mathrm{H}[\pi] = -\int \mathrm{d}p\, \pi(p)\log\pi(p) \tag{5}$$

Maximizing Eq. (5), subject to $\langle \mathbf{U}(p)\rangle_\pi = \mathbf{m}$, yields an exponential distribution of the form:

$$\pi(p) = Z(\theta)^{-1}\exp\left(-\theta^\top \mathbf{U}(p)\right) \tag{6}$$

where $Z(\theta)$ is called the partition function. This is what we dub *macrocanonical* model, for its resemblance to the macrocanonical distribution in statistical mechanics. The weights θ can be found by minimizing the Fisher divergence of π with respect to the data distribution, denoted ρ (Hyvärinen & Dayan (2005)):

$$\begin{aligned}\mathscr{L}(\pi) &= \left\langle \|\nabla\log\pi(p) - \nabla\log\rho(p)\|^2 \right\rangle_\rho \\ &= \left\langle \|\nabla\log\pi(p)\|^2 - 2\Delta\log\pi(p) \right\rangle_\rho + \mathrm{cst}\end{aligned} \tag{7}$$

which, for a model like Eq. (6) reduces to solving the linear system:

$$\mathbf{A}\theta = \mathbf{b} \tag{8}$$

$$\mathbf{A} = \left\langle \nabla\mathbf{U}(p)\, \nabla\mathbf{U}^\top(p) \right\rangle,\ \mathbf{b} = \langle \Delta\mathbf{U}(p)\rangle \tag{9}$$

where $\nabla\mathbf{U}(p) \in \mathbb{R}^{d,n}$, so that $\mathbf{A} \in \mathbb{R}^{d,d}$, which is usually manageable if the number of moments is not too large. In practice, solving Eq. (8) can be complicated because of the errors due to the approximation of the expected values by sample averages and because computing $\nabla\mathbf{U}(p)\, \nabla\mathbf{U}^\top(p)$ and $\Delta\mathbf{U}(p)$ might present excessive memory requirements if n is very large. In this work, these issues are dealt with in two, complementary ways:

1. the field p is sampled hierarchically following the algorithm proposed by Marchand *et al.* (2023) and Lempereur & Mallat (2026), which reduces the variance of the estimator considerably;
2. the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are estimated using the Hutchinson method, which can be done at a low memory cost.

In particular, calling $\mathbf{J} = \nabla\mathbf{U}(p)$ and $\mathbf{H}$ the Hessian of $\mathbf{U}(p)$, the estimators for $\mathbf{A}$ and $\mathbf{b}$ are:

$$\begin{aligned}\mathbf{A} &\approx \left\langle \frac{1}{n_{\mathrm{proj}}}\sum_{n=1}^{n_{\mathrm{proj}}} (\mathbf{J}\mathbf{z}_n)(\mathbf{J}\mathbf{z}_n)^\top \right\rangle \\ \mathbf{b} &\approx \left\langle \frac{1}{n_{\mathrm{proj}}}\sum_{n=1}^{n_{\mathrm{proj}}} \mathbf{z}_n^\top \mathbf{H}\mathbf{z}_n \right\rangle\end{aligned} \tag{10}$$

where the $\mathbf{z}_n$ are independent Rademacher random vectors. Because of the estimation technique, the matrix $\mathbf{A}$ is usually rank-deficient. Thus, the solution is given in terms of a Tikhonov-regularized least-squares problem:

$$\theta_\lambda^\star = \mathrm{argmin}\ \|\mathbf{A}\theta_\lambda - \mathbf{b}\|^2 + \lambda^2\|\theta_\lambda\|^2 \tag{11}$$

The present methodology is compared to the original implementation by Lempereur & Mallat (2026), which optimizes Eq. (7) directly by gradient descent using the Adam optimizer (Kingma & Ba (2017)).

Sampling

Sampling of the maximum entropy distribution is performed by following the Langevin diffusion equation, discretized using the Euler-Maruyama scheme:

$$p_{t+1} = p_t + \frac{\Delta t}{2}\mathbf{M}\nabla\log\pi(p_t) + \sqrt{\Delta t \mathbf{M}}Z,\ Z \sim \mathcal{N}(0,\mathbf{I}) \tag{12}$$

where Δt is the time step, Z a normally distributed random variable, and $\mathbf{M}$ a preconditioning matrix. The continuous-time Langevin equation has π as its stationary distribution—that is, in the limit of $t \to \infty$ the p_t are sampled from p_t regardless of the initial conditions—but the time discretization introduces a bias, which can be corrected using the accept-reject scheme of the Metropolis-Hastings algorithm (Geyer (2011)). The combination of these two procedures is called Metropolis-Adjusted Langevin Algorithm, or MALA. The case without preconditioning, where $\mathbf{M} = \mathbf{I}$, can be slow to converge, especially with non-convex probability π, which is the present case. To mitigate this, we employ the adaptive preconditioning strategy proposed by Titsias (2023), where $\mathbf{M}^{-1} = \left\langle \nabla \log \pi \, \nabla \log \pi^\top \right\rangle_\pi$, called the Fisher information matrix. In practice computing and inverting $\left\langle \nabla \log \pi \, \nabla \log \pi^\top \right\rangle_\pi$ is prohibitively expensive, so that $\mathbf{M}$ is estimated iteratively using a symmetric rank-1 update reminiscent of the SR1 optimization algorithm (Nocedal & Wright (2006)). The effectiveness of the preconditioner is examined in the results section.

Microcanonical models

The microcanonical model samples new pressure fields by directly matching the moments $\mathbf{m}$, as described by Cheng *et al.* (2024). It is based on the hypothesis that, all data lives in a small region $\Omega_\varepsilon \subseteq \mathbb{R}^n$ where:

$$\Omega_\varepsilon = \{p : \|\mathbf{U}(p) - \mathbf{m}\| \leq \varepsilon\}$$

and that, within this region, the distribution of p is uniform. This resembles the microcanonical model of statistical mechanics, which assigns equal probability to all microstates having the same macroscopic hamiltonian. The outline of the sampling algorithm is the following: start by sampling a random Gaussian field $p_0 \sim \mathcal{N}(0, \mathbf{I})$. Then, use an optimizer to iteratively change the values at every pixel to match the moments, i.e. to reduce the loss:

$$\mathscr{L}(p) = \|\mathbf{U}(p) - \mathbf{m}\|^2$$

Notice that, for generating a new field p, it is necessary to know the value of the moments $\mathbf{m}$. This represents an obstacle to the use of microcanonical models in situations where the moments are not known in advance. This is why we concentrate on the macrocanonical model, for which a possible generalization is given in the conclusions to this paper.

Dataset

We test the proposed methodology on the wall-pressure fluctuations on the a DNS simulation of a Controlled Diffusion (CD) airfoil at Reynolds number based on the chord $Re = 1.25 \times 10^5$, Mach number $M = 0.25$ and angle of attack $\alpha = 15°$. The compressible Navier-Stokes equations are solved on a structured overset grid using the solver HiPSTAR (Kim & Sandberg (2012)). Derivatives in the span-wise direction are computed using a spectral method, while the other directions are discretized using a fourth-order, wavenumber-optimized compact scheme. The external boundary conditions are imposed from a RANS solution of the full set-up of the corresponding experimental campaign, carried out in the anecoïc wind tunnel at Université de Sherbrooke. Details about the simulation can be found in (Ramo *et al.* (2026)). For the purpose of the present analysis, the wall-pressure fluctuations are saved in 192×192 matrices at fixed cord locations, where the first dimension represents time and the second the span-wise position. As a first step, we test the proposed methodology on a one-level reconstruction problem. The full fields are coarsened to a dimension 96×96 using Symlets with three vanishing moments. We learn to recover the lost high-frequency components using a maximum entropy model using the elements of the scattering covariance matrix as energy functions, as described in (Cheng *et al.* (2024)).

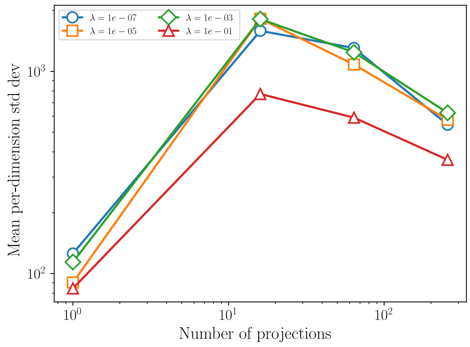

Figure 1. Average standard deviation per dimension of the estimated weights using the regularized least-squares algorithm.

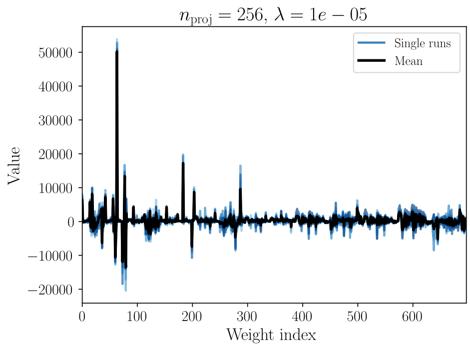

Figure 2. Computed weights θ using the regularized least-squares algorithm using $\lambda = 10^{-5}$ and 2^8 random samples.

RESULTS

Convergence of the estimator for θ

We compare the results of our algorithm, using the Hutchinson approximation followed by a regularized least-squares solve, with the gradient descent solution used by Lempereur & Mallat (2026). The least squares solution is computed using $\lambda = [10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}]$ and sampling $\mathbf{z}$ for 1, 2^4, 2^6, and 2^8 times. The results are shown in Fig. 1. The large variance is dominated by a few very large parameters, as shown in Fig. 2. We attribute the presence of large parameters to the intensive nature of the chosen energy functions: since $\mathbf{U}(p) = 1/d \sum_{z,t} \mathbf{u}(p(z,t))$, for some $\mathbf{u}$, the associated weights tend to scale with d, which in this case be as large as 192^2. While using $\mathbf{u}(p)$ directly as energy functions could be envisioned, this would trade a high variance in the estimation of the weights with that of the imposed moments $\mathbf{m}$. We leave the exploration of this tradeoff for future work. The results for the optimization-based algorithm are shown in Fig. 3. The growth of the variance with the number of steps and with the inverse of the step size testify of the difficulty of finding the optimal set of parameters to train the macrocanonical model using gradient descent on the score matching loss directly.

Finally, Fig. 5 shows the evolution of the computed weights along the chord of the airfoil, between the locations indicated in Fig. 4. We can see that, coherently with the slow evolution of the flow physics in the rear section of the airfoil, the weights evolve in a smooth fashion, even taking noise into consideration. In particular, we observe that the main peaks appear consistently at the same weight index, and have their size is of the same order of magnitude across all locations. While taming the estimation variance is still a challenge, this

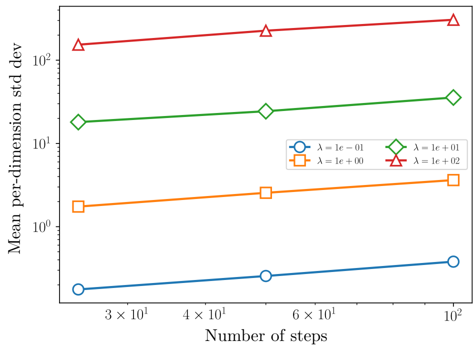

Figure 3. Average standard deviation per dimension of the estimated weights using the Adam-gradient-descent algorithm. Here, λ represents the step size.

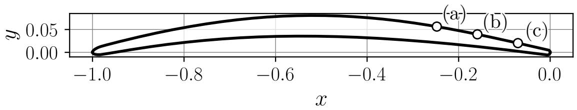

Figure 4. Geometry of the airfoil. The analyzed locations are highlighted with white circles.

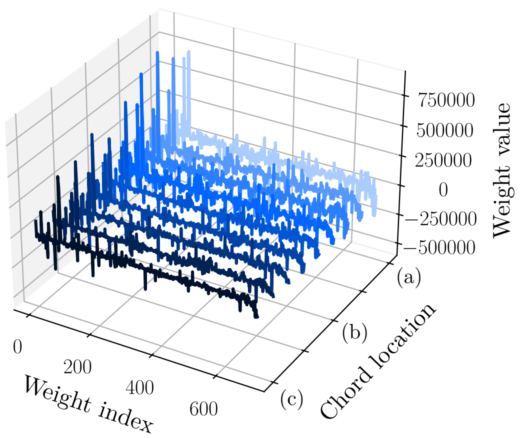

Figure 5. Evolution of the estimated weights along the chord. Weights estimated with $n_{\text{proj}} = 2^8$ and $\lambda = 10^{-3}$.

results confirms the consistency of the new estimation strategy.

Sampling

We use the optimal model estimated with the Hutchinson algorithm, as defined in the previous section, as the starting model. We now compare the sampling performance of the plain MALA sampler to the Fisher-adaptive MALA presented in the methodology. In Fig. 6 and 7 the evolution of the acceptance rate of the Metropolis-Hastings correction with the sampling step is shown, for five different step sizes—$\Delta t = [10^{-4}, 3\times10^{-4}, 10^{-3}, 3\times10^{-3}, 10^{-2}]$—each one with a batch size of 7, repeated 10 times. Notice, neither 0 nor 1 are optimal values for this metric. Indeed, in the first case no new sample is accepted, while in the second the chain is moving too slowly in the probability landscape. For the Langevin proposal, the optimal theoretical acceptance rate is around 0.574, see Roberts & Rosenthal (1998). It is clear that, while even the

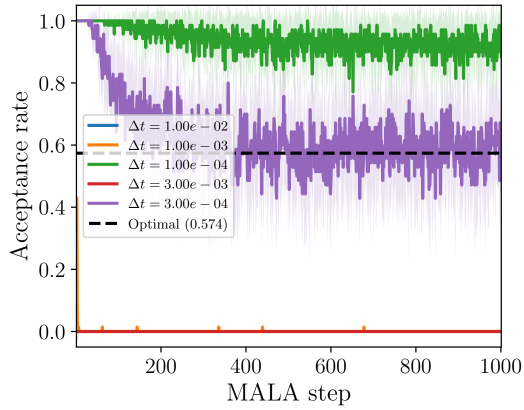

Figure 6. Acceptance rate over time for the simple MALA algorithm.

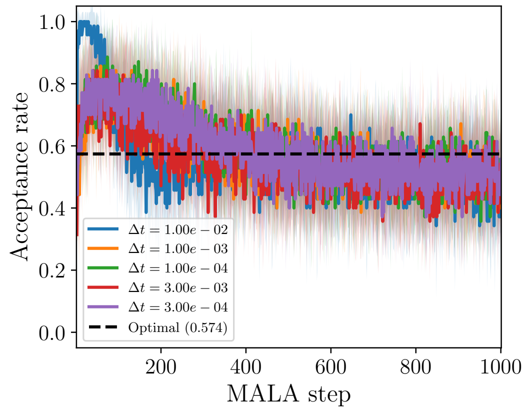

Figure 7. Acceptance rate over time for the simple Fisher-Adaptive MALA algorithm.

simple MALA sampler can achieve performances close to the theoretical optimum, the convergence is not guaranteed. On the other hand, the Fisher-Adaptive MALA can always achieve the optimal acceptance rate, and can converge to it fast even for suboptimal choice of step size.

Reconstruction

The algorithm is applied at three locations on the pressure side of the airfoil, as indicated in Fig. 4. The flow at all the three locations is subject to a moderate-to-intense adverse pressure gradient, but is nonetheless attached, see Ramo *et al.* (2026) for more details on the flow topology. The reconstructed field at the location (c) is shown in Fig. 8. The approximation is visually almost impossible to distinguish from the original, even in the high frequency features. This is confirmed by the analysis of the spectrum and of the coherence length, shown in Fig. 9, where good agreement is observed between the sampled and the original signal. Here, the vertical line separates the part signal that is being reconstructed from the part belonging to the original data.

As shown in Fig. 10, this trend is also observed at the (b) location, but the quality of the reconstruction is lower at the location (a). The second hump in the high frequency at the location (a) is due to the appearance of numerical artifacts, which appear to be highly localized. It is unclear, for the moment, how they arise, and they will be the object of a more in-depth analysis in future works.

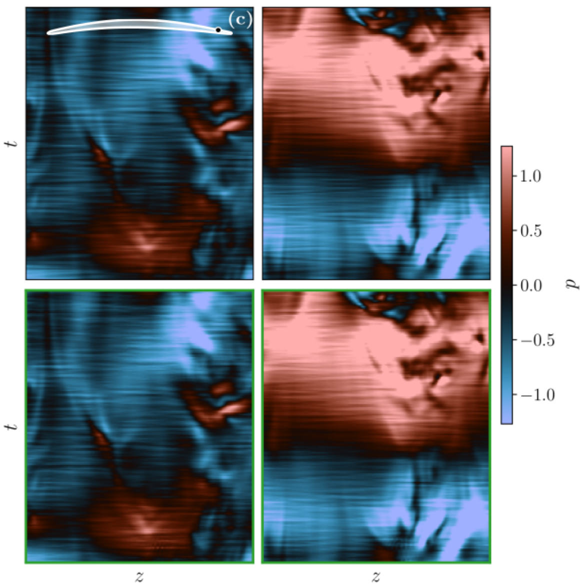

Figure 8. Reconstructed pressure fields at the location (c) using the hierarchical flow model, starting from downsampled samples from the DNS simulation solution. On the top the original pressure fields, and at the bottom the reconstructed ones.

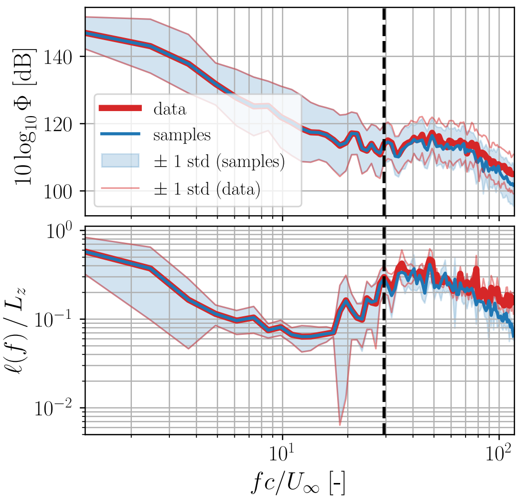

Figure 9. Spectrum (top) and coherence length (bottom) approximations at the location (c). The vertical line indicates where the spectrum is being reconstructed.

To judge if the issue with location (a) comes from an incorrect estimation of the weights or from a poor choice of energy functions, we compare the results with the microcanonical sampling model. For the optimization phase, we use the Adam optimizer for 500 iterations. The interested reader can read more about the connections and the issue of the equivalence of this sampling technique and the maximum-entropy distribution one in (Bruna & Mallat (2019)). The results for the spectrum at locations (a) and (c) are shown in Fig. 11.

CONCLUSIONS AND OUTLOOK

We compared two methodologies for predicting turbulent pressure fluctuations statistics by using generative models based on the scattering transform. It has been shown that this strateg is effective at reproducing the multiscale nature of turbulence, and is able to predict the autocorrelation and the coherence length well, with possible applications to trailing-edge noise prediction. We mostly focused on the extension of the hierarchical macrocanonical mode; proposed by Lempereur & Mallat (2026), using a novel estimation technique for the model weights based on the Hutchinson approximation. Furthermore, the sampling technique has been improved as well by using the Fisher-Adaptive MALA Monte Carlo sampler. Together, these improvements make the model more robust and faster to train and sample. The comparison with the microcanonical sampling model used by Cheng *et al.* (2024) shows that the present methodology is a promising candidate for the estimation of turbulent statistics from data.

At the moment, the model is not capable of estimating the statistics of unseen flow conditions. We plan to extend this methodology in that direction. In particular, this would entail that the probability π must be given conditioning on some flow

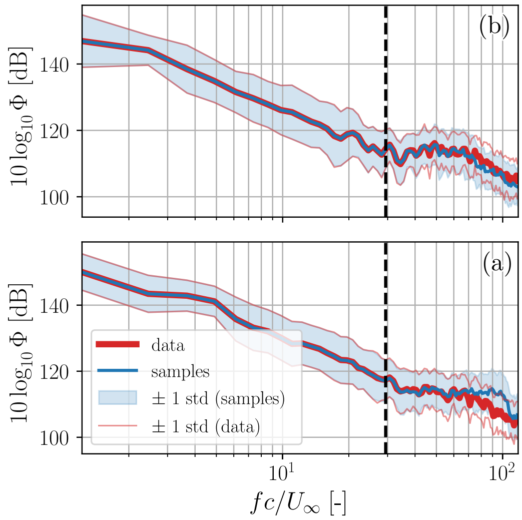

Figure 10. Comparison of the spectra at the locations (b), at the top, and (a), at the bottom.

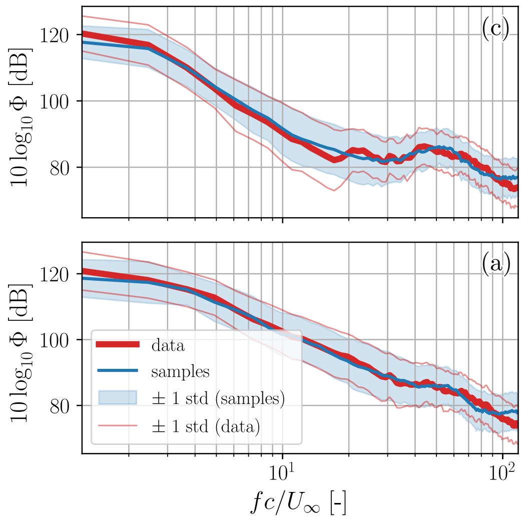

Figure 11. Predicted spectrum using the microcanonical sampling approach at the locations (a) and (c).

parameters $\mathbf{y}$. By Bayes theorem, this can be written as:

$$\pi(p|\mathbf{y}) = \frac{\pi(p)\pi(\mathbf{y}|p)}{\pi(\mathbf{y})}$$

where $\pi(\mathbf{y}|p)$ is a normal regression model that, taken an image of the wall-pressure fluctuations as input, predicts the overall flow parameters. The model $\pi(p)$, in general, does not follow the exponential form given in Eq. (6). For sampling, however, we only need $\nabla \log \pi(p)$, which after some algebra can be given as:

$$\nabla \log \pi(p) = -\langle \theta \rangle_{\theta|p} \nabla \mathbf{U}(p)$$

where the expectation $\langle \theta \rangle_{\theta|p}$ can be computed during the sampling by solving Eq. (8).

REFERENCES

Amiet, Roy K. 1976 Noise due to turbulent flow past a trailing edge. *Journal of sound and vibration* **47** (3), 387–393, publisher: Elsevier.

Bruna, Joan & Mallat, Stéphane 2019 Multiscale sparse microcanonical models. *Mathematical Statistics and Learning* **1** (3), 257–315.

Cheng, Sihao, Morel, Rudy, Allys, Erwan, Ménard, Brice & Mallat, Stéphane 2024 Scattering spectra models for physics. *PNAS nexus* **3** (4), pgae103.

Cheng, Sihao & Ménard, Brice 2021 How to quantify fields or textures? A guide to the scattering transform. ArXiv:2112.01288 [astro-ph].

Geyer, Charles J. 2011 Introduction to markov chain monte carlo. In *Handbook of Markov Chain Monte Carlo* (ed. Steve Brooks, Andrew Gelman, Galin L. Jones & Xiao-Li Meng), pp. 3–48. Chapman and Hall/CRC.

Hyvärinen, Aapo & Dayan, Peter 2005 Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research* **6** (4).

Jaynes, Edwin T 1957 Information theory and statistical mechanics. *Physical review* **106** (4), 620.

Kim, Jae Wook & Sandberg, Richard D 2012 Efficient parallel computing with a compact finite difference scheme. *Computers & Fluids* **58**, 70–87.

Kingma, Diederik P. & Ba, Jimmy 2017 Adam: A method for stochastic optimization.

Lam, Remi, Sanchez-Gonzalez, Alvaro, Willson, Matthew, Wirnsberger, Peter, Fortunato, Meire, Alet, Ferran, Ravuri, Suman, Ewalds, Timo, Eaton-Rosen, Zach, Hu, Weihua *et al.* 2023 Learning skillful medium-range global weather forecasting. *Science* **382** (6677), 1416–1421.

Lempereur, Etienne & Mallat, Stéphane 2026 Hierarchic flows to estimate and sample high-dimensional probabilities. *Applied and Computational Harmonic Analysis* p. 101854.

Marchand, Tanguy, Ozawa, Misaki, Biroli, Giulio & Mallat, Stéphane 2023 Multiscale data-driven energy estimation and generation. *Physical Review X* **13** (4), 041038.

Nocedal, Jorge & Wright, Stephen J 2006 *Numerical optimization*. Springer.

Ramo, Andrea Arroyo, Moreau, Stéphane, Bauerheim, Michaël & Sandberg, Richard D 2026 Direct noise simulation of an installed controlled diffusion airfoil at high angle-of-attack. *International Journal of Heat and Fluid Flow* p. 110272.

Roberts, Gareth O & Rosenthal, Jeffrey S 1998 Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60** (1), 255–268.

Roger, Michel & Moreau, Stéphane 2005 Back-scattering correction and further extensions of amiet's trailing-edge noise model. part 1: theory. *Journal of Sound and vibration* **286** (3), 477–506.

Shannon, Claude E. 1948 A mathematical theory of communication. *The Bell system technical journal* **27** (3), 379–423, publisher: Nokia Bell Labs.

Titsias, Michalis 2023 Optimal preconditioning and fisher adaptive langevin sampling. *Advances in Neural Information Processing Systems* **36**, 29449–29460.