IMPERIAL

ICLR

# Random Controlled Differential Equations

Francesco Piatti, Thomas Cass, William F. Turner

## Background

**(Neural) Controlled Differential Equations**
CDEs model a path $x_t : [0,T] \to \mathbb{R}^d$ by continuously evolving a state $y_t \in \mathbb{R}^N$ according to

$$y_t = y_0 + \int_0^t f(y_s)\, dx_s, \qquad y_0 \in \mathbb{R}^N, \tag{1}$$

where $f : \mathbb{R}^N \to \mathbb{R}^{N\times d}$ is a collection of vector fields.. If $f$ is parameterized by a neural network, this yields a Neural CDE, which can be viewed as a continuous-time analogue of an infinitely-deep ResNet.

**Random Infinitely-Deep ResNets [1]**
If the vector field $f$ in Eq. 1 is instead sampled at random and then frozen, one obtains a random continuous-time residual model, i.e. a randomly initialized infinitely-deep ResNet. Given an input path $x_t : [0,T] \to \mathbb{R}^d$, consider

$$dZ_t^N = \frac{1}{\sqrt{N}} \sum_{i=1}^{d} A_i\, \varphi(Z_t^N)\, dx_t^i, \qquad Z_0^N = z_0 \in \mathbb{R}^N, \tag{2}$$

where the matrices $A_i$ are random and $\varphi$ is a non-linearity. The terminal state $Z_T^N$ provides a random feature representation of the input path, on top of which only a linear readout is trained for classification or regression.

**Signatures and Signature Kernels**
The signature of a path $x_t : [0,T] \to \mathbb{R}^d$ is the collection of its iterated integrals

$$\mathsf{Sig}(x)_{0,T} = \left(1, \int_0^T dx_t, \int_{0<t_1<t_2<T} dx_{t_1} \otimes dx_{t_2}, \dots \right) \in T((\mathbb{R}^d)),$$

where $T((\mathbb{R}^d))$ is the tensor algebra. Signatures give an infinite-dimensional representation of sequential data, are universal, robust to irregular sampling, and are typically truncated in practice. Equipping the tensor algebra with a suitable inner product yields the signature kernel

$$K^{x,y}_{\mathsf{Sig}}(s,t) = \langle \mathsf{Sig}(x)_{0,s}, \mathsf{Sig}(y)_{0,t} \rangle_{T((\mathbb{R}^d))}, \tag{3}$$

which defines a powerful kernel method for time series. In practice, an RBF-lifted variant often yields stronger empirical performance [4].

**Rough Paths and Rough Signature Kernels**
Rough path theory extends controlled differential equations to highly irregular signals by enriching a path with its low-order iterated integrals [2]. Geometric $p$-rough paths ($G\Omega_p$) arise as limits, in rough-path topology, of bounded-variation paths equipped with their signatures. Their logarithms define log-signatures, a more compact representation retaining only Lie terms and removing algebraic redundancies. This framework also yields the rough signature kernel

$$K^{\mathbb{X},\mathbb{Y}}_{\mathsf{Sig}}(s,t) = \langle \mathsf{Sig}(\mathbb{X})_{0,s}, \mathsf{Sig}(\mathbb{Y})_{0,t} \rangle_{T((\mathbb{R}^d))}, \tag{4}$$

and a well-posed theory of rough differential equations under suitable regularity of the vector fields [2].

## Models

**Motivation and Contributions.** Building on the R-CDE framework introduced by [1], we investigate two natural extensions motivated by practice: RBF-lifted signature kernels often outperform the plain signature kernel, and rough-path dynamics can better capture highly irregular time series. Retaining the training-efficient random-reservoir viewpoint — a large random controlled system with only a linear readout trained — we:

1. develop two practically useful variants:
   - **RF-CDE**: random Fourier lifting before the dynamics,
   - **R-RDE**: a rough-path version based on a log-ODE discretisation.
2. prove that these models admit clean **infinite-width kernel limits**.
3. show that the resulting models are **training-efficient** and competitive on real and synthetic benchmarks.

**RF-CDE (Random Fourier Controlled Differential Equations)**
RF-CDE combines continuous-time random reservoirs with a random Fourier lift [3] of the input. Given a path $x_t : [0,T] \to \mathbb{R}^d$, we first lift it pointwise to a higher-dimensional path

$$X_t^F = \phi^F(x_t) \in \mathbb{R}^{2F},$$

where $\phi^F$ is a random Fourier feature map. The hidden state $Z_t^{N,F} \in \mathbb{R}^N$ then evolves according to

$$dZ_t^{N,F} = \frac{1}{\sqrt{N}} \sum_{i=1}^{2F} A_i\, \varphi(Z_t^{N,F})\, dX_t^{F,i}, \qquad Z_0^{N,F} = z_0. \tag{5}$$

where $\{A_i\}_{i=1}^{2F}$ is a collection of random matrices, independent across $i$ and from the RFF randomness, $X_t^{F,i}$ denotes the $i$-th component of the lifted path, and $\varphi$ is a non-linearity.

Thus RF-CDE is a random CDE driven by a nonlinear lift of the input path. The Fourier lift enriches the geometry seen by the reservoir, allowing the model to capture nonlinear similarities between paths.

In practice, we discretize Eq. 5 and tune the associated scaling and bias hyperparameters by grid search.

**R-RDE (Random Rough Differential Equations)**
To handle non-smooth signals, we work directly with geometric rough paths $\mathbb{X}$, which retain higher-order information through signatures or log-signatures. Given random matrices $\{A_i\}_{i=1}^d$, we first define the linear development

$$S_t^A(x) := \Gamma_A(\mathsf{Sig}(x)_{0,t}), \quad \Gamma_A(\mathbb{G}) = \sum_{w=i_1\cdots i_k} A_w \langle \mathbb{G}, w\rangle, \quad A_w = \frac{1}{N^{|w|/2}} A_{i_1} \cdots A_{i_k},$$

where $w = i_1 \cdots i_k$ ranges over words on the alphabet $\{1,\dots,d\}$. For smooth paths $x$, we then extend this by continuity to rough paths, obtaining a matrix-valued rough driver $S^A(\mathbb{X})$. The random feature path $Z^N(\mathbb{X}) \in \mathbb{R}^N$ is defined by

$$dZ_t^N(\mathbb{X}) = f(Z_t^N)\, dS_t^A(\mathbb{X}), \qquad Z_0^N = z_0, \tag{6}$$

where typically $f(z)[M] = M(\varphi(z))$.

In practice, we discretize it with a log–ODE scheme: on each interval $[t_i, t_{i+1}]$, we use the truncated log-signature $\log_m(\mathbb{X}_{t_i,t_{i+1}})$ and map its Lie coordinates to nested commutators of random matrices, yielding explicit updates that preserve the algebraic structure of the rough path.

## Theoretical Results

We establish that the proposed models are both well posed and mathematically interpretable through explicit infinite-width kernel limits.

**Theorem** (Existence and uniqueness of R-RDEs)**.** Let $\mathbb{X} \in G\Omega_p(\mathbb{R}^d)$ with $p \geq 1$ and $\varphi \in \mathrm{Lip}(\gamma)$ with $\gamma > p$. Then the R-RDE 6 admits a unique solution $Z^N \in C([0,T];\mathbb{R}^N)$, and the Itô–Lyons map $(\mathbb{X}, z_0) \mapsto Z^N$ is continuous in the rough-path topology.

**Kernel limits.**
Random reservoirs are not just heuristics: in the infinite-width limit, they converge to structured kernels on path space. In [1], the R-CDE model (Eq. 2) is shown to converge to the signature kernel. Here we prove analogous limits for our variants: RF-CDE converges to the RBF-lifted signature kernel, and R-RDE to the rough signature kernel.

**Theorem.** Let $x_t, y_t$ be differentiable paths on $[0,T]$ and let $Z_s^{N,F}(x), Z_t^{N,F}(y) \in \mathbb{R}^N$ solve Eq. 5 with $\varphi = \mathrm{id}$ and the same $\{A_i\}_{i=1}^{2F} \overset{\text{i.i.d.}}{\sim} \xi_N$ (independent of the RFF draw). Then, for every $s, t \in [0,T]$

$$\lim_{F\to\infty} \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\xi_N}\left[\left\langle Z_s^{N,F}(x), Z_t^{N,F}(y) \right\rangle_{\mathbb{R}^N}\right] = K^{x,y}_{\mathsf{Sig\text{-}RBF}}(s,t),$$

where $K^{x,y}_{\mathsf{Sig\text{-}RBF}}(s,t)$ denotes the RBF–lifted signature kernel.

**Theorem.** Let $\mathbb{X} \in G\Omega_p(\mathbb{R}^d)$ and $\mathbb{Y} \in G\Omega_q(\mathbb{R}^d)$ be geometric rough paths. Let $Z_s^N(\mathbb{X})$ and $Z_t^N(\mathbb{Y})$ be the solutions of Eq. 6 with $\varphi = \mathrm{id}$ and the same matrices $\{A_i\}_{i=1}^d \overset{\text{i.i.d.}}{\sim} \xi_N$. Then, for every $s, t \in [0,T]$

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{\xi_N}\left[\left\langle Z_s^N(\mathbb{X}), Z_t^N(\mathbb{Y}) \right\rangle_{\mathbb{R}^N}\right] = K^{\mathbb{X},\mathbb{Y}}_{\mathsf{Sig}}(s,t),$$

where $K^{\mathbb{X},\mathbb{Y}}_{\mathsf{Sig}}$ denotes the rough signature kernel (Eq. 4).

All proofs are given in the paper.

**Gaussian-process interpretation.**
With a fixed random reservoir and a trained linear readout, RF-CDE implements kernel ridge regression with kernel $N^{-1}\langle Z_s^{N,F}(x), Z_t^{N,F}(y)\rangle$, which converges to the RBF-lifted signature kernel as $N, F \to \infty$. Likewise, R-RDE converges to the rough signature kernel as $N \to \infty$. In both cases, this yields Gaussian-process priors over path-functionals with the corresponding limiting covariance kernels. This provides a clear interpretation of the models' inductive bias: they inherit the expressive structure of signature-based kernels while retaining the scalability of random-feature reservoirs.

## Empirical Results

**Benchmarks.** We compare primarily against Random Fourier Signature Features (RFSF) [5] in its two variants, Diagonal Projection (DP) and Tensorized Random Projection (TRP), and against SigPDE [4], a signature-kernel SVM method that requires Gram matrix inversion. The paper also reports results for a broader set of classical kernel baselines, as well as Neural CDE and Neural RDE in selected experiments.

**Multivariate time-series classification.** Across the 16 UEA datasets considered in the paper, **RF-CDE** is the strongest random-feature method on average, while remaining computationally light since only the readout is trained. The table below reports results with $N = 250$ features.

| Model | Avg. acc. ↑ | Avg. rank ↓ |
|---|---|---|
| R-CDE | 0.695 | 4.250 |
| **RF-CDE** | **0.741** | 3.062 |
| R-RDE | 0.708 | 4.125 |
| RFSF-DP | 0.726 | 3.406 |
| RFSF-TRP | 0.725 | 3.594 |
| SigPDE | 0.738 | **2.562** |

**Hurst-exponent classification** We also test the models on a fractional Brownian motion benchmark, where the task is to recover the Hurst exponent. We consider **V1** (raw trajectories) and **V2** (per-sample standardisation, so models must rely more on geometry than scale).

| Model | V1, $N = 64$ | V2, $N = 64$ |
|---|---|---|
| R-CDE | 0.870 | 0.635 |
| RF-CDE | 0.895 | 0.645 |
| **R-RDE** | **0.955** | **0.735** |
| RFSF-DP | 0.840 | 0.630 |
| RFSF-TRP | 0.895 | 0.650 |
| NCDE | 0.905 | 0.650 |
| NRDE | 0.920 | 0.675 |

**Ablation and robustness.** We also study ablations on the number of features and robustness to missing data: RF-CDE is particularly strong in the low-feature regime and shows the most stable degradation under random missing observations.

**References**

[1] N. M. Cirone, M. Lemercier, and C. Salvi. "Neural signature kernels as infinite-width-depth-limits of controlled resnets". In: International Conference on Machine Learning. PMLR. 2023, pp. 25358–25425.

[2] T. J. Lyons. "Differential equations driven by rough signals". In: Revista Matemática Iberoamericana 14.2 (1998), pp. 215–310.

[3] A. Rahimi and B. Recht. "Random features for large-scale kernel machines". In: Advances in neural information processing systems 20 (2007).

[4] C. Salvi, T. Cass, J. Foster, T. Lyons, and W. Yang. "The signature kernel is the solution of a goursat pde". In: SIAM Journal on Mathematics of Data Science 3.3 (2021), pp. 873–899.

[5] C. Toth, H. Oberhauser, and Z. Szabo. "Random Fourier Signature Features". In: SIAM Journal on Mathematics of Data Science 7.1 (2025), pp. 329–354.