FRACTALTS: TIME SERIES DATASET FOR CONTROLLED SCALING DYNAMICS AND MULTIFRACTAL PROCESSES

Anonymous authors
Paper under double-blind review

ABSTRACT

A major challenge in time-series machine learning (ML) is separating genuine learning of temporal structure from successful matching of low-order statistics. This problem is particularly important for fractional, fractal, and multifractal (MF) processes, whose defining properties are expressed across multiple scales. Two sequences may exhibit similar marginal distributions and short-range correlations while following different scaling laws. Existing datasets are predominantly collected from natural or engineered systems, where the generating mechanism and its scaling parameters are only partially known. Thus, model failures cannot be simply separated from uncertainty in the target properties. We introduce FRACTALTS, a theory-grounded benchmark dataset for controlled time-series research in which process families, scaling regimes, and stochastic parameters are explicitly specified at generation time. The core dataset contains independent realizations of monofractal fractional Brownian motion and six multifractal constructions spanning anti-correlated, uncorrelated, correlated, $1/f$, $1/f^2$, and non-stationary regimes. The generator exposes interpretable controls, including the Hurst exponent (H), multifractal intermittency, cascade parameter, spectral slope, envelope, and drift, and supports arbitrary numbers of independently seeded realizations. Rather than evaluating a particular downstream model, this paper establishes the dataset as a reproducible scientific testbed. A reproducible manifest and train/calibration/validation/test protocol is provided. Validation covers covariance, spectral scaling, Hurst behaviour, generalized scaling functions $h(q)$, multifractal spectra, and stationarity. The dataset is primarily designed for controlled evaluation of time-series generative models (GANs, VAEs, normalizing flows, and diffusion models), which are rarely tested on whether they preserve multiscale structure. It also supports representation learning, forecasting, anomaly detection, and estimator benchmarking under known multiscale dynamics.

1 INTRODUCTION

Time series are often evaluated through marginal distributions, short-range temporal dependence, forecasting accuracy, or downstream task performance. These criteria are useful but do not fully characterize processes whose identity is expressed through their behavior across temporal scales. Long-range dependence, self-similarity, intermittency, and multifractality can remain hidden when evaluation is restricted to a small set of moments or lags. A model may therefore reproduce local appearance while altering the scaling law that governs the process at larger or smaller scales.

A crucial problem is the lack of ground truth in real-world observations. For a measured signal, the generating mechanism is typically unknown. Even when a scaling exponent is estimated, finite-length estimation introduces uncertainty. Consequently, a benchmark based only on empirical observations cannot readily distinguish model failure from uncertainty in the target property. Controlled stochastic processes provide a complementary solution: their construction can expose the parameters of interest and allow independent realizations to be generated for training, calibration, and testing.

Classical stochastic-process theory provides several natural anchors for such a resource. Fractional Brownian motion (fBm) is a self-similar Gaussian process characterized by a single Hurst exponent and provides a canonical monofractal reference (Mandelbrot & Van Ness, 1968); it can be simulated exactly and efficiently through circulant embedding of the fractional-Gaussian-noise covariance

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(Davies & Harte, 1987; Dietrich & Newsam, 1997; Craigmile, 2003). Multifractal random walks (MRWs) introduce stochastic volatility and intermittency into the increment process while retaining an interpretable dependence structure (Bacry et al., 2001), extending the multiplicative-cascade picture of turbulence (Kolmogorov, 1941; Frisch & Parisi, 1985) and the thermodynamic formalism for singular measures (Halsey et al., 1986). Scaling structure is quantified using structure functions, detrended fluctuation analysis (DFA) and its multifractal extension (MFDFA), and wavelet-based estimators (Peng et al., 1994; Muzy et al., 1993; Kantelhardt et al., 2002; Jaffard et al., 2009; Abry & Veitch, 1998). These estimators are known to be sensitive to trends, finite length, and scale-range selection (Hu et al., 2001; Taqqu et al., 1995; Kantelhardt et al., 2001; Beran, 1994), which is precisely why a benchmark must treat their outputs as calibrated finite-sample quantities rather than exact constants. Surrogate-data analysis further provides a method to test whether apparent multifractality can be attributed to linear correlation or low-order effects (Theiler et al., 1992; Schreiber & Schmitz, 2000). Scaling laws of exactly this kind govern real signals in physiology (Peng et al., 1995; Ivanov et al., 1999), finance (Mandelbrot, 1997; Calvet & Fisher, 2002), geophysics and climate (Gagnon et al., 2006; Kantelhardt et al., 2006; Lovejoy & Schertzer, 2013), and network traffic (Leland et al., 1994), so a model that silently distorts them is failing on a property that matters in practice.

This gap is directly relevant to the rapidly growing literature on time-series generative models. Adversarial, variational, flow-based, and diffusion-based generators are now routinely used for synthesis, imputation, and probabilistic forecasting (Yoon et al., 2019; Esteban et al., 2017; Desai et al., 2021; Rasul et al., 2021; Tashiro et al., 2021; Yuan & Qiao, 2024; Crabbé et al., 2024; Naiman et al., 2024; Yang et al., 2026), and dedicated benchmarking libraries for them are beginning to appear (Ang et al., 2025). Yet these models are typically evaluated with marginal-distribution distances, discriminative "train-on-synthetic" scores, or two-sample tests (Heusel et al., 2017; Gretton et al., 2012), none of which directly certifies that long-range dependence, spectral color, or multifractal width has been preserved. A generator can therefore score well while quietly flattening the singularity spectrum or shifting the Hurst exponent. A controlled dataset with known scaling parameters turns this implicit failure mode into a measurable one.

Recent datasets published in ML venues demonstrate the value of carefully engineered scientific benchmarks. NoBOOM provides multiple real-world chemical-process datasets with detailed metadata and explicitly analyzes benchmark reliability issues (Wagner et al., 2025). TIDMAD couples a scientific time-series dataset with a domain-specific denoising setting (Fry et al., 2025), while Time-IMM targets irregular multimodal multivariate temporal data (Chang et al., 2025). Broad forecasting archives and classification collections (Godahewa et al., 2021; Bagnall et al., 2017) and recent fair-evaluation benchmarks (Qiu et al., 2024; Aksu et al., 2024) have likewise advanced the field, largely by improving coverage and curation of real data. These works establish a useful precedent: a dataset paper can make a substantive ML contribution when its data-generation or data-curation protocol addresses a clearly defined limitation of existing resources.

We introduce FRACTALTS, a controlled benchmark dataset designed around this principle. The benchmark does not claim that a synthetic process family is a substitute for real-world data. Instead, it supplies a complementary layer in which the target multiscale structure is known by construction and can be varied independently of dataset identity.

Our contributions are fourfold:

1. **A controlled dataset taxonomy.** We define seven core stochastic families spanning monofractal and multifractal regimes with anti-persistent, uncorrelated, persistent, colored-noise, Brownian/red-noise, and non-stationary dynamics.
2. **A parameterized and reproducible generator.** The release records process family, physical interpretation of each control parameter, random seed, sequence length, and generation configuration, allowing arbitrary numbers of independent realizations to be reproduced.
3. **A validation protocol for multiscale structure.** We specify independent tests of covariance, spectral scaling, Hurst behavior, generalized scaling, multifractal spectra, and stationarity, with calibration distributions estimated from held-out source realizations rather than treating finite-sample estimates as exact ground truth.
4. **A foundation for future model evaluation.** The resource is explicitly designed so that future generative, representation-learning, forecasting, and scientific-ML methods can be tested against controlled process-level ground truth.

Going forward, the paper first formalizes the benchmark design, then details the process families and generation pipeline, followed by validation, dataset organization, intended uses, limitations, and future work.

2 BENCHMARK DESIGN PRINCIPLES

We designed FRACTALTS around five principles.

1. **Mechanistic control.** Each core family is associated with a transparent stochastic construction. The objective is not merely to produce visually different signals, but to make the intended source of temporal structure explicit.
2. **Orthogonal structural variation.** Where possible, parameters controlling dependence, intermittency, spectral colour, and non-stationarity are varied independently. This enables controlled questions such as whether a method can distinguish $H < 0.5$ from $H > 0.5$, or whether it preserves multifractal breadth as intermittency changes.
3. **Independent calibration.** Scaling estimators are not exact at finite sequence lengths. We therefore separate nominal process parameters from empirically estimated finite-sample descriptors. A calibration split consists of independent source realizations not used to fit downstream models.
4. **Reproducibility by construction.** Every sequence is generated from a deterministic configuration and an explicit seed. A machine-readable manifest associates each sample with its family, parameters, length, split, and seed.
5. **Extensibility without taxonomy collapse.** The core benchmark focuses on fractal and multifractal stochastic structure. Physical noise mechanisms such as thermal, shot, partition, transit-time, atmospheric, solar, or cosmic noise can be incorporated later, but they are kept separate from the core taxonomy until their generators and structural properties have been independently validated.

3 DATASET COMPOSITION

The current generator contains seven registered core families, of which six have dedicated per-family reference generators released as documented notebooks; the non-stationary family is provided through the unified class-based generator. The reference release configuration contains 1,000 independent realizations per family, with sequence length $T = 4096$, yielding 7,000 core sequences. Each per-family generator accepts arbitrary n and T, so larger releases can be produced without changing the data model. The seven families are deliberately related rather than heterogeneous. This makes the benchmark suitable for controlled interpolation and extrapolation experiments: two datasets may differ by one structural parameter while sharing the rest of the construction.

3.1 MONOFRACTAL REFERENCE: FRACTIONAL BROWNIAN MOTION

Let $B_H(t)$ denote fractional Brownian motion with Hurst exponent $H \in (0, 1)$. Its increment process, fractional Gaussian noise (fGn), has covariance

$$\gamma_H(k) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right). \tag{1}$$

The reference generator draws fGn (or, optionally, the integrated fBm path) using the exact Davies–Harte circulant-embedding method as implemented in the `fbm` library (Flynn, 2020; Davies & Harte, 1987; Dietrich & Newsam, 1997), which embeds the Toeplitz covariance of Eqn. 1 in a circulant matrix, diagonalizes it by FFT, and draws samples with the exact second-order structure. Because fGn is Gaussian and therefore fully specified by its covariance, this construction is monofractal by design. To certify this at finite length, the generator wraps each draw in a rejection loop that recomputes the MFDFA singularity spectrum width ($\Delta\alpha$) and accepts the realization only if $\Delta\alpha < 0.05$; realizations with spurious finite-sample width are discarded and redrawn. The released monofractal set therefore consists exclusively of realizations that pass an explicit near-zero-width test, rather than relying on the theoretical guarantee alone.

Table 1: Core FRACTALTS families and their reference generators. "Acceptance criterion" is the per-realization structural test enforced by rejection sampling at generation time (Section 4): every released sequence is regenerated until it satisfies the stated bounds on the estimated Hurst exponent H (at q=2) and $\Delta\alpha$. Here λ^2 and λ denote the log-normal-multiplier variance and the spectral-cascade exponent, respectively; the notation follows each family's source notebook.

Family	Construction	Primary controls	Acceptance criterion
Monofractal	Exact fGn via Davies–Harte (`fbm` library)	H	$\Delta\alpha < 0.05$
MF anti-correlated	Fast Fourier Transform (FFT)-synthesized fGn (H<0.5) × log-normal multipliers	H, λ^2	$0<H<0.5$ and $\Delta\alpha>0.25$
MF uncorrelated	White noise × binomial (p) cascade	p	$0.48<H<0.52$ and $\Delta\alpha>0.7$
MF correlated	Spectral fGn (H>0.5) × log-normal spectral cascade	H, λ	$0.5<H<1$ and $\Delta\alpha>0.25$
MF $1/f$ (pink)	Spectral synthesis with amplitude $f^{-(H-1/2)}$, H=1	H	$0.99<H<1.01$ and $\Delta\alpha>0.25$
MF $1/f^2$ (Brownian)	Cumulative sum of Gaussian white noise	n/a	$1.49<H<1.51$ and $\Delta\alpha>0.25$

3.2 MULTIFRACTAL FAMILIES WITH CORRELATED BASE AND MULTIPLICATIVE MODULATION

Three families impose multifractal intermittency on a correlated Gaussian base by multiplication, but differ in how the base and the modulation are constructed. In each case the multiplication in the time domain couples an interpretable linear-dependence structure to a scale-dependent volatility, and the released realizations are those that pass the family's structural acceptance test.

Anti-correlated ($H < 0.5$). The base is fractional Gaussian noise synthesized by FFT from the covariance of Eqn. 1 with $H < 0.5$, giving anti-persistent increments. Multifractality is imprinted by pointwise multiplication with i.i.d. log-normal multipliers of variance λ^2 (default $\lambda^2 = 0.02$), so that

$$X(t) = \mathrm{fGn}_H(t) \cdot \exp\big(\mathcal{N}(-\lambda^2/2, \lambda^2)\big). \tag{2}$$

Acceptance requires $0 < \widehat{H} < 0.5$ and $\widehat{\Delta\alpha} > 0.25$.

Correlated ($H > 0.5$). The base is a persistent correlated Gaussian noise synthesized in the frequency domain with amplitude profile $i^{-(H-1/2)}$, and the modulation is a log-normal cascade synthesized in the frequency domain with amplitude profile $i^{-\lambda}$ and then exponentiated and mean-normalized. A smaller spectral-cascade exponent λ (default $\lambda = 0.01$) widens the singularity spectrum. Acceptance requires $0.5 < \widehat{H} < 1$ and $\widehat{\Delta\alpha} > 0.25$.

Uncorrelated (white-noise base). The base is i.i.d. Gaussian white noise ($H \approx 0.5$), and multifractality is introduced by a deterministic binomial (multiplicative) cascade with weight parameter p built by repeated Kronecker products, $w \leftarrow w \otimes [p,\, 1-p] \times 2$ over $\log_2 T$ levels. The product is standardized to zero mean and unit variance. Acceptance requires $0.48 < \widehat{H} < 0.52$ and $\widehat{\Delta\alpha} > 0.7$, i.e. a genuinely broad spectrum on top of linearly uncorrelated increments.

3.3 MULTIFRACTAL COLORED-SPECTRUM CONSTRUCTIONS

Two families target a prescribed spectral color directly. The *pink* ($1/f$) family uses spectral synthesis: for a target exponent $H = 1$ it assigns Fourier amplitudes $f^{-(H-1/2)}$ with random phases, imposes Hermitian symmetry for a real signal, and inverse-transforms; the realization is standardized. Acceptance requires $0.99 < \widehat{H} < 1.01$ and $\widehat{\Delta\alpha} > 0.25$. The *Brownian* ($1/f^2$) family forms the cumulative sum of Gaussian white noise (Wiener path) and standardizes it; acceptance requires $1.49 < \widehat{H} < 1.51$ and $\widehat{\Delta\alpha} > 0.25$. These are best regarded as *controlled constructions with certified descriptors* rather than closed-form finite-sample multifractals: the multifractal breadth of the released realizations is guaranteed operationally by the acceptance test, not by an idealized infinite-scale formula.

Algorithm 1 Reference dataset generation with structural acceptance

Require: family c, number of instances N, length T, parameters θ, acceptance criterion $\mathcal{A}_c$, master seed s

Initialize seeded random generator $G \leftarrow \mathrm{RNG}(s)$
Resolve family constructor F_c and validated defaults θ_c; set $\theta^\star \leftarrow \theta_c \cup \theta$
for $i = 1, \ldots, N$ **do**
  repeat
    $x \leftarrow F_c(T; \theta^\star, G)$ {draw a candidate}
    $(\widehat{H}, \widehat{\Delta\alpha}) \leftarrow \mathrm{MFDFA}(x;\ \text{scales 10–100},\ q \in [-5, 5] \setminus \{0\},\ \text{order } 1)$
  until $\mathcal{A}_c(\widehat{H}, \widehat{\Delta\alpha})$ is satisfied
  $x_i \leftarrow x$; store x_i and metadata $(c, i, T, \theta^\star, s_i, \widehat{H}, \widehat{\Delta\alpha})$
end for
return $\{x_i\}_{i=1}^N$ and manifest =0

3.4 Non-stationary multifractal construction

The non-stationary family, provided through the unified class-based generator, combines a multifractal random walk with a time-varying variance envelope and a deterministic trend:

$$Y(t) = X(t)\, E(t) + D(t), \tag{3}$$

where $E(t)$ is a smooth periodic envelope and $D(t)$ is an increasing drift. This family is designed to distinguish multiscale structure from stationarity. It is intentionally challenging because global estimators can conflate multifractal scaling with local variance changes or deterministic trends.

4 Generation Pipeline

Each family is produced by a dedicated reference generator (released as a documented notebook for the six core families, and through the unified class-based generator for the non-stationary family). All generators share a common structure: a seeded random-number generator, a family-specific construction, and a *rejection-sampling acceptance loop* that certifies every released realization against explicit structural criteria. For the Gaussian building blocks we adopt exact spectral simulation (Davies & Harte, 1987; Wood & Chan, 1994) in preference to faster approximate synthesis schemes (Paxson, 1997), so that the released monofractal reference matches its analytical covariance to numerical precision rather than only in the large-sample limit.

In the *"SUPPLEMENTARY MATERIALS"* section, Figure 1 shows one representative realization from each of the seven core families, drawn with a fixed seed at the reference length $T = 4096$. Table 3 reports empirical structural descriptors for the six per-family reference generators, measured directly on the raw construction (before the rejection-sampling acceptance loop) over $N = 100$ independent realizations per family at $T = 4096$.

The acceptance loop is the distinguishing feature of the pipeline. After each candidate realization is constructed, the generator runs MFDFA (Gorjão et al., 2022) on it, estimates the Hurst exponent $\widehat{H}$ at $q = 2$ and the singularity spectrum width $\widehat{\Delta\alpha}$ via the Legendre transform, and accepts the realization only if it satisfies the family's criterion (Table 1); otherwise it is discarded and a new candidate is drawn. The MFDFA configuration used for acceptance is fixed across all families: scales $s \in \{10, \ldots, 100\}$ chosen as $\mathrm{unique}(\lfloor \mathrm{logspace}(1, 2, 20) \rfloor)$, moment orders $q \in \{-5, \ldots, 5\}$ on a 41-point grid excluding $q = 0$, and first-order polynomial detrending. Consequently, the released dataset is not merely generated but *certified*: every sequence provably meets a stated bound on its measured scaling descriptors at the benchmark's own estimation settings.

The per-family notebooks export each realization as an individual MATLAB `.mat` file (keyed `data`) under a family-named directory, and the unified generator additionally supports compressed NumPy archives and consolidated `.mat` export. The release includes a canonical manifest in CSV or JSONL format. At minimum, each record contains:

```
sample_id, family, split, seed, length, H, intermittency, cascade_p,
                spectral_H, H_measured, dalpha_measured
```

with null values for parameters that do not apply to a given family. The two measured fields record the MFDFA descriptors that each realization satisfied at acceptance time, so downstream users can filter or stratify by certified structural properties.

4.1 Reference splits

For the default benchmark release we recommend four independent pools: training, validation, calibration, and test. A nominal configuration is 1,000 / 200 / 1,000 / 1,000 realizations per family, respectively, when sufficient computation and storage are available. The 1,000-realization public corpus can be used as a compact release, while the generator can regenerate larger split-specific pools. Crucially, no split should be created by slicing a single long realization into many samples; independent realization-level splitting is required to prevent leakage of shared stochastic trajectories.

4.2 Parameter grid

The reference per-family generators use the following defaults, each paired with its acceptance criterion (Table 1): monofractal $H = 0.75$; anti-correlated $H = 0.25$, $\lambda^2 = 0.02$; uncorrelated $p = 0.7$; correlated $H = 0.75$, $\lambda = 0.01$; pink $H = 1.0$; Brownian (implicit $H \approx 1.5$). Structural validation and any leaderboard evaluation use MFDFA with scales $s \in \{10, \ldots, 100\}$, the 41-point q-grid on $[-5, 5] \setminus \{0\}$, and order-1 detrending, matching the acceptance loop. Parameter sweeps (for example varying H across families, λ^2 or λ for intermittency, and p for cascade heterogeneity) can enlarge the grid; parameter-out-of-distribution (OOD) subsets hold out at least one interior or boundary value from the training grid while retaining the complete manifest.

5 Structural Validation

The core contribution of FractalTS is not simply that the code can produce diverse curves; it is that the curves demonstrably exhibit the properties claimed by their labels. This is enforced twice: once at generation time, where every realization must pass the family's MFDFA acceptance criterion before it enters the dataset (Section 4), and again at release time through the aggregate validation described here. Validation is performed at three levels: second-order statistics, multiscale scaling, and stationarity/mechanism controls. The primary multiscale estimator throughout is MFDFA (Kantelhardt et al., 2002; Gorjão et al., 2022) with the fixed configuration used for acceptance (scales 10–100, moments $q \in [-5, 5] \setminus \{0\}$ on a 41-point grid, order-1 detrending, Legendre transform for the singularity spectrum), so that reported descriptors are directly comparable to the acceptance bounds.

5.1 Second-order and spectral validation

For each family and parameter condition, we estimate the empirical mean, variance, lagged autocorrelation, and one-sided power spectral density (PSD) as secondary descriptors. Families with an explicit spectral target (pink, $S(f) \propto f^{-1}$; Brownian, $S(f) \propto f^{-2}$) additionally report a fitted log-PSD slope $\widehat{\beta}$ over a fixed frequency band with the DC component excluded. The primary structural descriptors, however, are the MFDFA quantities $\widehat{H}$ (at q=2) and $\widehat{\Delta\alpha}$, because these are the quantities the acceptance loop certifies. Reports include confidence intervals over independent realizations rather than a single example trace.

For fGn, empirical covariance is compared to Eqn. 1. For a Gaussian process with the intended covariance, both the covariance curve and its finite-sample uncertainty should match the reference within tolerance. This check is particularly important for FFT-based simulation because an incorrect normalization can preserve correlation shape while changing variance by orders of magnitude.

5.2 Hurst and generalized scaling validation

The benchmark reports estimates of the Hurst exponent H and, for multifractal families, generalized scaling $h(q)$ over a fixed set of moments q. For the monofractal reference, the expected structural condition is a nearly flat $h(q)$ curve. For multifractal families, the condition is non-zero scale

dependence in $h(q)$ together with a measurable singularity spectrum width after calibration for finite-length effects.

Let $\widehat{h}_i(q)$ be the estimate for sequence i. We report

$$\overline{h}(q) = \frac{1}{N}\sum_{i=1}^{N}\widehat{h}_i(q), \qquad \mathrm{CI}_{95}(q) = \mathrm{bootstrap}_{95\%}\{\widehat{h}_i(q)\} \tag{4}$$

For a scalar summary we use the curve deviation

$$E_h = \frac{1}{|\mathcal{Q}|}\sum_{q\in\mathcal{Q}}\left|\overline{h}(q) - h_{\mathrm{cal}}(q)\right|, \tag{5}$$

where h_{cal} is estimated from an independent calibration pool and intervals are obtained by nonparametric bootstrap over realizations (Efron, 1979). This is an evaluation quantity for the dataset's future benchmark use, not a generative-model result.

5.3 Multifractal spectrum validation

We estimate the singularity spectrum $f(\alpha)$ using MFDFA and, where available, an independent wavelet-leader estimator. The primary release records $\alpha_{\min}$, $\alpha_{\max}$, and $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$, but the complete curve is also retained. Using two estimators provides an important robustness check because multifractal estimates can depend on detrending, scale selection, and boundary handling.

A shuffled-source control is generated for selected sequences. Shuffling destroys temporal ordering while approximately preserving the one-point marginal distribution. If broad multifractal structure survives shuffling unchanged, the benchmark should not attribute that width solely to long-range temporal organization. Surrogate-data analysis motivates this control (Schreiber & Schmitz, 2000).

5.4 Stationarity validation

The non-stationary family is evaluated by rolling mean and variance, local fluctuation scale, and explicit stationarity diagnostics. The benchmark stores these descriptors as sequence-level metadata. The purpose is not to assign a binary label using a single hypothesis test; rather, it is to quantify the degree and form of non-stationarity so that future methods can be evaluated against an interpretable target.

5.5 Finite-sample calibration

An important methodological choice is to separate analytical construction parameters from estimated sample statistics. Let $\phi(x)$ be a structural descriptor such as $\widehat{H}$, $\widehat{\beta}$, or $\widehat{\Delta\alpha}$. Even if the generating parameter θ is fixed, $\phi(x)$ has a finite-sample distribution $P_{\phi|\theta,T}$. We estimate this distribution using independent calibration realizations and store its mean, standard deviation, median, and central confidence interval.

This turns a benchmark assertion from "the estimate must equal the nominal parameter" into the weaker and more defensible statement "the estimate should be consistent with the calibrated finite-sample distribution." This distinction is essential for a scientific dataset and also makes the resource more useful to downstream researchers who may use different estimators.

6 Dataset Quality Gates

Before a public release, every family/parameter condition should pass a fixed set of quality gates. We recommend the following order.

1. **Numerical validity:** finite values, no unexpected NaNs or infinities, and deterministic regeneration under repeated seeds.
2. **Amplitude validity:** empirical variance and range are within documented tolerances, especially for FFT-based constructions.

3. **Correlation validity:** empirical auto-correlation function/covariance agrees with the intended dependence regime.
4. **Spectral validity:** estimated PSD slope agrees with the declared colored-noise target over the preregistered frequency band.
5. **Scaling validity:** fBm is approximately monofractal, while multifractal families show calibrated non-trivial scale dependence.
6. **Mechanism controls:** shuffled/surrogate controls behave consistently with the intended source of the measured multiscale structure.
7. **Cross-seed reproducibility:** aggregate statistics remain stable when the master seed changes.

No family should be described as "exact" solely from implementation comments. In particular, FFT normalization, circulant-embedding conventions, endpoint handling, and scale-band selection must be verified empirically before final data release.

7 Baseline Scientific Uses

Although the present paper deliberately does not benchmark generative models, the dataset supports several immediate scientific tasks.

1. **Process-family identification.** Given an unlabeled sequence, a classifier can be tested on whether it distinguishes monofractal, multifractal, colored-noise, and non-stationary regimes.
2. **Scaling-parameter estimation.** Regression methods can estimate H, intermittency, or spectral slope from observed sequences. Because the generating parameters are known, both point accuracy and calibration can be measured.
3. **Representation learning.** The parameterized family structure allows representation methods to be evaluated on whether latent geometry reflects continuous changes in H, intermittency, spectral slope, or stationarity.
4. **OOD structural recognition.** Held-out parameter values provide a controlled test of interpolation and extrapolation in process space. This is different from random train/test splitting because the test distribution is defined by an unseen structural condition.
5. **Methodological validation of estimators.** Researchers developing fractal or multifractal estimators can compare bias, variance, robustness, and computational cost under known process families.

These uses are deliberately model-agnostic. In future work, the same resource can be used to evaluate whether synthetic-data generators preserve these properties.

8 Comparison to Existing Dataset Papers

A useful benchmark dataset should offer something that is difficult to obtain from existing real-world collections. NoBOOM demonstrates the value of publicly releasing rich, carefully analyzed time-series data when industrial measurements are normally inaccessible (Wagner et al., 2025). TIDMAD illustrates a complementary scientific-resource design in which the dataset is connected to an explicit physical denoising problem (Fry et al., 2025). Time-IMM addresses another limitation of conventional collections by explicitly representing irregular multimodal multivariate observations (Chang et al., 2025).

FractalTS differs primarily in the source of ground truth. Rather than relying on metadata attached to naturally occurring observations, it constructs the data from controlled stochastic mechanisms. Its contribution is therefore the ability to ask causal-mechanistic questions about time-series structure while retaining the scale and reproducibility of a machine-learning benchmark.

The design also follows a broader lesson from recent benchmark work: dataset quality should be evaluated explicitly rather than assumed from the existence of many samples. A benchmark should document its generation process, leakage risks, source assumptions, statistical pathologies, and limitations, in the spirit of structured dataset documentation (Gebru et al., 2021). FractalTS consequently treats validation and calibration as part of the dataset itself. It is also complementary to emerging generative-model benchmarking libraries (Ang et al., 2025): those efforts standardize

how generators are compared, whereas FRACTALTS supplies inputs whose multiscale ground truth is known exactly, so that any such comparison can additionally certify preservation of scaling structure.

9 LIMITATIONS

1. **Controlled processes are not real-world data.** The seven families span a useful subset of temporal phenomena but cannot represent the semantic, physical, or multivariate complexity of domains such as medicine, climate, finance, or industrial control.
2. **Finite-length scaling estimates are uncertain.** Even for analytically defined processes, estimated Hurst and multifractal descriptors depend on sequence length, estimator, scale range, and preprocessing. This is why the benchmark stores empirical calibration distributions rather than treating nominal parameters as exact observed estimates.
3. **Not every family has a closed-form finite-sample spectrum.** The spectral-synthesis and cascade-based multifractal families are intentionally constructive and should not be overclaimed as exact realizations of an idealized infinite-scale multifractal process; their multiscale properties are guaranteed operationally, through the acceptance criteria, rather than analytically.
4. **Rejection sampling induces a selection.** Because each family retains only realizations whose measured descriptors fall within an acceptance window, the released ensemble is a conditioned subset of the raw construction rather than an unconditional sample from it. This is deliberate (it is what certifies the labels), but users should treat the released marginal over $\widehat{\Delta\alpha}$ as truncated by design, and should regenerate with the published criteria rather than assuming an unconditional generative distribution.
5. **Parameter identifiability may be imperfect.** Different mechanisms can produce similar low-order statistics. This is not a flaw to hide; it is a useful challenge for methods that claim to identify scaling mechanisms.
6. **Core scope is intentionally narrow.** Physical noise categories such as thermal, shot, partition, burst, transit-time, atmospheric, solar, cosmic, or quantum noise are candidates for a future extension, but mixing them into the core benchmark without independent physical and statistical validation would weaken the taxonomy.

10 FUTURE WORK

The principal next step is to use FRACTALTS as a controlled evaluation substrate for time-series generative models. Future studies can ask whether GANs (Goodfellow et al., 2014; Yoon et al., 2019), variational autoencoders (Kingma & Welling, 2014; Desai et al., 2021), normalizing flows (Rezende & Mohamed, 2015), diffusion models (Ho et al., 2020; Yuan & Qiao, 2024; Crabbé et al., 2024), or classical statistical baselines preserve not only marginal and local temporal statistics but also Hurst behavior, generalized scaling $h(q)$, multifractal spectra, spectral slopes, and stationarity profiles. Because the generating parameters are known, such studies can move beyond distributional realism to *parameter-level fidelity*: measuring, for example, the bias a diffusion model introduces into the recovered Hurst exponent or the fraction of multifractal width it discards.

A second direction is conditional generation in which process parameters such as H and intermittency are supplied as explicit conditions. This would permit parameter recovery tests and controlled interpolation within the generative model's conditioning space. A third direction is multivariate extension, where cross-channel dependence and joint multifractality become additional benchmark dimensions. A fourth is a physically grounded noise extension with independently documented generators for thermal, shot, partition, transit-time, burst, atmospheric, solar, cosmic, and quantum noise. Such additions will be included only after their numerical and structural properties are validated using the same quality gates defined in Section 6.

ETHICS STATEMENT

The core dataset is entirely synthetic and does not contain human-subject data, personal information, or proprietary measurements. The principal ethical consideration is scientific reliability: labels should not imply stronger mathematical guarantees than have been validated empirically. We therefore

distinguish analytical design targets from finite-sample estimates and recommend full disclosure of failed generation or validation cases.

AI USE STATEMENT

AI-assisted tools were used during manuscript drafting and language editing and for literature review (e.g., finding related work). The authors remain responsible for the technical content, verification of mathematical statements, experimental validation, and final submission.

11 REPRODUCIBILITY STATEMENT

The benchmark is designed to be regenerated from code, parameters, and random seeds. The reference generators depend on the standard scientific-Python stack (Harris et al., 2020; Virtanen et al., 2020) together with the `fbm` (Flynn, 2020) and `MFDFA` (Gorjão et al., 2022) libraries, and the supplementary release contains the six per-family generator notebooks, the unified class-based generator, parameter manifests, an exact environment specification, validation scripts, and checksums for released data files. FRACTALTS is anonymously available from `https://shorturl.at/LWfja`. The paper's protocol defines all primary structural descriptors, the MFDFA configuration, and the splitting rules. This separation of generation and evaluation artifacts allows an independent researcher to reproduce the dataset without access to hidden state or private data.

ACKNOWLEDGMENTS

Omitted for double-blind review.

REFERENCES

Patrice Abry and Darryl Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, 1998. doi: 10.1109/18.650984.

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. GIFT-Eval: A benchmark for general time series forecasting model evaluation. *arXiv preprint arXiv:2410.10393*, 2024.

Yihao Ang, Qiang Huang, Yifan Bao, Anthony K. H. Tung, and Zhenjie Huang. GenTS: A comprehensive benchmark library for generative time series models. *arXiv preprint arXiv:2605.17804*, 2025.

Emmanuel Bacry, Julien Delour, and Jean-François Muzy. Multifractal random walk. *Physical Review E*, 64(2):026103, 2001. doi: 10.1103/PhysRevE.64.026103.

Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660, 2017. doi: 10.1007/s10618-016-0483-9.

Jan Beran. *Statistics for Long-Memory Processes*. Chapman and Hall/CRC, 1994. ISBN 9780412049019.

Laurent E. Calvet and Adlai J. Fisher. Multifractality in asset returns: Theory and evidence. *Review of Economics and Statistics*, 84(3):381–406, 2002. doi: 10.1162/003465302320259420.

Ching Chang, Jeehyun Hwang, Yidan Shi, Haixin Wang, Wei Wang, Wen-Chih Peng, and Tien-Fu Chen. Time-IMM: A dataset and benchmark for irregular multimodal multivariate time series. In *Advances in Neural Information Processing Systems 38, Datasets and Benchmarks Track*, 2025.

Jonathan Crabbé, Nicolas Huynh, Jan Stanczuk, and Mihaela van der Schaar. Time series diffusion in the frequency domain. In *International Conference on Machine Learning*, 2024.

Peter F. Craigmile. Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24(5): 505–511, 2003. doi: 10.1111/1467-9892.00318.

Robert B. Davies and David S. Harte. Tests for Hurst effect. *Biometrika*, 74(1):95–101, 1987. doi: 10.1093/biomet/74.1.95.

Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. TimeVAE: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*, 2021.

C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4): 1088–1107, 1997. doi: 10.1137/S1064827592240555.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26, 1979. doi: 10.1214/aos/1176344552.

Cristóbal Esteban, Stephanie L. Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional GANs. *arXiv preprint arXiv:1706.02633*, 2017.

Christopher Flynn. fbm: Exact methods for simulating fractional Brownian motion and fractional Gaussian noise. Python package, `https://pypi.org/project/fbm/`, 2020. Implements the Davies–Harte, Cholesky, and Hosking methods.

Uriel Frisch and Giorgio Parisi. Fully developed turbulence and intermittency. In *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, pp. 84–88. North-Holland, 1985.

Jessica Fry, Xinyi Fu, Zhenghao Fu, Kaliroe Pappas, Lindley Winslow, and Aobo Li. TIDMAD: Time series dataset for discovering dark matter with AI denoising. In *Advances in Neural Information Processing Systems 38, Datasets and Benchmarks Track*, 2025.

Jean-Sébastien Gagnon, Shaun Lovejoy, and Daniel Schertzer. Multifractal earth topography. *Nonlinear Processes in Geophysics*, 13(5):541–570, 2006. doi: 10.5194/npg-13-541-2006.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92, 2021. doi: 10.1145/3458723.

Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I. Webb, Rob J. Hyndman, and Pablo Montero-Manso. Monash time series forecasting archive. In *Advances in Neural Information Processing Systems 34, Datasets and Benchmarks Track*, 2021.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, 2014.

Leonardo Rydin Gorjão, Galib Hassan, Jürgen Kurths, and Dirk Witthaut. MFDFA: Efficient multifractal detrended fluctuation analysis in Python. *Computer Physics Communications*, 273: 108254, 2022. doi: 10.1016/j.cpc.2021.108254.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Thomas C. Halsey, Mogens H. Jensen, Leo P. Kadanoff, Itamar Procaccia, and Boris I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Physical Review A*, 33(2):1141–1151, 1986. doi: 10.1103/PhysRevA.33.1141.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems 30*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33*, 2020.

Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. Effect of trends on detrended fluctuation analysis. *Physical Review E*, 64(1):011114, 2001. doi: 10.1103/PhysRevE.64.011114.

Plamen Ch. Ivanov, Luís A. Nunes Amaral, Ary L. Goldberger, Shlomo Havlin, Michael G. Rosenblum, Zbigniew R. Struzik, and H. Eugene Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399(6735):461–465, 1999. doi: 10.1038/20924.

Stéphane Jaffard, Bruno Lashermes, and Patrice Abry. Wavelet leaders: A new method to estimate the multifractal singularity spectra. *Physica A: Statistical Mechanics and its Applications*, 388(14):2793–2805, 2009. doi: 10.1016/j.physa.2009.03.043.

Jan W. Kantelhardt, Eva Koscielny-Bunde, Henio H. A. Rego, Shlomo Havlin, and Armin Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 295(3–4):441–454, 2001. doi: 10.1016/S0378-4371(01)00144-3.

Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1–4):87–114, 2002. doi: 10.1016/S0378-4371(02)01383-3.

Jan W. Kantelhardt, Eva Koscielny-Bunde, Diego Rybski, Peter Braun, Armin Bunde, and Shlomo Havlin. Long-term persistence and multifractality of precipitation and river runoff records. *Journal of Geophysical Research: Atmospheres*, 111(D1):D01106, 2006. doi: 10.1029/2005JD005881.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014.

Andrey N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akademii Nauk SSSR*, 30:301–305, 1941.

Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994. doi: 10.1109/90.282603.

Shaun Lovejoy and Daniel Schertzer. *The Weather and Climate: Emergent Laws and Multifractal Cascades*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139093811.

Benoit B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, 1997. doi: 10.1007/978-1-4757-2763-0.

Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968. doi: 10.1137/1010093.

Jean-François Muzy, Emmanuel Bacry, and Alain Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E*, 47(2):875–884, 1993. doi: 10.1103/PhysRevE.47.875.

Ilan Naiman, Nimrod Berman, Itai Pemper, Idan Arbiv, Gal Fadlon, and Omri Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. In *Advances in Neural Information Processing Systems 37*, 2024.

Vern Paxson. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*, 27(5):5–18, 1997. doi: 10.1145/269790.269792.

Chung-Kang Peng, Sergey V. Buldyrev, Shlomo Havlin, Michael Simons, H. Eugene Stanley, and Ary L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2):1685–1689, 1994. doi: 10.1103/PhysRevE.49.1685.

Chung-Kang Peng, Shlomo Havlin, H. Eugene Stanley, and Ary L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87, 1995. doi: 10.1063/1.166141.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, and Bin Yang. TFB: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17(9): 2363–2377, 2024. doi: 10.14778/3665844.3665863.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*, 2021.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, 2015.

Thomas Schreiber and Andreas Schmitz. Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3–4):346–382, 2000. doi: 10.1016/S0167-2789(00)00043-9.

Murad S. Taqqu, Vadim Teverovsky, and Walter Willinger. Estimators for long-range dependence: An empirical study. *Fractals*, 3(4):785–798, 1995. doi: 10.1142/S0218348X95000692.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSDI: Conditional score-based diffusion models for probabilistic time series imputation. In *Advances in Neural Information Processing Systems 34*, 2021.

James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D: Nonlinear Phenomena*, 58 (1–4):77–94, 1992. doi: 10.1016/0167-2789(92)90102-S.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Dennis Wagner, Fabian Hartung, Justus Arweiler, Aparna Muraleedharan, Indra Jungjohann, Arjun Nair, Steffen Reithermann, Ralf Schulz, Michael Bortz, Daniel Neider, Heike Leitte, Joachim Pfeffinger, Stephan Mandt, Sophie Fellenz, Torsten Katz, Fabian Jirasek, Jakob Burger, Hans Hasse, and Marius Kloft. NoBOOM: Chemical process datasets for industrial anomaly detection. In *Advances in Neural Information Processing Systems 38, Datasets and Benchmarks Track*, 2025. doi: 10.52202/085713-3499.

Andrew T. A. Wood and Grace Chan. Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432, 1994. doi: 10.1080/10618600.1994.10474655.

Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Shirui Pan, and Qingsong Wen. A survey on diffusion models for time series and spatio-temporal data. *ACM Computing Surveys*, 58(8):1–39, 2026. doi: 10.1145/3736631.

Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In *Advances in Neural Information Processing Systems 32*, 2019.

Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation. In *International Conference on Learning Representations*, 2024.

Table 2: Recommended reference release. Values should be updated to the final released manifest.

Property	Reference value
Core families	7
Reference realizations per family	1,000
Core sequences	7,000
Sequence length	4,096
Features per sequence	1
Independent primary seed streams	7 or more
Calibration set	Independent source realizations
Storage formats	NumPy/NPZ + CSV/JSON metadata
Generator parameterization	Family-specific, manifest-recorded

A SUPPLEMENTARY MATERIALS

A.1 DATASET STATISTICS AND RELEASE STRUCTURE

The public release is organized by process family and parameter condition rather than by arbitrary file number. Each family contains the raw time series, a parameter manifest, and precomputed calibration descriptors. A compact directory structure is:

```
FractalTS/
  data/
    monofractal/
    multifractal_anti_correlated/
    multifractal_uncorrelated/
    multifractal_correlated/
    multifractal_pink_noise/
    multifractal_nonstationary/
    multifractal_brownian/
  metadata/
    manifest.csv
    calibration.json
    parameter_grid.json
  code/
    noise_generator.py
    validators/
  docs/
    datasheet.md
    benchmark_protocol.md
```

The final release should also contain checksums for data files and a machine-readable software environment. For every generated artifact, users should be able to recover the family, parameter values, random seed, sequence length, generator version, and validation status.

A.2 REFERENCE FAMILY NAMES AND INTERFACE

The reference release provides six dedicated per-family generator notebooks (`monofractal`, `multifractal_anti_correlated`, `multifractal_uncorrelated`, `multifractal_correlated`, `multifractal_pink_noise`, `multifractal_brownian`) and a unified class-based generator that additionally provides `multifractal_nonstationary` and a broader library of noise types. Each per-family generator takes an instance count and a sequence length, applies the family's construction and rejection-sampling acceptance loop, and writes each accepted realization as an individual MATLAB `.mat` file (keyed `data`) under a family-named directory. The generators depend on `numpy` and `scipy` (for `.mat` I/O), the `fbm` library (Flynn, 2020) for exact fractional-Gaussian-noise synthesis in the monofractal family, and the `MFDFA` library (Gorjão et al., 2022) for the acceptance-time multifractal analysis.

A.3 Representative Realizations

Figure 1 shows one representative realization from each of the seven core families, drawn with a fixed seed at the reference length $T = 4096$. The panels use the actual per-family reference generators (and the unified generator for the non-stationary family), so they depict the released data rather than idealized sketches. The visual signatures match the intended regimes: the monofractal fGn is stationary and noise-like; the anti-correlated family is rough with rapid sign reversals; the uncorrelated family shows the sparse, localized bursts characteristic of a binomial cascade acting on white noise; the correlated family is visibly smoother and more persistent; the pink ($1/f$) family exhibits slow, long-range wandering; the Brownian ($1/f^2$) family is a smooth integrated random walk; and the non-stationary family displays an explicit variance envelope superimposed on a drift.

A.4 Measured Validation Table

Table 3 reports empirical structural descriptors for the six per-family reference generators, measured directly on the raw construction (before the rejection-sampling acceptance loop) over $N = 100$ independent realizations per family at $T = 4096$. Estimation uses the benchmark's acceptance-time MFDFA configuration: scales $s \in \{10, \ldots, 100\}$, the 41-point moment grid $q \in [-5, 5] \setminus \{0\}$, and order-1 detrending, with the singularity spectrum width $\widehat{\Delta\alpha}$ obtained from the Legendre transform of $\tau(q) = q\,h(q) - 1$. Confidence intervals are nonparametric bootstrap 95% intervals over realizations (Efron, 1979).

Two points follow from the measurements. First, the measured Hurst exponent $\widehat{H}$ closely tracks each notebook's target across the full range of regimes, from strongly anti-persistent (0.27 vs. target 0.25) through uncorrelated (0.59; the binomial-cascade modulation biases the q=2 slope slightly above 0.5), persistent (0.74 vs. 0.75), pink (1.00), and Brownian (1.49 vs. 1.50). Second, the raw-construction singularity width $\widehat{\Delta\alpha}$ varies substantially across families: the binomial-cascade uncorrelated family is intrinsically broad ($\widehat{\Delta\alpha} \approx 1.25$), whereas the spectral pink, Brownian, monofractal, and correlated constructions are intrinsically narrow at these short scales ($\widehat{\Delta\alpha} \approx 0.11$–$0.15$). This is exactly why the pipeline applies rejection sampling: the acceptance criteria (last column) retain only realizations whose *measured* width meets the family's threshold, so the released monofractal set is certified to $\Delta\alpha < 0.05$ and the released multifractal sets to $\Delta\alpha > 0.25$ (or > 0.7 for the uncorrelated family), even when the pre-rejection mean does not. The benchmark thus ships operationally certified descriptors rather than relying on idealized infinite-scale formulas.

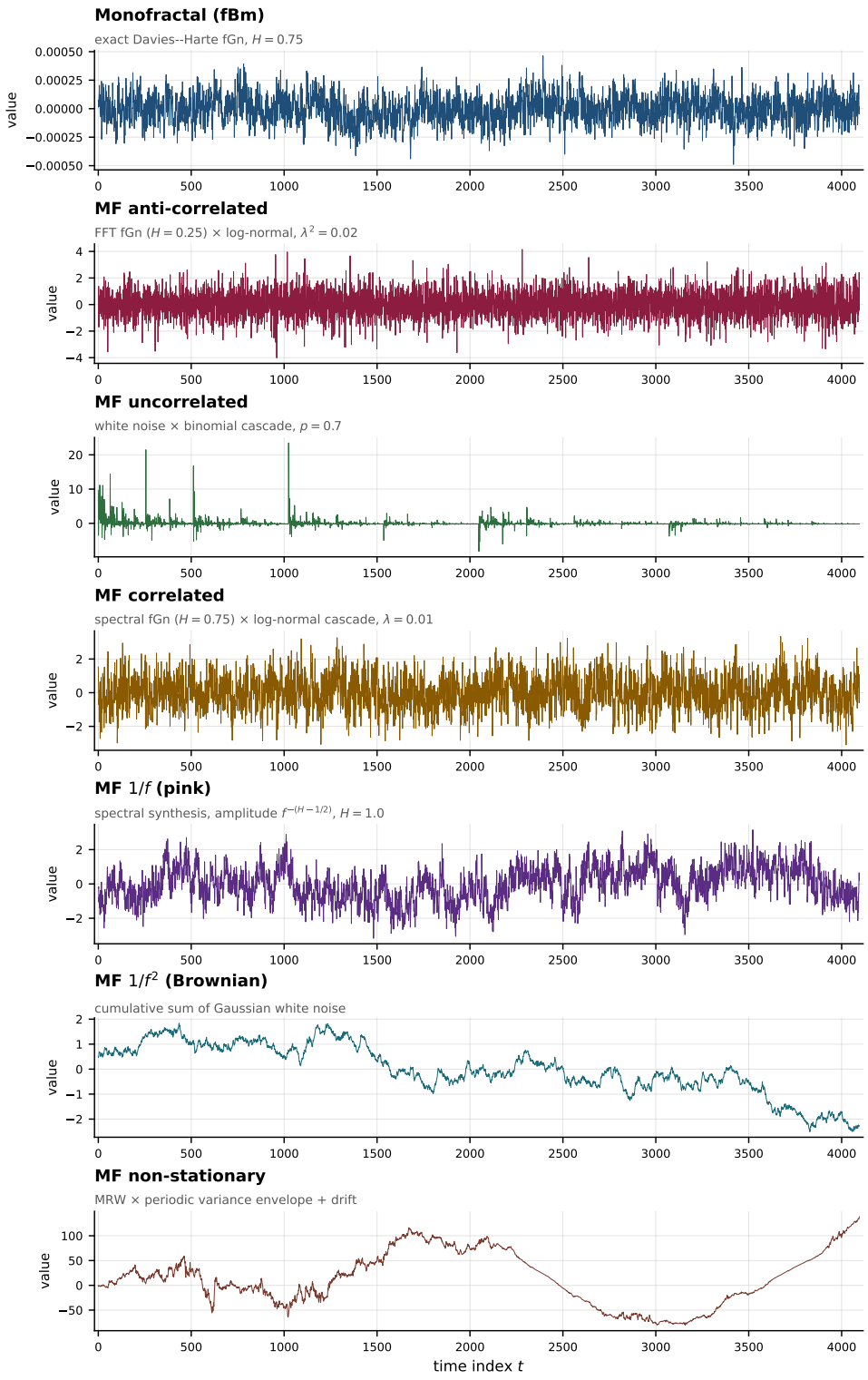


Figure 1: One representative realization of each of the seven FRACTALTS families at $T = 4096$ (fixed seed). Each panel is annotated with the family name and the construction used by its reference generator. Vertical scales differ across panels because the families are standardized differently by construction (several are normalized to unit variance, whereas the monofractal and non-stationary families are shown in their native units).

Table 3: Structural descriptors of the six per-family reference generators, measured on the *raw construction* (before rejection sampling) over $N = 100$ independent realizations at $T = 4096$, with nonparametric bootstrap 95% confidence intervals in brackets. Estimation uses the benchmark's acceptance-time MFDFA settings (scales 10–100, $q \in [-5, 5] \setminus \{0\}$ on a 41-point grid, order-1 detrending). $\widehat{H}$ is the generalized Hurst exponent at q=2; $\widehat{\Delta\alpha}$ is the singularity spectrum width. The measured $\widehat{H}$ closely tracks each notebook's target Hurst value. The last column lists the acceptance criterion enforced at generation time; because rejection sampling retains only realizations meeting these bounds, the *released* sets satisfy the stated $\Delta\alpha$ thresholds exactly, whereas the raw-construction $\widehat{\Delta\alpha}$ shown here is the pre-rejection mean.

Family	Generator	Target H	$\widehat{H}$	$\widehat{\Delta\alpha}$	Acceptance criterion
Monofractal	Davies–Harte fGn	0.75	0.76 [0.75, 0.76]	0.13 [0.11, 0.14]	$\Delta\alpha<0.05$
MF anti-correlated	fGn × log-normal	0.25	0.27 [0.27, 0.28]	0.23 [0.22, 0.24]	$0<H<0.5,\ \Delta\alpha>0.25$
MF uncorrelated	white × cascade	0.50	0.59 [0.58, 0.60]	1.25 [1.19, 1.33]	$0.48<H<0.52,\ \Delta\alpha>0.7$
MF correlated	spectral fGn × cascade	0.75	0.74 [0.74, 0.75]	0.13 [0.12, 0.15]	$0.5<H<1,\ \Delta\alpha>0.25$
MF $1/f$ (pink)	spectral synth.	1.00	1.00 [0.99, 1.00]	0.11 [0.09, 0.13]	$0.99<H<1.01,\ \Delta\alpha>0.25$
MF $1/f^2$ (Brownian)	cumsum(white)	1.50	1.49 [1.48, 1.49]	0.15 [0.13, 0.17]	$1.49<H<1.51,\ \Delta\alpha>0.25$