Learning Stochastic Models of Turbulence

Challenging the Log-Normal Hypothesis with Data

Gautier Tandeau de Marsac

DEP MEE, IMT Atlantique, France

December 12, 2025

Outline

The Challenge of Turbulence and K41 Theory

The Fundamental Goal

Understanding the statistical structure of fully developed turbulence remains a major challenge in classical physics.

- **K41 Theory (Kolmogorov, 1941):**
  - Hypothesized a universal local structure based on a constant mean energy dissipation rate $\langle \epsilon \rangle$.
  - Focuses on the longitudinal velocity increment at scale l:

$$\delta_l u(x) = u(x + l) - u(x)$$

  - **Prediction:** Variance scales as a power law $\langle (\delta_l u)^2 \rangle \sim l^{2/3}$ based on dimensional analysis.
- **The Limitation:** Experiments show that small-scale velocity fluctuations are strongly **non-Gaussian** (Intermittency), contradicting the uniform dissipation hypothesis.

K62 Refinement and the Log-Normal Hypothesis

Intermittency Phenomenon

Energy dissipation is not uniform but fluctuates significantly in space and time.

- **K62 Theory (Kolmogorov-Oboukhov):** The logarithm of the local energy dissipation rate, $\log \epsilon_r$, is assumed to follow a **Normal (Gaussian)** distribution.
- **Modern Formalism:** This hypothesis underpins the multifractal formalism used to capture intermittency.
- **Stochastic Model (Chevillard et al., 2012):** Velocity increments are modeled as a stochastic product:

$$\delta_l u \approx x_1 e^{x_2}$$

Where x_1 is a Gaussian noise and e^{x_2} is a log-normal multiplier.

Project Aim

Research Question

Are the components x_1 and x_2 in the relation $\delta_l u = x_1 e^{x_2}$ necessarily Gaussian?

Our Objective:

- Characterize the laws governing velocity increments by **challenging the classical log-normal hypothesis**.
- Adopt a **data-driven approach** to learn the probability distributions directly from experimental data.
- Investigate if a more complex, **non-Gaussian structure** is required to accurately describe turbulence intermittency.

Framework: Multifractal Random Walk (MRW)

We utilize the MRW framework where $\delta_l u \approx x_1 e^{x_2}$.

- x_1 **(Fractional Gaussian Noise):** Represents linear correlations.
- x_2 **(Log-dissipation):** Derived from the convolution of an uncorrelated source noise $n(x)$ over the integral scale L.

The Innovation

Standard theory assumes the source noise $n(x)$ is Gaussian. **We aim to learn its true probability density function (PDF) from data.**

Dataset:

- Modane wind tunnel ($R_\lambda = 2500$, $\langle v \rangle = 20.5$ m/s).
- 256 samples of velocity increment moments across 100 scales (4×10^{-5} s to 0.24 s).

Optimization Strategy

We minimize the Mean Squared Error (MSE) between computed and experimental moments in three stages:

1. **Baseline Log-Normal Model:** Assume $x_1, x_2 \sim \mathcal{N}$. Optimize physical parameters (c_1, c_2) to fit only the variance (S_2).
2. **Higher-Order Constraints:** Incorporate **flatness** into the loss function to test the Gaussian assumption against intermittency metrics.
3. **Learned Distribution (Non-Parametric):**
   - Relax the Gaussian assumption for x_2.
   - Use the **Gumbel-Softmax relaxation** technique to differentiably learn the optimal shape of the $n(x)$ distribution (discretized into 300 bins).

Optimization Phase 1: Variance Fitting

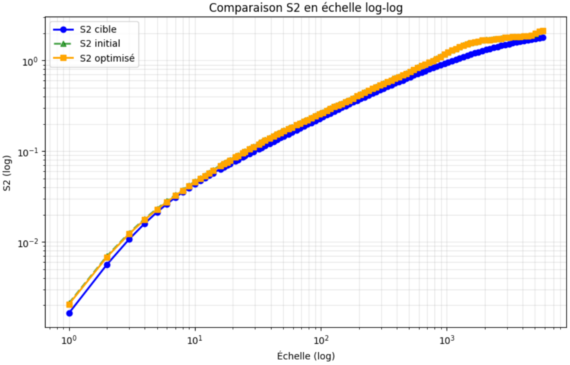


Figure: Variance spectrum S_2 (Log-Log).

Procedure:

- We assume standard Gaussian models for both x_1 and x_2.
- We optimize physical parameters c_1 and c_2 to minimize error on Variance (S_2).

Results:

- **Excellent Fit:** The optimized model (orange) perfectly overlaps the target (blue).
- **Parameters:** Converged to $c_1 \approx 0.33$ and $c_2 \approx 0.03$.
- **Conclusion:** The Gaussian model captures the energy scaling (2nd order) correctly.

Optimization Phase 2: Flatness Mismatch

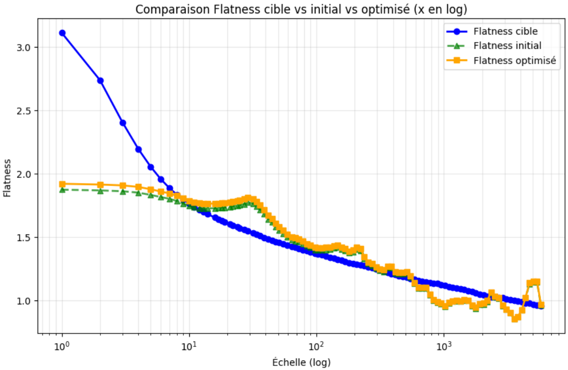


Figure: Flatness factor (Log-Linear).

The Problem:

- Even after optimizing c_1, c_2 on both variance and flatness:
- **Small Scales ($< 10^1$):** Significant deviation. The Gaussian model (orange/green) saturates around $\mathcal{F} \approx 1.9$, while experimental data (blue) reaches > 3.0.
- **Implication:** A purely Gaussian source noise $n(x)$ cannot generate strong enough intermittency.

Need for a learned, non-Gaussian distribution.

Major Result: The Learned Noise Distribution

When the gaussian assumption was relaxed for x_2, the optimization converged to a non-trivial distribution, deviating significantly from the Gaussian prior.

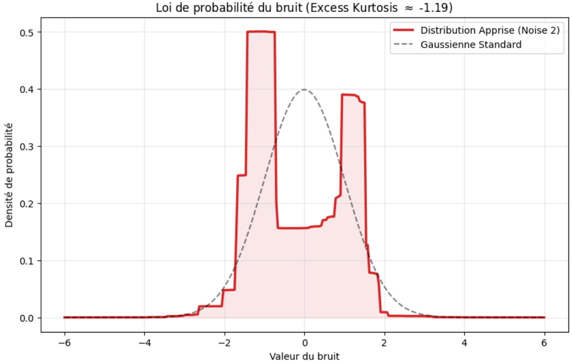


Figure 1: Learned PDF (red) vs Standard Gaussian (dashed).

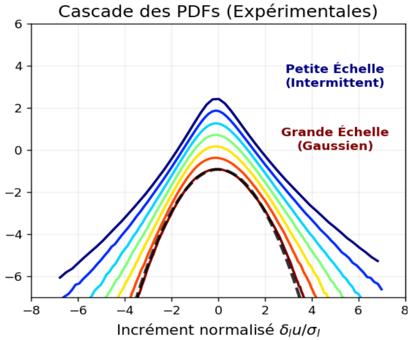
Key Observations:

- **Shape:** Trimodal uniform shape.
- **Statistics:** Negative Excess Kurtosis of -1.19.
- **Implication:** This structural deviation is a key feature required to capture intermittency at small scales.

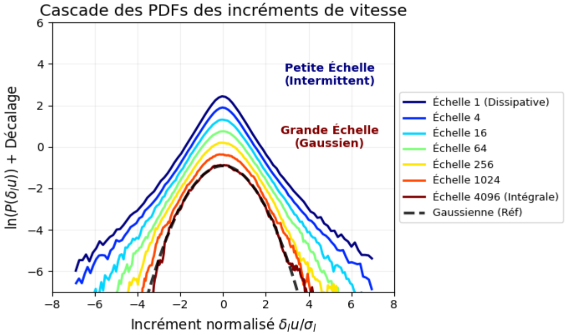
Visualizing Intermittency: The PDF Cascade

Comparison of velocity increment PDFs across scales.

Experimental Data (Target)



Learned MRW Model

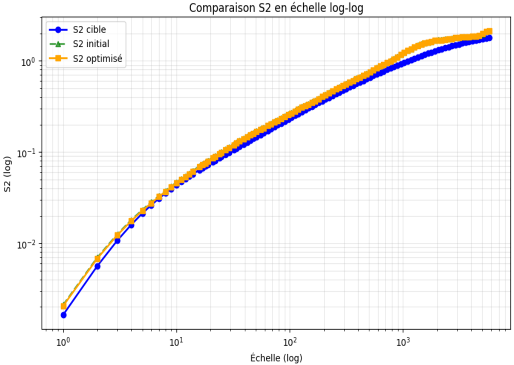


Transition from Gaussian to heavy-tailed PDFs.

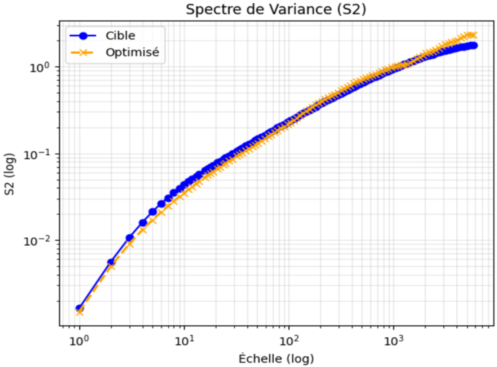
Model Performance: Variance and Flatness

The learned non-Gaussian model achieves remarkable agreement with experimental data.

Variance (Gaussian)



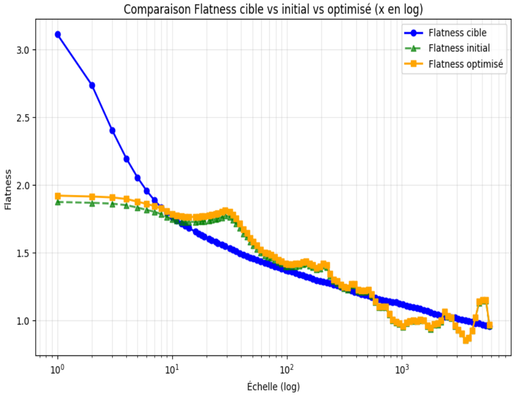
Variance (Learned)



Intermittency coefficient converged to $c_2 \approx 0.02$, consistent with turbulence phenomenology.

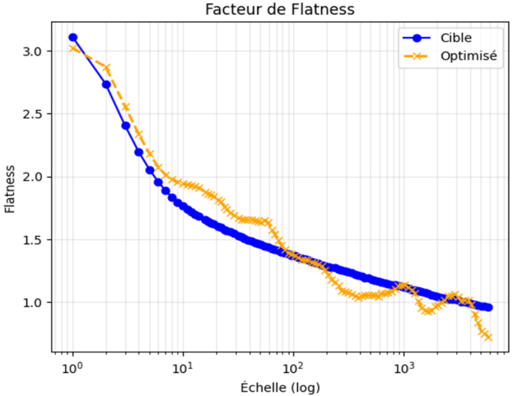
The learned non-Gaussian model achieves remarkable agreement with experimental data.

Flatness (Gaussian)



Significant deviation in low scales

Flatness (Learned)



Accurately captures scale-dependence

Conclusion

- **Challenging the Standard:** We questioned the classical log-normal hypothesis for the intermittency noise $n(x)$.
- **Data-Driven Insight:** By directly learning the PDF from Modane wind tunnel data, we identified a specific **trimodal distribution**.
- **Validation:** This non-Gaussian model successfully reproduces both the variance scaling and the complex behavior of flatness across scales.
- **Takeaway:** Accurate modeling of fully developed turbulence requires relaxing the Gaussian assumption in favor of learned, data-informed distributions.

Thank you for your attention.

Appendix: Learned Trimodal Distribution

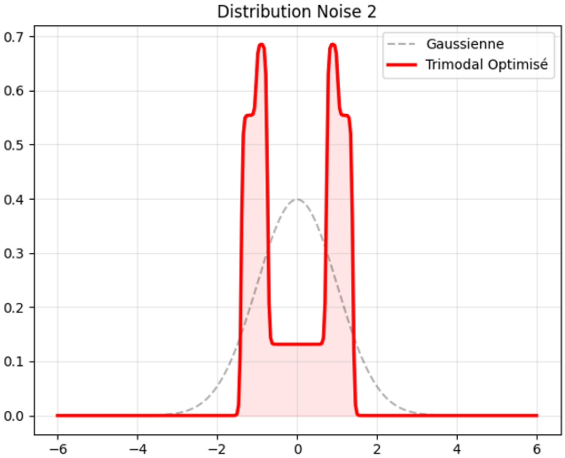


Figure: The optimized probability density function with a uniform-trimodal structure.

Appendix: Moments Fit (Trimodal Model)

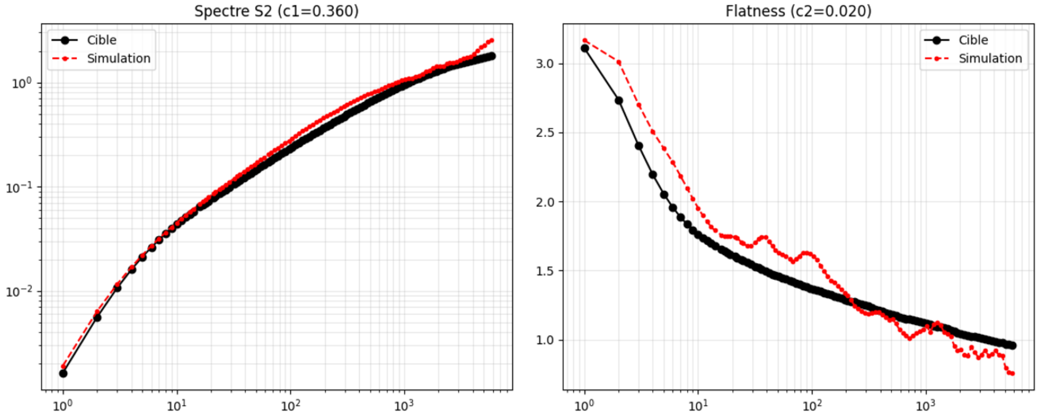


Figure: Comparison of Variance (S_2) and Flatness statistics between the target (black) and the trimodal simulation (red).