



HAL
open science

Analog-Based Forecasting of Turbulent Velocity: Relationship between Unpredictability and Intermittency

Ewen Frogé, Carlos Granero-Belinchon, Stéphane G. Roux, Nicolas B.
Garnier, Thierry Chonavel

► **To cite this version:**

Ewen Frogé, Carlos Granero-Belinchon, Stéphane G. Roux, Nicolas B. Garnier, Thierry Chonavel.
Analog-Based Forecasting of Turbulent Velocity: Relationship between Unpredictability and Inter-
mittency. 2024. hal-04694737v2

HAL Id: hal-04694737

<https://hal.science/hal-04694737v2>

Preprint submitted on 29 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Analog-Based Forecasting of Turbulent Velocity: Relationship between Unpredictability and Intermittency

E. Frogé^{1,2,*}, C. Granero-Belinchón^{1,2}, S. G. Roux³, N. B. Garnier³, and T. Chonavel¹

¹ *Department of Mathematical and Electrical Engineering,
IMT Atlantique, Lab-STICC, UMR CNRS 6285,*

655 Av. du Technopôle, Plouzané, 29280, Bretagne, France.

² *Odyssey, Inria/IMT Atlantique, 263 Av. Général Leclerc, Rennes, 35042, Bretagne, France. and*

³ *Univ Lyon, ENS de Lyon, CNRS, Laboratoire de Physique,
46, allée d'Italie, Lyon, F-69364, Rhône-Alpes, France*

This study evaluates the performance of analog-based methodologies to predict, in a statistical way, the longitudinal velocity in a turbulent flow. The data used comes from hot wire experimental measurements from the Modane wind tunnel. We compared different methods and explored the impact of varying the number of analogs and their sizes on prediction accuracy. We illustrate that the innovation, defined as the difference between the true velocity value and the prediction value, highlights particularly unpredictable events that we directly link with extreme events of the velocity gradients and so to intermittency. A statistical study of the innovation indicates that while the estimator effectively seizes linear correlations, it fails to fully capture higher-order dependencies. The innovation underscores the presence of intermittency, revealing the limitations of current predictive models and suggesting directions for future improvements in turbulence forecasting.

I. INTRODUCTION

For a given temporal stochastic process X , Cramér defined *innovation* to describe the new random inputs into the system and to quantify the unpredictability in the time evolution of X [1, 2]. For a deterministic dynamical system, if a solution exists and is known, the lack of noise leads to a vanishing innovation. However, in chaotic systems, innovation rarely vanishes and can be used to quantify the degree of predictability of the process and to identify unpredictable localized events related to the system's dynamics.

Fluid turbulence is one example of a chaotic system with very complex multiscale dynamics, in which the velocity field is impossible to derive deterministically [3–8]. Turbulent velocity presents long-range correlations, indicating non-local interactions [9], and intermittency leads to extreme events in the velocity gradients [9, 10]. The chaotic dynamics of turbulence showcases extreme sensitivity to initial conditions, where a small noise due to experimental errors or numerical approximations, rapidly amplify [11]. Small thermal fluctuations can also lead to different solutions [12]. Unpredictability in turbulent flows is an inherent feature tied to the stochasticity of their dynamics [13, 14]. As a consequence, predictability in turbulence should be addressed with probabilistic forecasting methods that produce possible future states modeled by a probability density function (PDF) [15].

We propose to characterize the predictability of turbulent velocity in terms of innovation, highlighting that difficult-to-predict events are related to bursts in velocity gradients and thus to intermittency. In this context, we adapt analog-based estimators [11] previously used in meteorological [16, 17] and dynamical systems applications [18] to predict the next value in a hot-wire velocity measurement from a grid turbulence experiment in the Modane wind tunnel [19]. Very importantly the sensitivity of turbulent dynamics to very small disturbances means that close initial conditions may still yield widely different trajectories. This highlights the importance of probabilistic forecasting approaches. Consequently, we use analog-based approaches to estimate the mean and variance of a PDF characterizing the next turbulent velocity, that define respectively the predicted velocity and its associated uncertainty. We compute the corresponding innovation as the difference between the real signal and the mean analog prediction. We then analyze the statistical properties of the velocity, velocity increment, innovation, and the cumulative sum of innovations by examining their second-order structure function and flatness. On the one hand, the second order structure function is a second-order statistic that characterizes the energy distribution across scales. On the other hand, the flatness is a higher-order statistic that characterizes the significance of extreme events across scales.

While the analog-based innovation predictions can effectively capture and predict non-extreme dynamics by removing some linear dependencies, extreme events in velocity gradients due to intermittency are not accurately predicted by the analog method. This method falls short in accounting for higher-order dependencies. The innovation process shows a nearly white spectrum with a non-Gaussian, fat-tailed probability distribution and retains higher-order

* Correspondence: ewen.froge@imt-atlantique.fr

dependencies. Extreme events in innovation are localized in time and appear concomitantly with extreme velocity gradient events. Moreover, these extreme events are also concomitant with higher values of the variance of the analog forecast, and so of the variance of the innovation. Intermittency leads to unpredictability of turbulent velocity in terms of both the mean prediction and its associated uncertainty.

This article is structured into three sections. In section *Turbulence & Intermittency* we discuss intermittency, its characterization, and the turbulent data used. Section *Innovation* introduces the new proposed framework and the estimators we compute. Finally, section *Results* evaluates the performance of the different estimators and discusses the statistics of innovation on real experimental data.

II. TURBULENCE & INTERMITTENCY

The Richardson cascade picture of three-dimensional turbulence identifies three domains of scale: the integral domain, which contains the large scales where energy is injected; the inertial range, where energy cascades from large to small scales; and the dissipative domain where energy is dissipated at smaller scales.

In the case of three-dimensional fully developed turbulence, the multifractal formalism [9, 20, 21] describes in the inertial domain a power law behavior of the structure functions $S_p(l)$ of the turbulent velocity field as a function of the scale:

$$S_p(l) = \mathbb{E}[(\delta_l v)^p] \propto l^{\zeta(p)} \quad (1)$$

where $\delta_l v(x) = v(x+l) - v(x)$ is the velocity increment of size l and $\zeta(p)$ is the scaling exponent.

The Kolmogorov's 1941 (K41) theory [22] originally proposed a linear scaling exponent $\zeta(p) = ph$, with a single Holder exponent $h = \frac{1}{3}$ i.e. a monofractal velocity field. This model does not take into account the intermittency phenomenon highlighted by experiments [19, 23]. Intermittency is characterized by a deformation of the probability density function of the velocity increments from Gaussian at large scales to non-Gaussian with heavy tails at small scales [24]. This behavior reflects the presence of extreme events, which become more significant at smaller scales. It implies multifractality: a non-linear scaling exponent $\zeta(p)$ with a dependency of h upon p $\zeta(p) = ph(p)$. An intermittent model of turbulence was proposed by Kolmogorov and Obukhov (KO62) [25, 26] and later studied within the multifractal formalism [10, 21, 24, 27–29].

The energy distribution across scales is characterized by the second-order structure function $S_2(l)$ which, for scales in the inertial domain, is proportional to $l^{2/3}$ up to intermittent corrections [9]. $S_2(l)$ characterizes second-order statistics of the turbulent velocity field in the same way as autocorrelation and power spectrum [9].

The flatness characterizes the relative importance of extreme events and allows for a measure of the deformation of the PDF. It is defined for a given scale l as

$$F(l) = \frac{S_4(l)}{3S_2(l)^2} \quad (2)$$

It thus characterizes higher-order statistics of velocity. In a monofractal Gaussian field, the flatness remains constant across scales and is equal to 1. In contrast, turbulence shows an increasing flatness with decreasing scale l , indicating a higher prevalence of extreme events or bursts at smaller scales [30–32]. The increased flatness at smaller scales due to the departure from Gaussian statistics is a key indicator of intermittency.

A. Modane Turbulent Velocity Dataset

We use an Eulerian longitudinal velocity measurement from an experimental grid turbulence setup in the Modane wind tunnel [19]. The full velocity measure spans 1000 seconds with a sampling frequency of $f_s = 25$ kHz. Velocity measurements were obtained using hot-wire anemometry. The Taylor scale Reynolds number of the flow is $R_\lambda = 2500$.

To relate temporal measurements to spatial properties, we employ Taylor's hypothesis of frozen turbulence, which allows us to interpret temporal variations in the velocity signal as spatial variations using $v(x, t) = v(x - Vt, 0)$, where v is the velocity, x is the space coordinate in the longitudinal direction, defined along the mean velocity of the flow, t is the time coordinate and $V = 20.5 \text{ m s}^{-1}$ is the mean velocity of the flow.

From previous studies, the integral scale of the flow is $L = 2350dl$ and the Kolmogorov scale [33] of the flow is $\eta = 5dl$ [34], where $dl = V/f_s$. In the following, all results are presented in function of t/T with $T = L/V$.

III. INNOVATION

A. Theoretical Description

For a stochastic process $X = \{x_t\}$ sampled at intervals equally separated by dt , we define the innovation ϵ_t as:

$$\epsilon_t = x_t - \mathbb{E}[x_t|x_{-\infty:t-dt}] \quad (3)$$

where $\mathbb{E}[x_t|x_{-\infty:t-dt}]$ is the expected value of x_t conditioned on its complete past $x_{-\infty:t-dt}$.

We interpret ϵ_t as the novel, unpredictable component of x_t at time t [1, 2]. Thus, innovation acts as a limit to the error of an optimal prediction, providing insight into the degree of unpredictability of the system as a function of time. Very interestingly, innovation can point out specific unpredictable events localized in time.

In practice, the expected value $\mathbb{E}[x_t|x_{-\infty:t-dt}]$ cannot be conditioned on an infinite past and we restrict the conditioning on a finite past interval $t \in [t - pdt, t - dt]$, leading to the approximation:

$$\epsilon_t^{(p)} = x_t - \mathbb{E}[x_t|x_{t-pdt:t-dt}] \quad (4)$$

where p indicates the finite duration of the past horizon.

B. Analogs

For a given time t , we consider the temporal sub-sequence $\vec{x}_t^{(p)}$ of size p of the stochastic process X as:

$$\vec{x}_t^{(p)} = \begin{pmatrix} x_{t-dt} \\ \vdots \\ x_{t-pdt} \end{pmatrix} \quad (5)$$

Two such sub-sequences $\vec{x}_t^{(p)}$ and $\vec{x}_{t'}^{(p)}$ considered at two different times t and t' are called analogs if they are close enough in terms of some given p -dimensional distance.

Based on Poincaré's theorem [35], we assume that 1) analogs exist if the time series is sufficiently long; and 2) close analogs will lead to close successors. Then for a given t , the set of successors $x_{t'}$ of the analogs $\vec{x}_{t'}^{(p)}$ of $\vec{x}_t^{(p)}$ can be used to predict in a statistical way the successor x_t . This approach relies on the similarity of past sequences and their subsequent outcomes for making predictions.

In the context of the present study, analog forecasting serves as a methodological framework for estimating the probability density function $q(x_t|x_{t-pdt:t-dt})$ and its associated expected value $\mathbb{E}[x_t|x_{t-pdt:t-dt}]$ that appears in the definition of innovation.

C. Analog-based predictions and innovations

Given $\vec{x}_t^{(p)}$ and a fixed integer k , we define the analogs of $\vec{x}_t^{(p)}$ as the set $\{\vec{x}_{t'_i}^{(p)}\}_{1 \leq i \leq k}$ of the k historical sequences with $t'_i < t$ that are the closest to $\vec{x}_t^{(p)}$ in terms of the Euclidean distance, as suggested in [36, 37].

The mean $\hat{x}_t^{(p)}$ and variance σ_t^2 of the analog-based estimation of $q(x_t|x_{t-pdt:t-dt})$ are derived from the set of selected analogs and represent respectively the prediction $\hat{x}_t^{(p)} = \mathbb{E}[x_t|x_{t-pdt:t-dt}]$ and its uncertainty. Different estimators for $\hat{x}_t^{(p)}$ and σ_t^2 exist in the literature [36].

D. Average

The original method for analog prediction computes a weighted average of the successors $\{x_{t'_i}\}_{1 \leq i \leq k}$ of the analogs [38, 39]. This leads to the prediction $\hat{x}_t^{(p)}$ with variance σ_t^2 :

$$\hat{x}_t^{(p)} = \sum_{i=1}^k w_{ii} x_{t'_i}, \quad \sigma_t^2 = \sum_{i=1}^k w_{ii} (x_{t'_i} - \hat{x}_t^{(p)})^2 \quad (6)$$

with the corresponding innovation $\hat{\epsilon}_t^{(p)} = x_t - \hat{x}_t$.

The weights w_{ii} are defined as the components of a diagonal weight matrix \mathbf{W} :

$$w_{ii} = \frac{e^{-\|\vec{x}_t^{(p)} - \vec{x}_{t'_i}^{(p)}\|/\lambda}}{\sum_{j=1}^k e^{-\|\vec{x}_t^{(p)} - \vec{x}_{t'_j}^{(p)}\|/\lambda}} \quad (7)$$

with $\lambda = \text{median}(\{\|\vec{x}_t^{(p)} - \vec{x}_{t'_i}^{(p)}\|\}_{1 \leq i \leq k})$, following [36, 37].

This averaging method tends to draw predictions towards the average of the k successors, resulting in estimations of the innovation that are higher in unexplored regions[37].

E. Linear Regression (LR)

Following [36], we then make an analog-based prediction using a weighted linear regression where the analogs are used as explanatory variable and their successors are the explained one.

For a given number k of analogs, the minimization problem can be written as $\min_{\beta_t} \|\vec{y} - \mathbf{W}\mathbf{X}_p\beta_t\|$ where:

- \vec{y} is the vector of analog successors and \mathbf{X}_p is a $k \times (p+1)$ matrix containing all analog vectors with a first column of ones:

$$\vec{y} = \begin{pmatrix} x_{t'_1} \\ \vdots \\ x_{t'_k} \end{pmatrix}, \quad \mathbf{X}_p = \begin{pmatrix} 1 & x_{t'_1-dt} & \dots & x_{t'_1-pdt} \\ \vdots & \vdots & & \vdots \\ 1 & x_{t'_k-dt} & \dots & x_{t'_k-pdt} \end{pmatrix} \quad (8)$$

- \mathbf{W} is a $k \times k$ diagonal matrix which specifies the weight of each analog. The non-zero elements w_{ii} of \mathbf{W} are defined using eq.7.
- β_t is the coefficients vector of the weighted linear regression that depend on t through the definition of the analogs, that we search for.

The ordinary least squares estimator for the minimization gives:

$$\hat{\beta}_t = (\mathbf{X}_p^T \mathbf{W} \mathbf{X}_p)^{-1} \mathbf{X}_p^T \mathbf{W} \vec{y} \quad (9)$$

The analog prediction $\hat{x}_t^{(p)}$ with variance σ_t^2 then reads:

$$\hat{x}_t^{(p)} \equiv \hat{\beta}_t^T \vec{x}_t^{(p)}, \quad \sigma_t^2 = \sum_{i=1}^k w_{ii} (\xi_{t,i})^2 \quad (10)$$

where $\vec{\xi}_t = \vec{y} - \mathbf{W}\mathbf{X}_p\hat{\beta}_t$ are the residues, and the corresponding innovation is $\hat{\epsilon}_t^{(p)} = x_t - \hat{\beta}_t^T \vec{x}_t^{(p)}$.

This linear regression approach in phase space establishes a relationship from past states to their successors based on historical analogs. It has two parameters: the number k of neighbors and the dimension p of the sub-sequences. Note that for the inversion (eq.9) to be possible, one should have $k \geq p$.

F. Normalized LR

One can focus on trends rather than values by removing the mean value of the sub-sequences, allowing the analog search to prioritize patterns over analogs' values [40]. This method is thus identical to the linear regression method after replacing all coordinates x_{t_j} of sub-sequences $\vec{x}_t^{(p)}$ by $x_{t_j} - \bar{x}_t^{(p)}$ with $\bar{x}_t^{(p)} \equiv \frac{1}{p} \sum_{1 \leq j \leq p} x_{t-jdt}$.

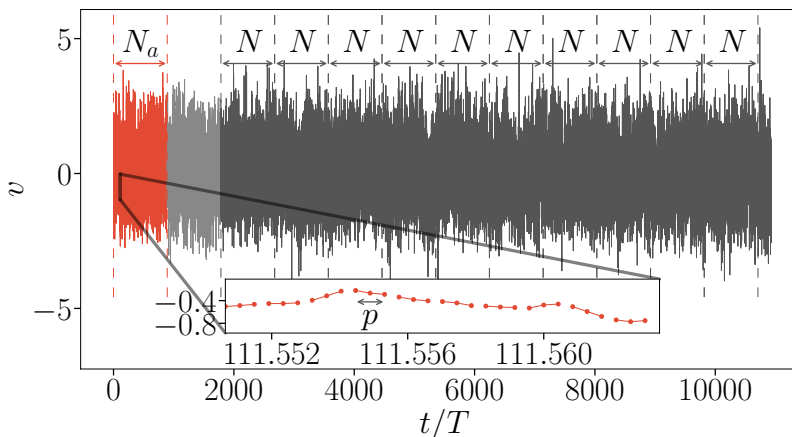


FIG. 1. Visualization of Modane turbulent velocity data usage for the innovation $\epsilon_t^{(p)}$. The first red section of size N_a represents the historical record used for analog search. The subsequent ten black sections are the $n_{\text{real}} = 10$ realizations of size N over which the innovation is computed. Vertical dashed lines indicate the boundaries of the historical record and of the realizations. The inset shows a zoom of the historical record with non-overlapping analogs candidates and a highlighted analog of length $p = 3$.

IV. RESULTS

A. Analogs Prediction

In this section, we compare the performance of the three analog-based predictions presented above when used on Modane turbulent velocity signal v . We also evaluate the impact of k and p in the performance, in order to determine the optimal values for k and p .

We compute the corresponding innovations $\epsilon_t^{(p)}$ for different methodologies and k and p combinations. This is done on $n_{\text{real}} = 10$ realizations of length $N = 2^{21} \simeq 900T$ samples of v , see figure 1. The database containing the historical record for analog searching is chosen to precede and not overlap with the portion over which the predictions and innovations are computed and it contains a total of $N_a = N$ samples, where N_a has been selected to ensure that further increasing the size of the historical record does not significantly affect the results. Overlap between analogs is not permitted, ensuring that the selected analogs are distinct. The results are computed over the n_{real} realizations.

In order to quantify the performance of the estimator over a full time series, we compute the Mean Squared Error (MSE) between the Modane turbulent velocity and the mean prediction $\frac{1}{N} \sum_{j=1}^N (v_j - \hat{v}_j^{(p)})^2$. This MSE is equal to the variance in time of the innovation ϵ_t .

We can consider the naive prediction $\hat{v}_t = v_{t-dt}$ which assumes a locally constant speed and hence doesn't account for any dynamics or even any change in the process. This assumption leads to the MSE: $\frac{1}{N} \sum_{j=1}^N (v_j - v_{j-1})^2$, which is nothing but the variance of the increment $\delta_{dt}v$. As a consequence, we can use the increment as a baseline of the behavior of the innovation $\epsilon_t^{(p)}$ for the three analog-prediction methods presented earlier.

In Figure 2 we compare the performance of the three analog prediction methods as a function of k for different values of p . "Normalized LR" configuration (orange) consistently shows the lowest MSE, indicating superior performance. Regular "LR" (blue) yields similar results. For these two estimators: 1) the MSE increases when p increases due to the curse of dimensionality leading to increased computational complexity and convergence issues and 2) the MSE decreases when k increases until reaching a plateau at 0.026. "Average" (green) method without linear regression shows higher MSE which increases with the number k of neighbors.

Increasing the size p of the sub-sequences over $p = 3$ does not improve performance, even when k is accordingly increased. This is because three-point statistics already capture most of the dynamics of turbulent flows [41].

The lower variance in time of innovation —equal to MSE— compared to the one of the increment (horizontal continuous black line) indicates that predictions are indeed containing relevant information on the dynamics. The existence of a minimum plateau should be interpreted cautiously since we are using only a single-point velocity measurement, which may not probe the full 3D complexity of turbulence. In line with previous works [37], using linear regression for analogs significantly enhances prediction accuracy, with the linear regression model far outperforming averaging methods.

Based on this performance analysis, we use the "Normalized LR" method in the following, with $p = 3$ and $k = 70$.

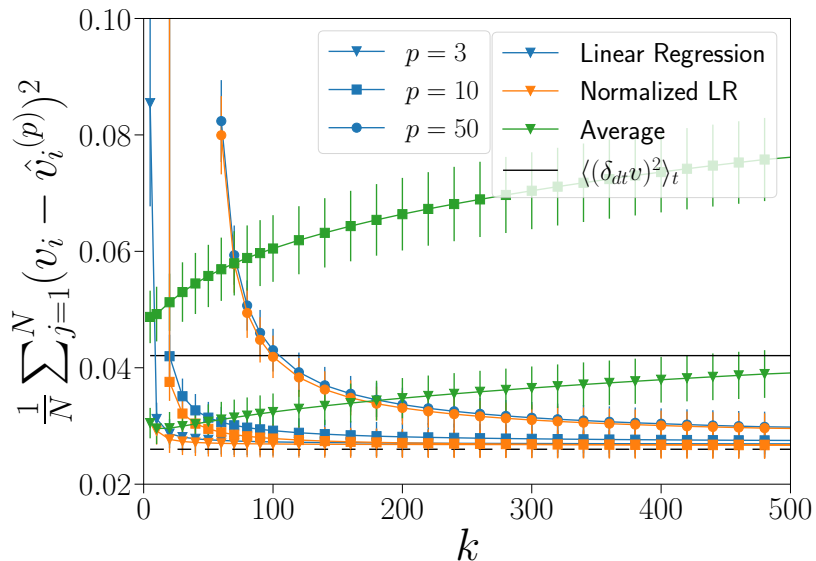


FIG. 2. Mean Squared Error $\frac{1}{N} \sum_{j=1}^N (v_j - \hat{v}_j^{(p)})^2$ as a function of the number k of neighbors used in the kNN algorithm. The performance is shown for different sizes of the past: $p = 3, 10, 50$ and for different methods: Linear Regression (blue), Normalized LR (orange), Averaging LR (green). Vertical lines correspond to the average value and errorbars to the standard deviation computed over n_{real} realizations. Increment variance (continuous black line) is present as a baseline for informative prediction. A dashed line is present to highlight the minimum value towards which the estimator converges.

B. Innovation Statistics

Figures 3a) and 3b) show respectively the typical evolution of the Modane velocity v and of its increment $\delta_{dt}v$ over multiple integral scales. The increment exhibits intermittent behavior with calm regions and bursty regions that highlight its non-Gaussianity. Figures 3c) and 3e) show respectively the innovation $\epsilon_t^{(p)}$ and its cumulative sum $\sum_{t'=0}^{t'=t} \epsilon_{t'}^{(p)}$. The innovation mirrors the behavior of the increment but with reduced variance. The concomitant occurrence of bursts in both the increment and the innovation time series suggests that extreme events in the increments induced by intermittency lead to an increase in the unpredictability of the velocity values v_t knowing their past $(v_{t-dt}, v_{t-2dt}, v_{t-3dt})$. This is coherent with previous results illustrating the challenge of predicting extreme events [30–32]. Figure 3 d) illustrates the variance of the innovation, highlighting its intermittent nature. The variance of the innovation shows spike patterns that match with bursts in the increment and innovations signals. Furthermore, we observed that $\mathbb{E}[\log(\sigma^2) | \log((\delta_{dt}v)^2)]$ increases with $\log((\delta_{dt}v)^2)$ in a non-linear way. This confirms that regions of high variance of innovation correspond to the presence of intermittent bursts. This simultaneous peaking underscores that unpredictability is tied with intermittency. The cumulative sum of the innovation is non-stationary and seems to behave like a random walk, which might indicate that the linear part of the dynamics is correctly predicted in $\hat{v}_t^{(p)}$ and therefore removed from the innovation $\epsilon_t^{(p)}$.

Figures 4a) and 4b) present the evolution of the second-order structure function and the flatness, with the time-scale τ of the increment, for the Modane turbulent velocity v_t (red), its increment $\delta_{dt}v$ (blue), the innovation $\epsilon_t^{(p)}$ (green), and its cumulative sum (yellow). We relate our results obtained in time to Kolmogorov-Obukhov theories in space using Taylor’s hypothesis [25, 26].

The logarithm of the second order structure function $\ln(S_2(\tau/T))$ of the increment $\delta_{dt}v$ and the innovation ϵ (blue and green curves respectively) behave very similarly with the time-scale τ . However, it is completely constant across scales for the innovation while it increases at small scales for the increment, until reaching a plateau.

This increase of the second order structure function S_2 of the increment at small scales reveals the existence of linear correlations within this process. In contrast, the flat behavior of S_2 for the innovation (in green) indicates that linear correlations have been eliminated and accurately predicted by the analog method.

For the velocity (in red), we recognize the three following distinct regions: the integral domain at larger scales, characterized by a plateau and the largest value of S_2 ; the inertial range with the $\tau^{2/3}$ scaling from K41 theory (eq.1) with $\zeta(2) = 2h$ and $h = 1/3$ up to intermittent corrections; and the dissipative domain at smaller scales with a $(\tau/T)^2$ scaling. For the cumulative sum of the innovation (in yellow), we observe two distinct regions with different power laws: one with an exponent close to 0.7 from the smallest scales up to $\tau \sim T$ and one with an exponent 1 at

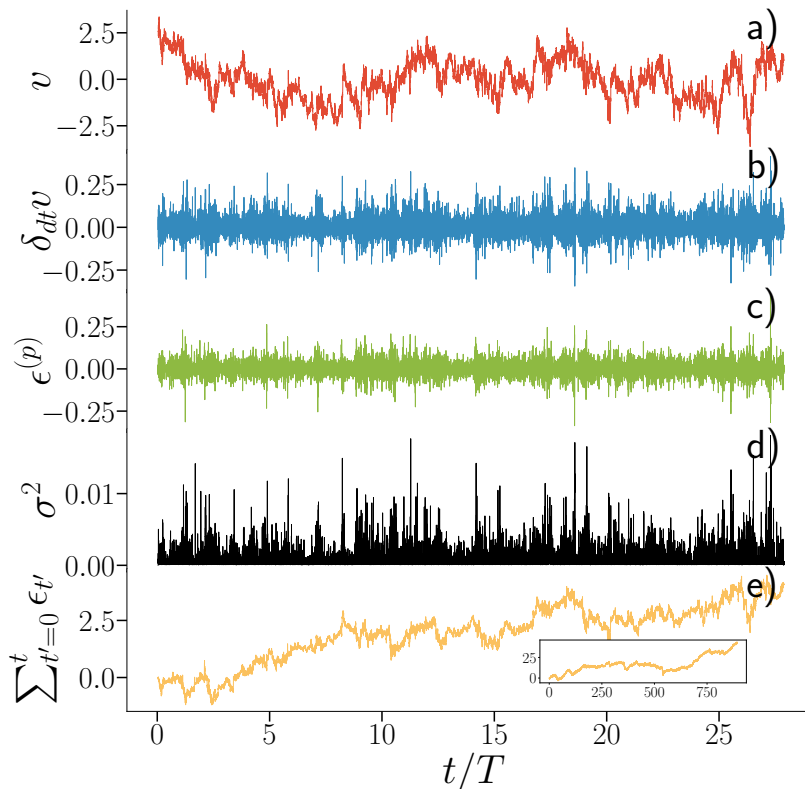


FIG. 3. Time series of the Modane velocity v (a), its smallest increment $\delta_{dt}v$ (b), the innovation ϵ (c), innovation variance σ^2 (d), and the cumulative sum of innovation $\sum_{t'=0}^t \epsilon_{t'}$ (e). The insert shows the cumulative sum of innovation over a longer duration.

larger scales, identical to a Brownian motion.

Finally, the logarithm of the flatness $\ln(F(\tau/T))$ is presented in fig.4b). The flatness of the increments $\delta_{dt}v$ (in blue) and of the innovation ϵ (in green) is always larger than one, indicating non-Gaussian behavior at all scales. The flatness of the innovation is larger because simple linear correlations have been predicted, which reduces S_2 . This implies that the unpredicted extreme events due to intermittency have a greater relative impact, resulting in increased flatness. For the velocity (in red) the flatness tends to 1 at large scales, indicating Gaussian behavior, while being substantially larger at smaller scales, revealing intermittency [42]. The cumulative sum of the innovation (in yellow) shows a power-law behavior for scales larger than $\ln(\tau/T) \sim -3$ with an exponent close to -0.1 . For smaller scales, the evolution of the flatness is faster. Such behavior indicates the persistence of higher-order correlations within the innovation signal, despite the removal of most, if not all, linear correlations.

V. CONCLUSION

In this study, we used analog-based predictions of turbulent velocity measurements from the Modane wind tunnel and computed the innovation in order to analyze the predictability of turbulent velocity. A first application confirmed that linear regression-based analog prediction better accounts for complex dynamics such as turbulence, compared to the standard averaging method, as suggested by state-of-the-art studies. Our analysis then illustrated that extreme events in the turbulent velocity gradient correspond to extreme events in the innovation, indicating a direct relationship between intermittency-related bursts of the velocity signal and harder to predict velocity values. These extreme events coincide also with high values of the uncertainty of the prediction. The correspondence between high values of the innovation and high values of the uncertainty of the prediction could be illustrative of the intrinsic stochasticity of turbulence. Thus in turbulence, predictions should better exploit the information of the pdf $q(v_t|v_{t-pdt:t-dt})$, not considering only its first order moment.

Moreover, a multi-scale analysis revealed that the analog-based estimator of turbulent velocity effectively predicts and removes most linear dependencies in the turbulent velocity signal, as evidenced by the flat second-order structure function of the corresponding innovation. The second-order structure function of the cumulative sum of the innovation

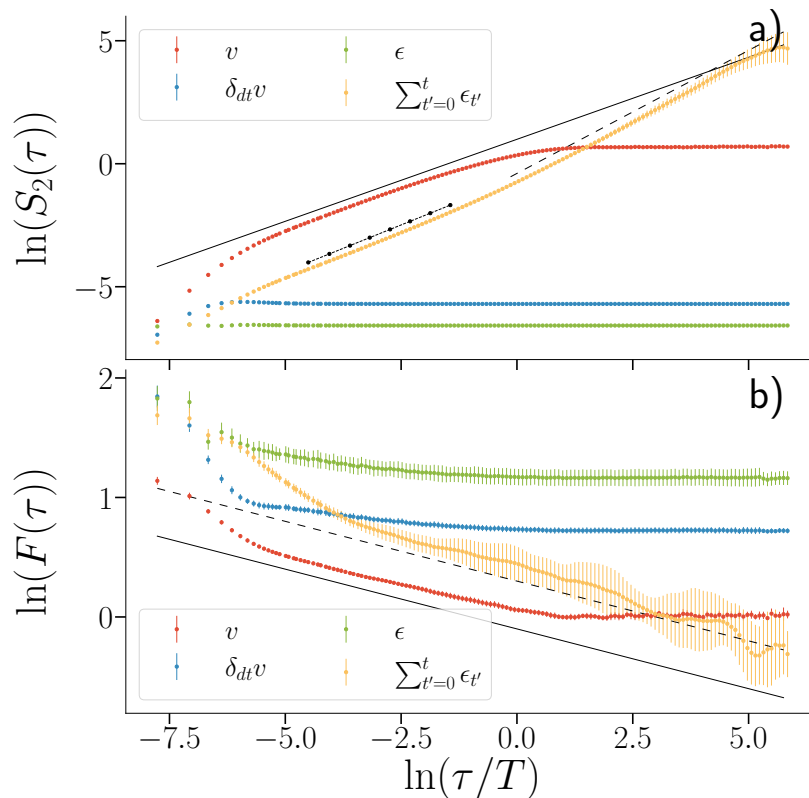


FIG. 4. Modane moments: a) second-order structure function $S_2(\bullet)$ and b) flatness $S_4(\bullet)/3 \cdot S_2(\bullet)$ for the velocity v (red), the associated smallest increment $\delta_{dt}v$ (blue), innovation ϵ (green), and innovation sum (yellow). Lines correspond to the average value and errorbars to the standard deviation computed over n_{real} realizations. Computation are done over 10 realizations, each of 892 integral scales. The black lines correspond to power laws: a) continuous: $(\tau/T)^{2/3}$, dashed: $(\tau/T)^1$, round markers: $(\tau/T)^{0.77}$ and b) continuous and dashed: $(\tau/T)^{-0.1}$ with different offsets.

does not exhibit the typical dissipative and integral domains. Instead, it shows two distinct regions with different power laws: one with a slope close to 0.7 and another with a slope 1. This suggests that while the estimator captures linear dynamics, it struggles with high-order dependencies. The flatness analysis further confirms that high-order dependencies remain largely intact, as indicated by the same power law of exponent -0.1 for the cumulative sum of the innovation as for the turbulent velocity itself. These findings highlight the limitations of the current analog-based estimator in fully capturing high-order interactions and extreme events. This is an illustration of the limitations of the hypotheses behind analog-based forecasting when applied to turbulence. Future research should focus on introducing more physical, e.g. multifractal, models in the selection of analogs to better predict high-order interactions and better capture the full complexity of turbulent flows.

ACKNOWLEDGMENTS

The code used for the estimation of analog-based predictions and innovations is available in : <https://gitlab.imt-atlantique.fr/e22froge/multi-scale-causality>

The authors wish to thank P. Tandeo for enriching discussions. This work was supported by the French National Research Agency (ANR-21-CE46-0011-01), within the program ‘‘Appel à projets générique 2021’’.

-
- [1] H. Cramér, On the linear prediction problem for certain stochastic processes, *Arkiv för Matematik* 4 (1) (1960) 45–53.
 [2] J. L. Doob, *Stochastic Processes*, Wiley, 1953.

- [3] J. Peinke, M. Klein, A. Kitte, A. Okninsky, J. Parisi, . E. Roessler, On chaos, fractals and turbulence, *Physica Scripta* T49 (1993) 672–676.
- [4] A. Crisanti, M. H. Jensen, A. Vulpiani, G. Paladin, Intermittency and predictability in turbulence, *Physical Review Letters* 70 (2) (1993) 166–169.
- [5] A. Crisanti, M. H. Jensen, G. Paladin, A. Vulpiani, Predictability of velocity and temperature fields in intermittent turbulence, *Journal of Physics A: Mathematical and General* 26 (23) (1993) 6943.
- [6] E. Aurell, G. Boffetta, A. Crisanti, G. Paladin, A. Vulpiani, Predictability in systems with many characteristic times: The case of turbulence, *Physical Review E* 53 (3) (1996) 2337–2349.
- [7] G. Boffetta, A. Celani, Predictability in chaotic systems and turbulence, *Le Journal de Physique IV* 08 (PR6) (1998) Pr6–146.
- [8] G. Boffetta, S. Musacchio, Chaos and Predictability of Homogeneous-Isotropic Turbulence, *Physical Review Letters* 119 (5) (2017) 054102.
- [9] U. Frisch, *Turbulence: The Legacy of A.N. Kolmogorov*, Cambridge University Press, 1995.
- [10] C. Meneveau, K. R. Sreenivasan, The multifractal nature of turbulent energy dissipation, *Journal of Fluid Mechanics* 224 (1991) 429–484.
- [11] E. N. Lorenz, Atmospheric Predictability as Revealed by Naturally Occurring Analogues, *Journal of the Atmospheric Sciences* 26 (4) (1969) 636–646.
- [12] D. Ruelle, Microscopic fluctuations and turbulence, *Physics Letters A* 72 (2) (1979) 81–82.
- [13] L. Biferale, A. A. Mailybaev, G. Parisi, Optimal subgrid scheme for shell models of turbulence, *Physical Review E* 95 (4) (2017) 043108.
- [14] D. Bandak, A. A. Mailybaev, G. L. Eyink, N. Goldenfeld, Spontaneous Stochasticity Amplifies Even Thermal Noise to the Largest Scales of Turbulence in a Few Eddy Turnover Times, *Physical Review Letters* 132 (10) (2024) 104002.
- [15] T. Li, L. Biferale, F. Bonaccorso, M. Buzzicotti, L. Centurioni, Stochastic Reconstruction of Gappy Lagrangian Turbulent Signals by Conditional Diffusion Models (Oct. 2024). [arXiv:2410.23971](https://arxiv.org/abs/2410.23971).
- [16] H. M. van den Dool, A New Look at Weather Forecasting through Analogues, *Monthly Weather Review* 117 (10) (1989) 2230–2247.
- [17] Z. Toth, Long-Range Weather Forecasting Using an Analog Approach, *Journal of Climate* 2 (6) (1989) 594–607.
- [18] G. Sugihara, R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature* 344 (6268) (1990) 734–741.
- [19] H. Kahalerras, Y. Malecot, Y. Gagne, B. Castaing, Intermittency and Reynolds number, *Physics of Fluids* 10 (1998) 910–921.
- [20] U. Frisch, G. Parisi, On the singularity structure of fully developed turbulence, *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* 01 (1985) 71–88.
- [21] G. Paladin, A. Vulpiani, Anomalous scaling laws in multifractal objects, *Physics Reports* 156 (4) (1987) 147–225.
- [22] A. N. Kolmogorov, The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, *Proceedings: Mathematical and Physical Sciences* 434 (1890) (1991) 9–13.
- [23] Y. Gagne, E. J. Hopfinger, U. Frisch, A New Universal Scaling for Fully Developed Turbulence: The Distribution of Velocity Increments, in: P. Coulet, P. Huerre (Eds.), *New Trends in Nonlinear Dynamics and Pattern-Forming Phenomena: The Geometry of Nonequilibrium*, Springer US, New York, NY, 1990, pp. 315–319.
- [24] B. Castaing, Y. Gagne, E. J. Hopfinger, Velocity probability density functions of high Reynolds number turbulence, *Physica D: Nonlinear Phenomena* 46 (2) (1990) 177–200.
- [25] A. N. Kolmogorov, A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number, *Journal of Fluid Mechanics* 13 (1962) 82–85.
- [26] A. M. Obukhov, Some specific features of atmospheric turbulence, *Journal of Fluid Mechanics* 13 (1962) 77–81.
- [27] L. Chevillard, B. Castaing, A. Arneodo, E. L  v  que, J. F. Pinton, S. G. Roux, A phenomenological theory of Eulerian and Lagrangian velocity fluctuations in turbulent flows, *Comptes Rendus Physique* 13 (9) (2012) 899–928.
- [28] B. Castaing, Y. Gagne, M. Marchand, Log-similarity for turbulent flows?, *Physica D: Nonlinear Phenomena* 68 (3) (1993) 387–400.
- [29] J. Delour, J. Muzy, A. Arn  odo, Intermittency of 1D velocity spatial profiles in turbulence: A magnitude cumulant analysis, *European Physical Journal B* 23 (2001) 243–248.
- [30] B. Dubrulle, Beyond Kolmogorov cascades, *Journal of Fluid Mechanics* 867 (2019) P1.
- [31] D. Buaria, A. Pumir, E. Bodenschatz, P. K. Yeung, Extreme velocity gradients in turbulent flows, *New Journal of Physics* 21 (4) (2019) 043004.
- [32] H. K. Moffatt, Extreme events in turbulent flow, *Journal of Fluid Mechanics* 914 (2021) F1.
- [33] A. Kolmogorov, The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds’ Numbers, *Akademiia Nauk SSSR Doklady* 30 (1941) 301–305.
- [34] C. Granero-Belinchon, S. G. Roux, N. B. Garnier, Scaling of information in turbulence, *EuroPhysics Letters* 115 (5) (2016) 58003.
- [35] H. Poincar  , Sur le probl  me des trois corps et les   quations de la dynamique, *Acta mathematica* 13 (1) (1890) A3–A270.
- [36] P. Platzer, P. Yiou, P. Naveau, P. Tandeo, Y. Zhen, P. Ailliot, J.-F. Filipot, Using local dynamics to explain analog forecasting of chaotic systems, *Journal of the Atmospheric Sciences* 78 (7) (2021) 2117–2133.
- [37] R. Lguensat, P. Tandeo, P. Ailliot, M. Pulido, R. Fablet, The Analog Data Assimilation, *Monthly Weather Review* 145 (2017) 4093–4107.
- [38] S. Kruizinga, A. H. Murphy, Use of an Analogue Procedure to Formulate Objective Probabilistic Temperature Forecasts

- in The Netherlands, *Monthly Weather Review* 111 (11) (1983) 2244–2254.
- [39] L. D. Monache, T. Nipen, Y. Liu, G. Roux, R. Stull, Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions, *Monthly Weather Review* 139 (11) (2011) 3554–3570.
- [40] T. P. Barnett, R. W. Preisendorfer, Multifield Analog Prediction of Short-Term Climate Fluctuations Using a Climate State vector, *Journal of the Atmospheric Sciences* 35 (10) (1978) 1771–1787.
- [41] J. Peinke, M. R. R. Tabar, M. Wächter, The Fokker–Planck Approach to Complex Spatiotemporal Disordered Systems, *Annual Review of Condensed Matter Physics* 10 (Volume 10, 2019) (2019) 107–132.
- [42] C. Granero-Belinchon, S. G. Roux, N. B. Garnier, Kullback-Leibler divergence measure of intermittency: Application to turbulence, *Physical Review E* 97 (2018) 013107.