Signatures, trees and pdes

DataSig

A rough path between mathematics and data science

Terry Lyons

CIRM - 12 March 2021

with many others … but particularly Cris, Patrick, James, James, Varun, Cris, Maud, Peter, Roly, Sam,… Tom, Harald, Hao, Patricia

University of Oxford · The Alan Turing Institute · Imperial College London · UCL · UKRI Engineering and Physical Sciences Research Council

A DataSig vision

DataSig

A rough path between mathematics and data science

"We channel our research around developing the mathematics needed to model and understand complex streams of non-stationary multimodal data. We build prototypes that have real world value to develop this understanding. This is only possible because of significant collaboration and partnership."

DataSig

A rough path between mathematics and data science

Modelling behavior of evolving systems

A public domain collaboration on detecting malware

DataSig

A rough path between mathematics and data science

DataSig

Terry

Cris

Patricia

Imanol

Peter

Maud

Roly

Thomas

Varun

The Alan Turing Institute

DataSıg: an EPSRC/UKRI 5-year program grant

DataSıg
A rough path between mathematics and data science

Mathematics

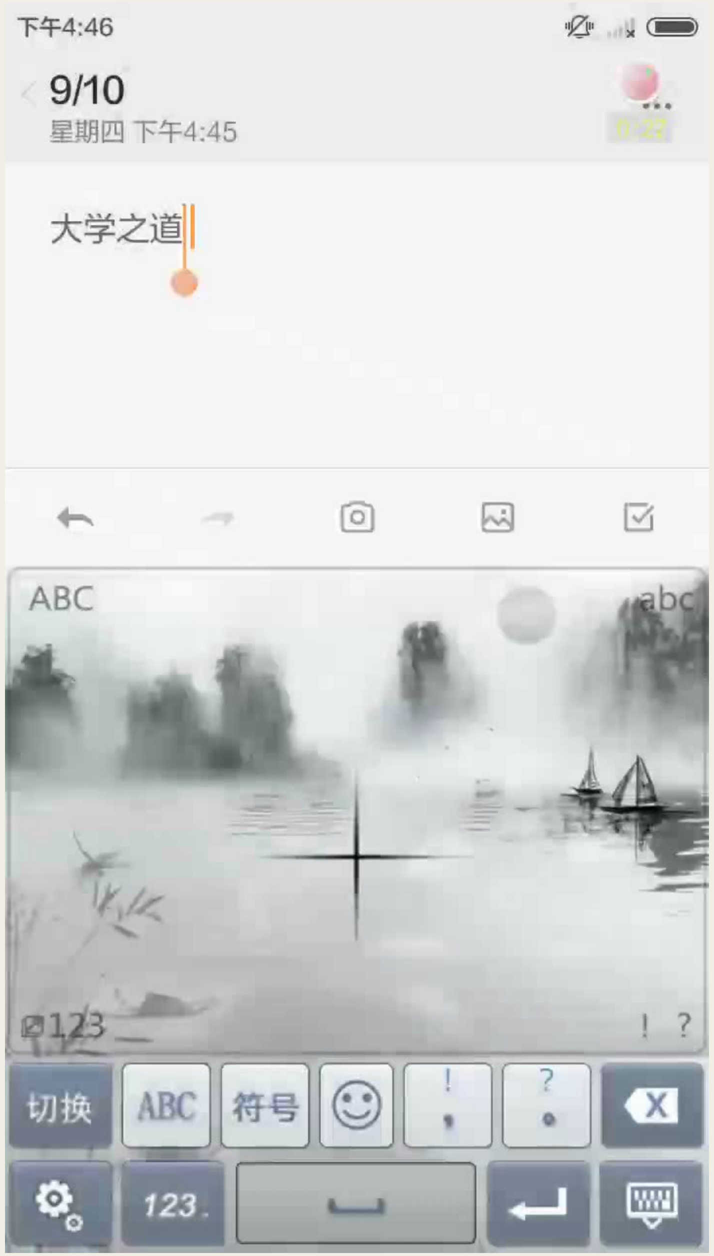
- rough path theory and signatures
- describing the interactions between complex systems from the top down
- extending the calculus of differential equations to complex contexts

Data science

- the notion of an unparameterized path captured by the order of events
- clean and minimal universal feature sets - (expected) signature
- the notion of a neural controlled differential equation
- the notion of a pde-kernel
- a principled mathematical framework that allows further innovation (e.g. simulation)

Embedded contexts

- streamed data is everywhere; Chinese handwriting, hospital wards, event logs …



Streamed data

- a character drawn on the screen of an iPhone
- an order book
- a piece of text
- progression through hospital record
- astronomical data
- video of a person moving
- an evolving stream of emotions
- ICU data to detect sepsis
- the evolving stock position in a supermarket or computer switch

Ensembles of streamed data

- the event log of processes generated by malware
- the behaviour of crowds
- the evolution of cancer cell lines

Key questions

- understand what you have observed
- predict the distribution of what is happening next
- identify anomalies

DataSig

A rough path between mathematics and data science

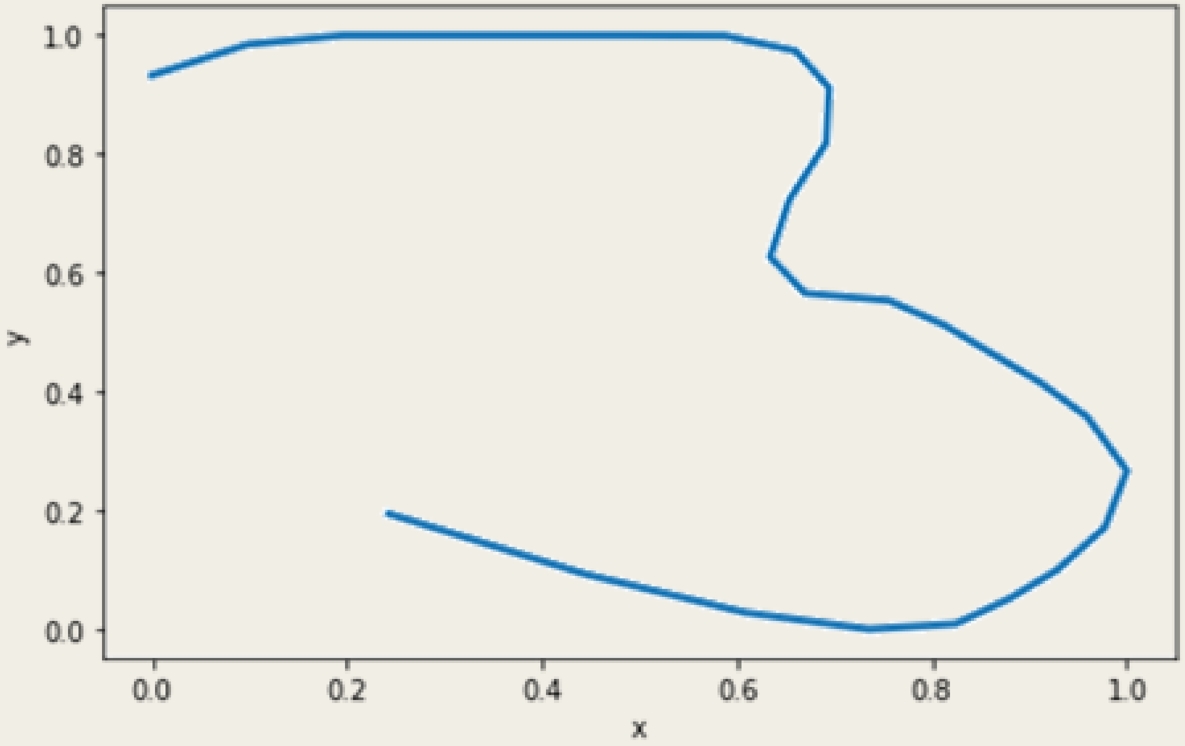
DataSig

A rough path between mathematics and data science

Some maths of evolving systems

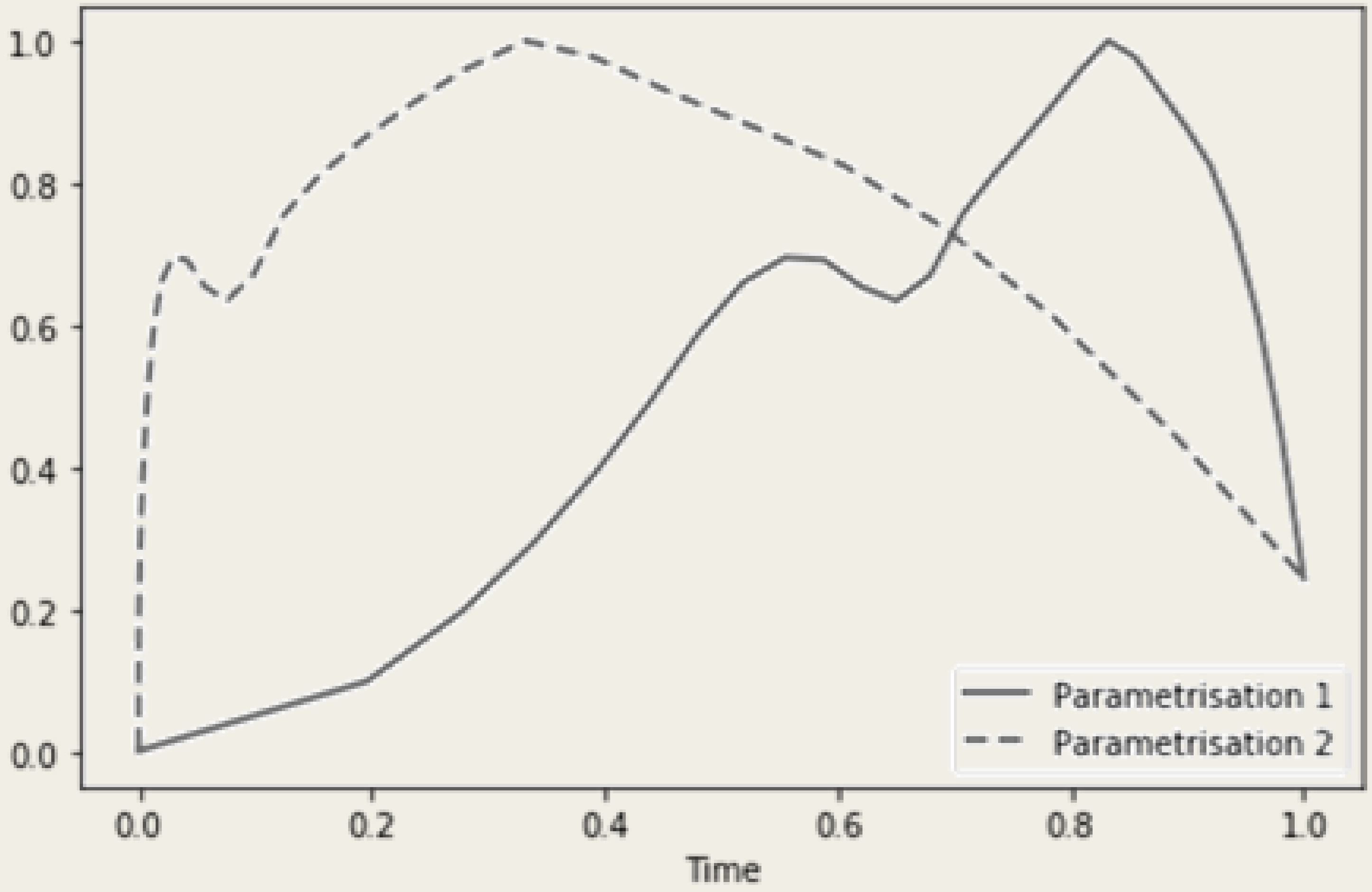
Data science does not like symmetry

- Re-parameterisation is a huge symmetry group
- Multimodal streams modulo re-parameterisation form a group
- Representing this group in the tensor algebra provides a faithful feature set and removes the symmetry
- Drawn from old mathematics, new tools, signature and log signature, and new maths describing the functions on streams



Different sampling procedures

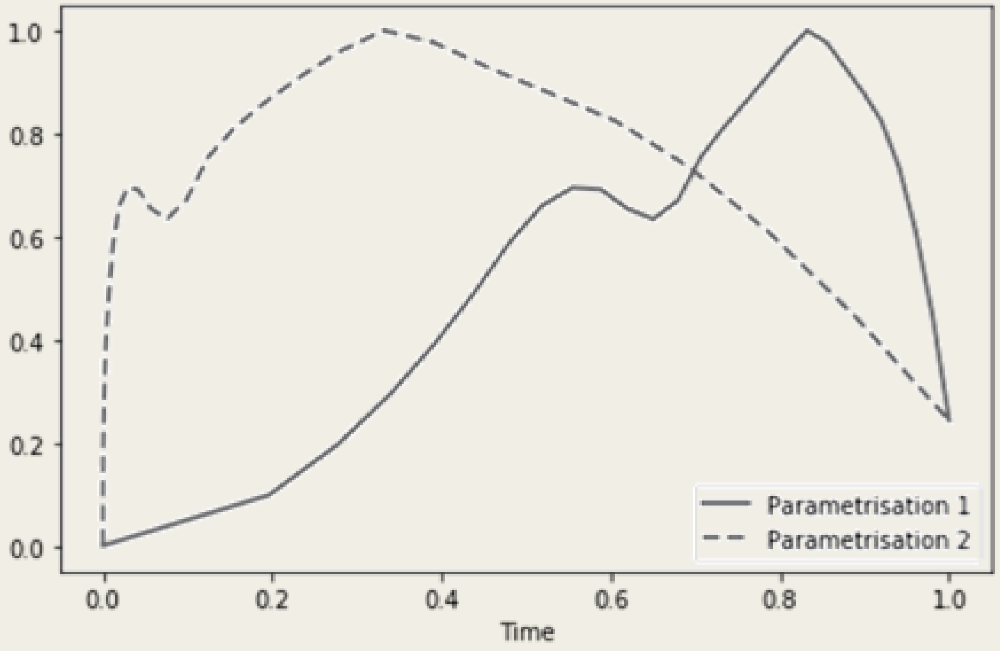
- The letter “3” is drawn from top to bottom
- The x coordinate of the evolving symbol sampled differently (at uneven speeds)

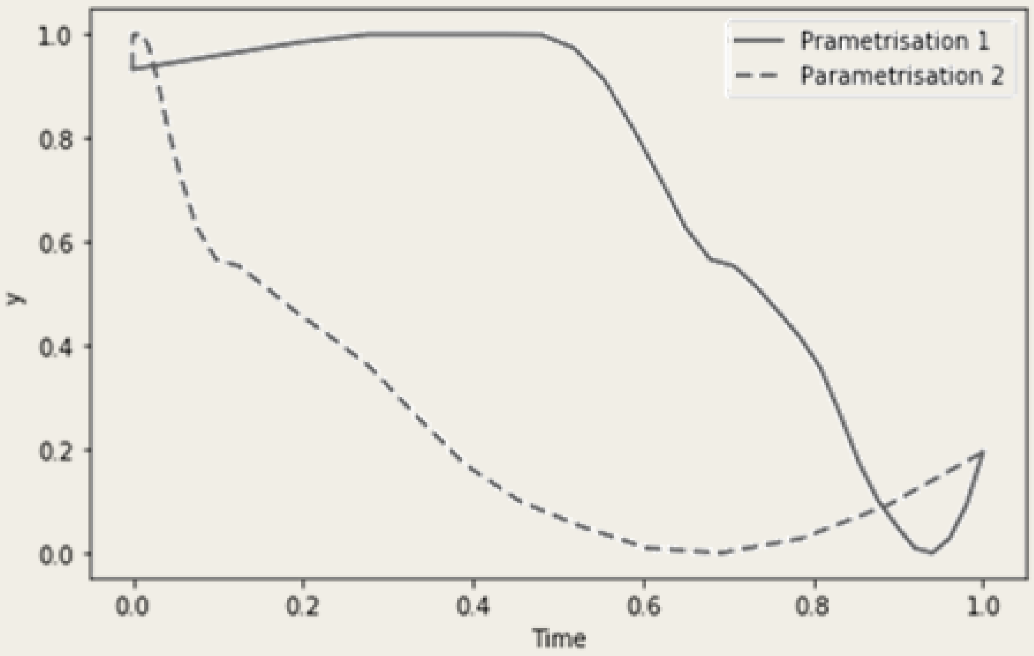


Different sampling procedures

The number “3” x, y coordinates – same picture drawn at two different speeds

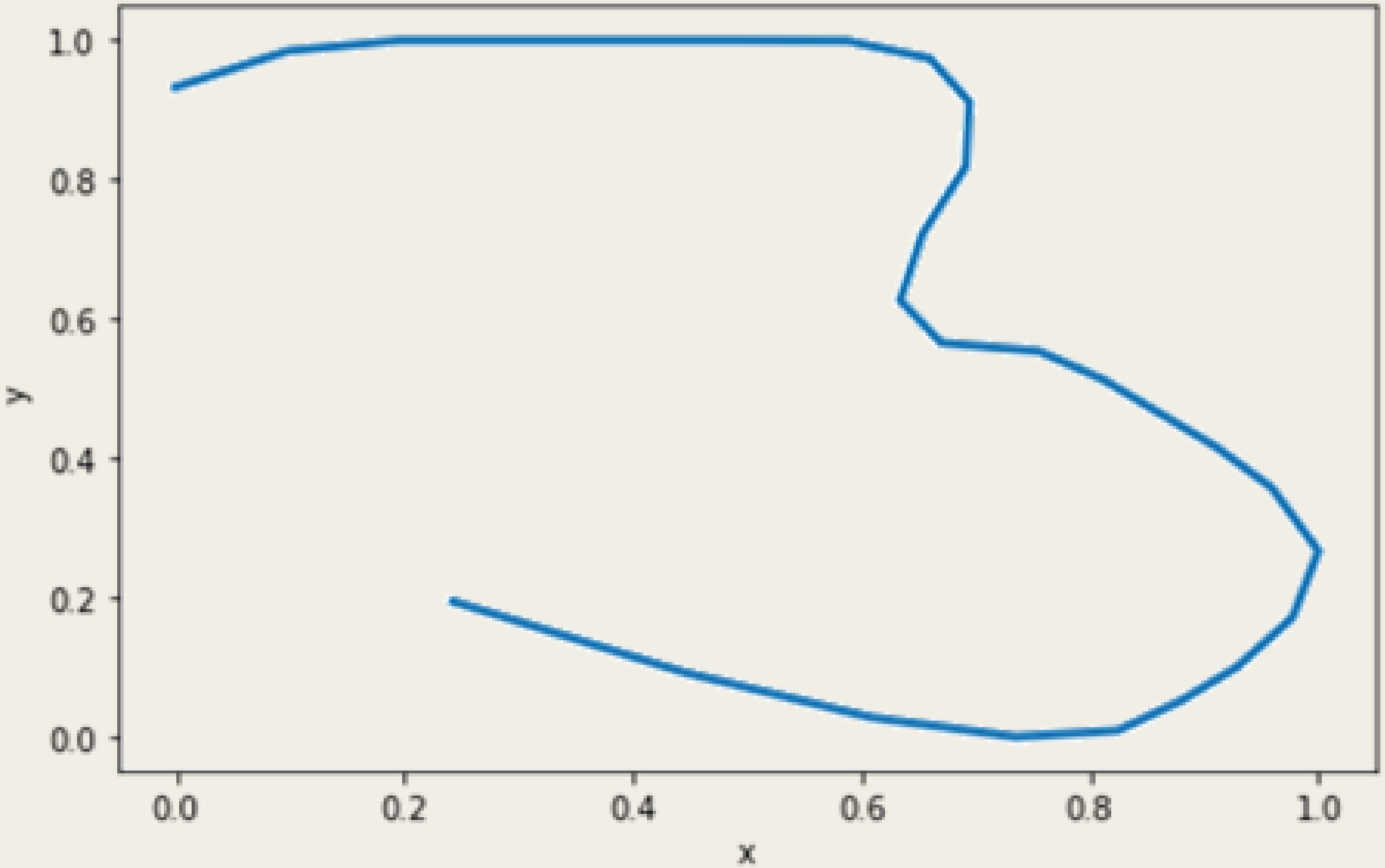
- no consistent wavelets
- reparameterisations do not form a linear space!





Different sampling procedures

- The letter “3” is drawn from top to bottom
- How does one describe the three or any path modulo the symmetry of parametrisation?



The signature of a path describes an unparameterised stream γ

A rough path between mathematics and data science

Signature is a *top down* description for unparameterised paths that describes a path segment through its effects of stylised nonlinear systems

$$dS = S \otimes d\gamma$$

It filters out the infinite dimensional noise of resampling allowing prediction and classification with *much* smaller learning sets.

It gives fixed dimensional feature sets regardless of the sample points.*

* missing data/varying parameterisation not issues although inadequacy of sampling may be

The signature - faithful and universal features describing an unparameterised stream

The signature of a stream γ over $I = [s, t]$ defined by $\sum_{k=0}^{\infty} S_k$ where $S_0 = 1$ and

$$S_k(\gamma, I) := \iint\limits_{s<u_1<\cdots<u_k<t} d\gamma_{u_1} d\gamma_{u_2} \dots d\gamma_{u_k}$$

These “Fourier-like” features exactly describe the *unparameterised* stream (Hambly Lyons Annals Math 2010) up to appropriate null sets.

Projected controlled differential equations are universal models

$$< e, Y_t > \text{where } dY_u = f(Y_u) dX_u$$

Analysis, geometry, combinatorial Hopf/dendriform/sensor algebras

DataSig
A rough path between mathematics and data science

Signature leads to linear space of real valued sensors $\langle \boldsymbol{e} | \boldsymbol{S_I} \rangle$ on streams

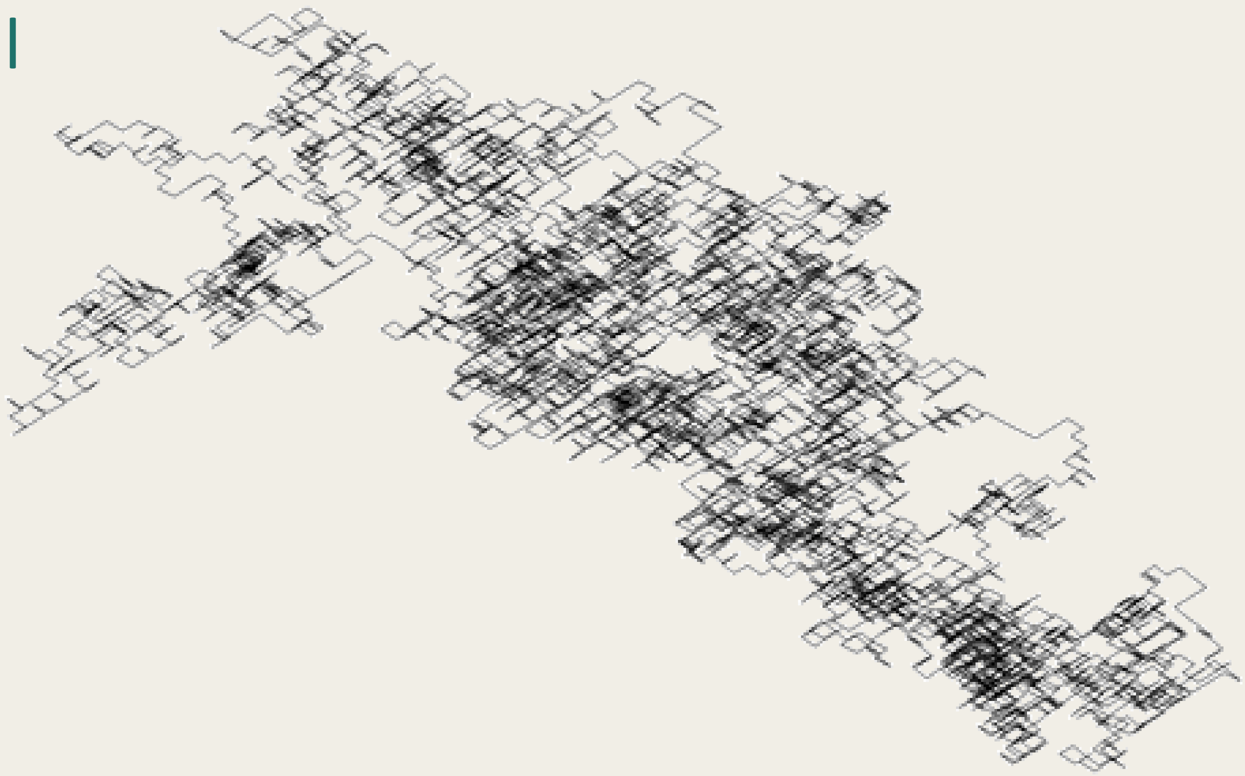
Pointwise multiplication and integration of these functionals

$$\langle \boldsymbol{\alpha} | \boldsymbol{\gamma} \rangle \langle \boldsymbol{\beta} | \boldsymbol{\gamma} \rangle = \langle \boldsymbol{\alpha} \sqcup\!\sqcup \boldsymbol{\beta} | \boldsymbol{\gamma} \rangle$$

$$\int \langle \boldsymbol{\alpha} | \boldsymbol{\gamma} \rangle \, \boldsymbol{d} \langle \boldsymbol{\beta} | \boldsymbol{\gamma} \rangle = \langle \boldsymbol{\alpha} \prec \boldsymbol{\beta} | \boldsymbol{\gamma} \rangle$$

can usefully be described in purely algebraic language.

The log signature is structurally important.



Analysis, geometry, combinatorial Hopf/dendriform/sensor algebras

- The *Signature* is a faithful embedding of the *unparametrized* stream into a vector space
- Continuous functions on streams can be well approximated by linear functionals on signatures
- The Expected Signature describes the ensemble of paths
- The log-signature describes paths without redundancy
- There is a natural pde kernel

Analysis, geometry, combinatorial Hopf/dendriform/sensor algebras

DataSig

A rough path between mathematics and data science

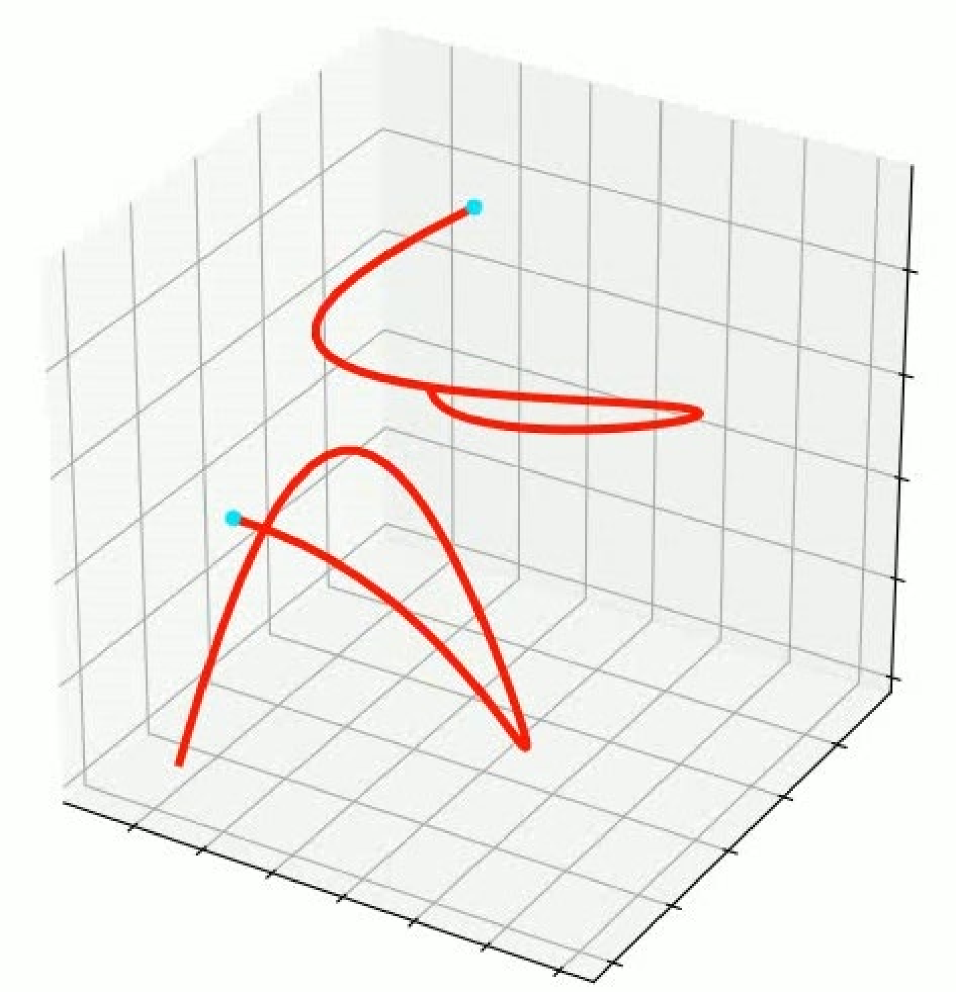
- Let be x, y be unparametrized paths in H and consider the bilinear form $K(x, y) \coloneqq \langle S(x), S(y) \rangle$. Franz J. Kiraly, Harald Oberhauser; JMLR 20(31):1-45, 2019 gave a kernel trick for the tangent kernel for the truncated signature embedding.
- Salvi et al. then identified the Goursat pde as the kernel trick for the untrucated kernel and gave analytic sense for it even for rough paths:

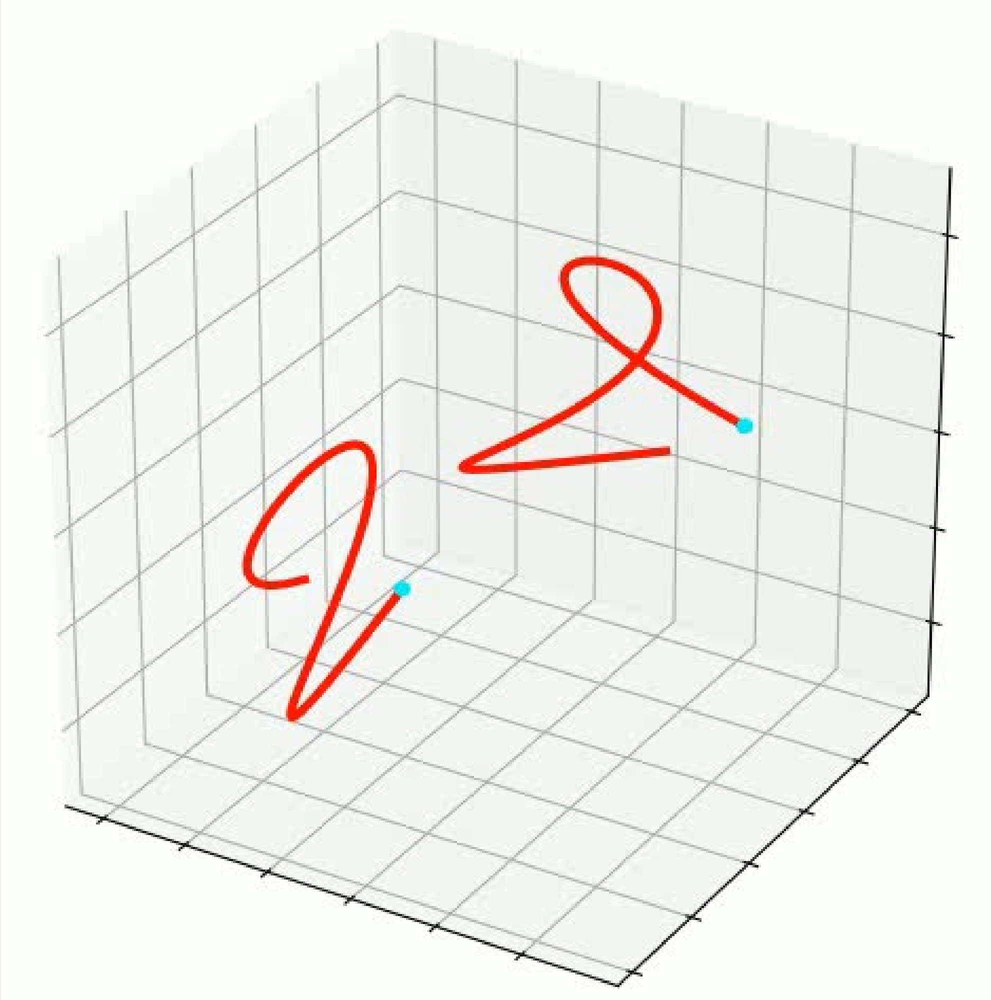
$$\frac{\partial^2 K(x|_{[u_0,u]}, y|_{[v_0,v]})}{\partial u \partial v} = \langle \dot{x}, \dot{y} \rangle K\left(x|_{[u_0,u]}, y|_{[v_0,v]}\right)$$

Recovering the curves from the signature

DataSig

A rough path between mathematics and data science

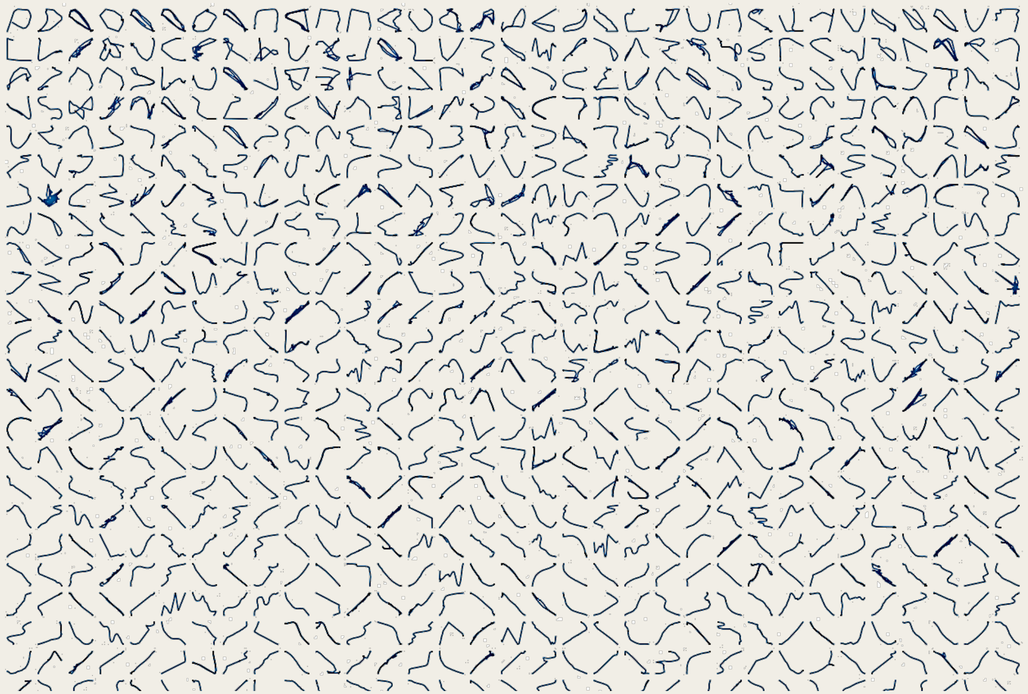




Weixin Yang, Jaiwei Chang

Fermanian

Our data



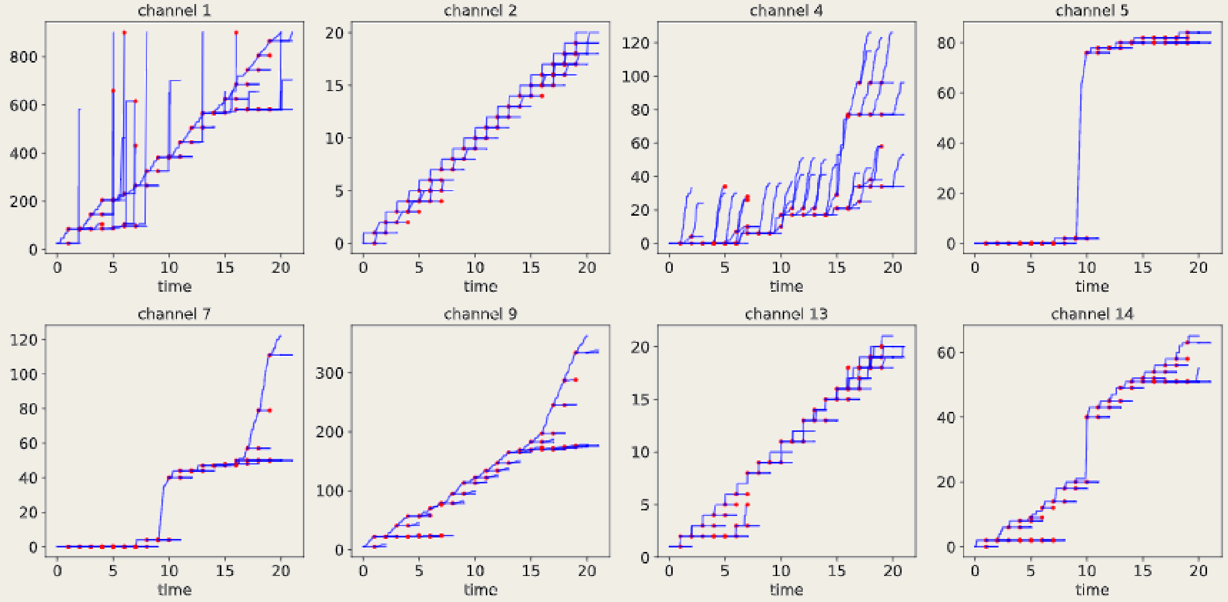
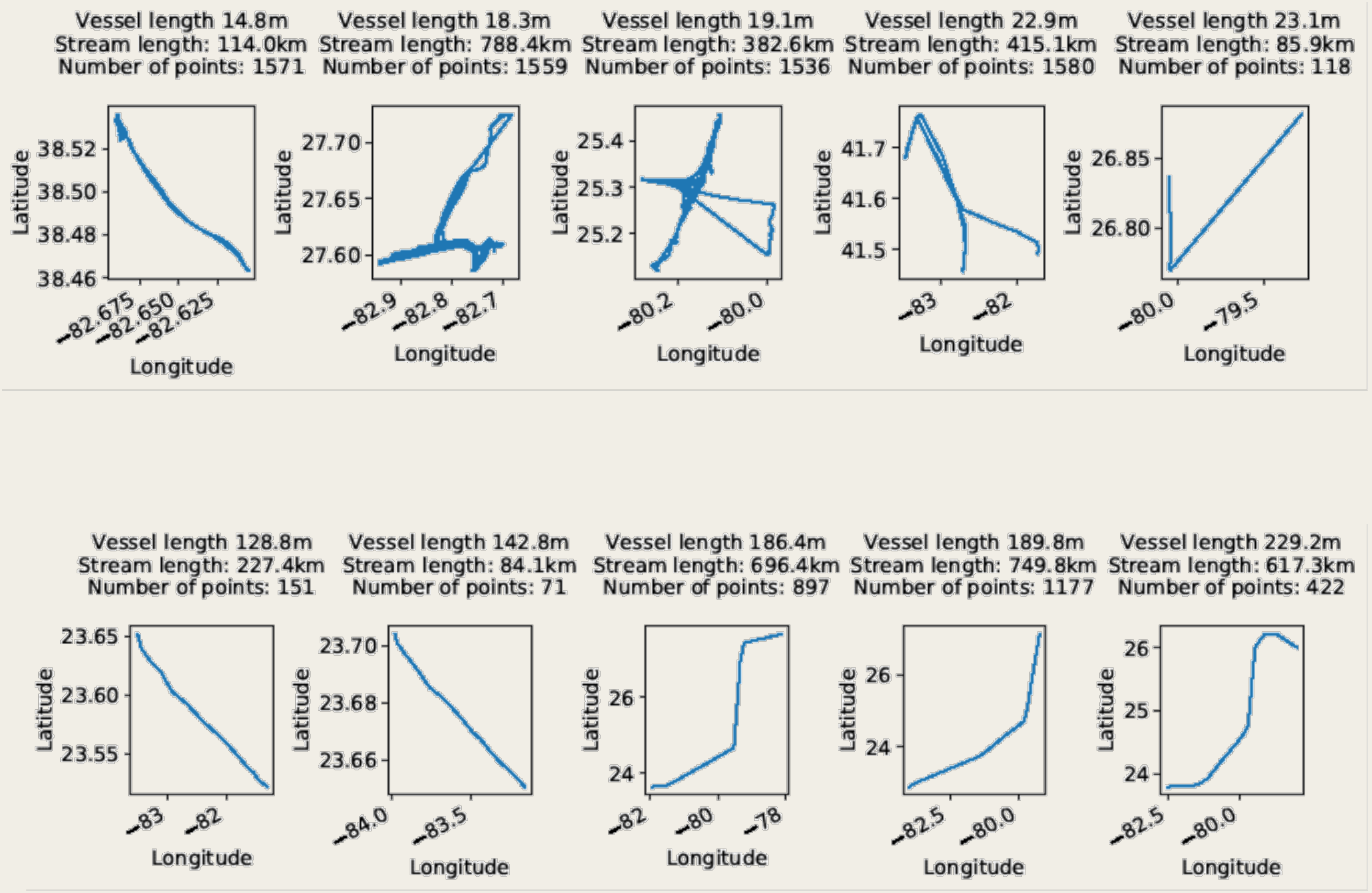


Fig. 1. Visual representation of selected channels of one single streaming tree. Each plot represents the evolution in time of the value of a given channel of the streaming tree, on its various branches. A red dot indicates a point where the currently-tracked process sets off a child process, causing the tree to branch.

Paths or Signatures?
Long/short vessel

- The vectorisation of unparametrized streams allows efficient use of many standard methods.
- For example, anomaly detection can be successfully applied this to these real-world shipping trajectories example.



Neural Controlled differential equations

DataSig

A rough path between mathematics and data science

All fields

Use controlled differential equations to model policies

DataSig

A rough path between mathematics and data science

Train a neural f so that

$$dy = f(y)d\gamma, y_0 = a$$

Where a represents the current state and γ determines the policy. Allows learning counterfactuals.

Neural Rough Differential Equations for Long Time Series

https://arxiv.org/abs/2009.08295

James Morrill, Cristopher Salvi, Patrick Kidger, James Foster, Terry Lyons

Ensembles of paths

DataSig

A rough path between
mathematics and data science

Process tree example:
Expected signatures of clouds of paths

DataSig

A rough path between mathematics and data science

Developed a way to apply expected signature techniques by viewing processes as trees evolving over time (eg the crop yield prediction task).

Predicting the yield of wheat crops over a region from the longitudinal measurements of climatic variables recorded across different locations of the region.

Eurostat dataset containing the total annual regional yield of wheat crops in mainland France - divided in 22 administrative regions - from 2015 to 2017.

Process tree example:
Expected signatures of clouds of paths

DataSig
A rough path between mathematics and data science

Viewing a cloud of interacting paths evolving over time as an expected signature, it can be merged with other channels, and the process repeated. PDE kernels can manage dimension. The crop yield prediction task matches this model: AISTATS 2021 arxiv.org/pdf/2006.05805.pdf

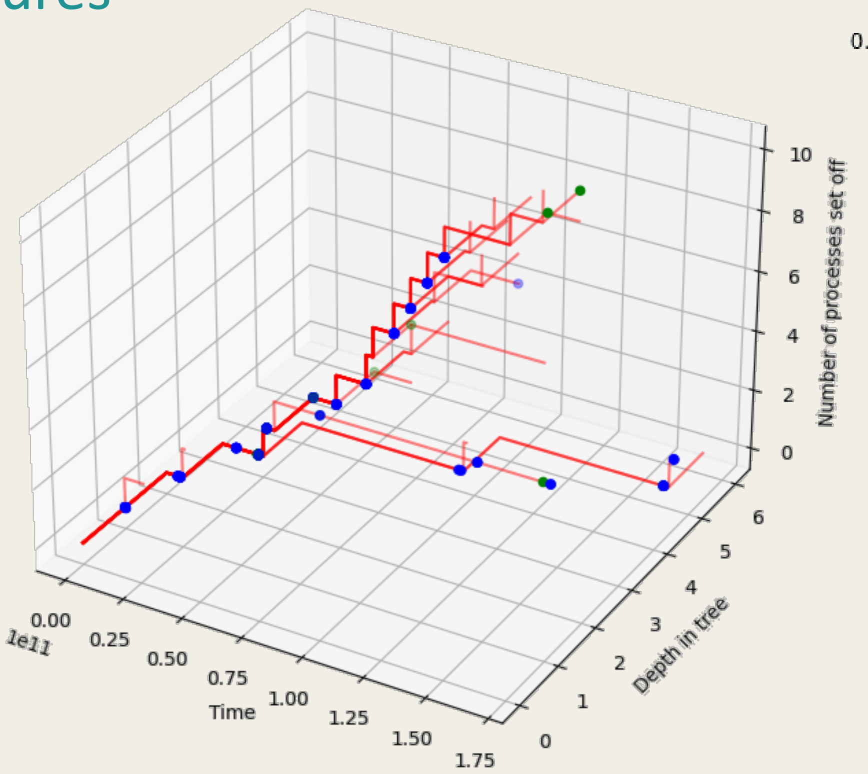
The climatic measurements (temperature, soil humidity and precipitation) are extracted from the GLDAS database (Rodell et al, 2004), are recorded every 6 hours at a spatial resolution of 0.25° × 0.25°, and their number varies across regions. Add regional policy information, etc.

Model	MSE	MAPE
Baseline	2.38	23.31
DeepSets	2.67	22.88
DR-RBF	.82	13.18
DR-Matern32	.82	13.18
DR-GA	.72	12.55
KES	**.65**	**12.34**
SES	**.62**	**10.98**

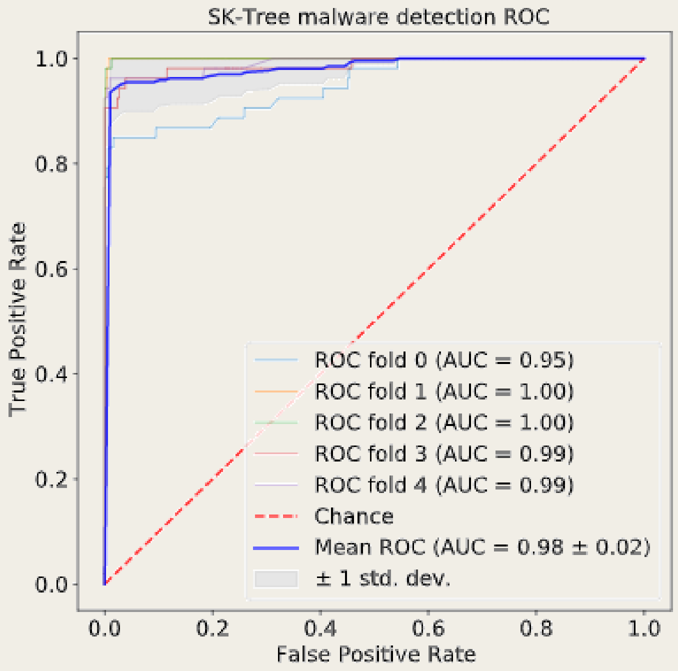
Process tree example:
Expected signatures of clouds of paths

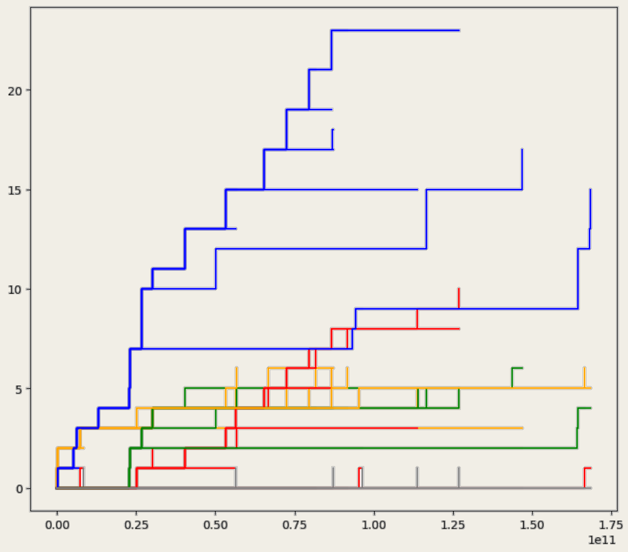
Developed SK-tree structure to apply standardised expected signature techniques to host-based event logs, by viewing processes as trees evolving over time analysed as expected signatures through a PDE kernel.
[2102.07904.pdf (arxiv.org)](2102.07904.pdf)

We demonstrate the SK-Tree to detect malicious events on a portion of the publicly available DARPA OpTC dataset, achieving an initial AUROC score of 98% for a supervised question.



ROC evaluation of the SK-Tree binary classifier on the OpTC data





Communication

DataSig

A rough path between
mathematics and data science

Landmark-based action recognition

To communicate our methodology, and aside from our papers, with their software we are constructing notebooks with introductory examples of what we can do.

People moving can easily be anonymized to landmarks. It is a static process. The moving stick people still contain information.

Peter Foster has put together a simple notebook you can run that demonstrates viable approaches to recognizing these actions that can be trained on small datasets.

https://www.datasig.ac.uk/examples

Thank you!

$$dy = f(y)d\gamma$$