SHAPELETS FOR SIMULATION

DYNAMIC TRAJECTORY ANALYSIS

AUTHORS

Graham Laidler[1],
Dr. Nicos Pavlidis[1],
Prof. Barry Nelson[2]

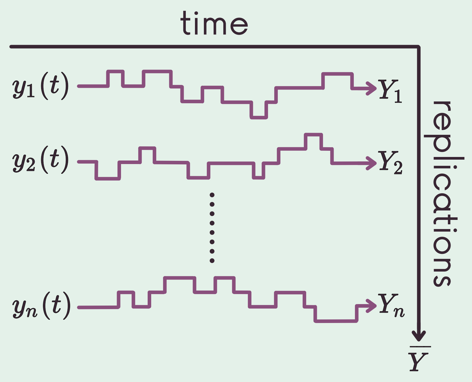
AFFILIATIONS

[1]Lancaster University
[2]Northwestern University

Introduction

We consider discrete-event simulation over a finite time horizon.

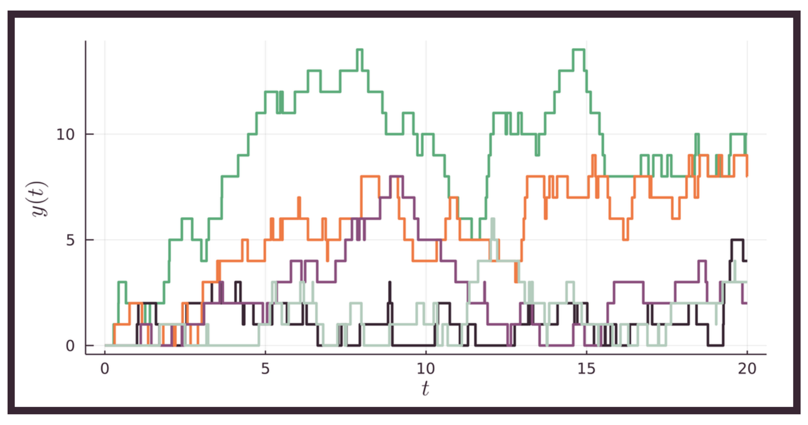
Simulation is an inherently dynamic process, although output analysis traditionally focuses on long-run performance. This averages out the dynamic behaviour over time and replications.



Instead, we can look at the full trajectories of a performance measure from each replication.

$$y : [0, T] \to \mathbb{R}$$

We assume these to be piecewise constant functions of simulation time.



E.g. queue size trajectories from 5 replications of an M/M/1 queue

Objective

Our objective is to unlock deeper, time-dynamic analyses from the underlying trajectories of simulation output.

Simulation is often used to compare alternative system designs. When alternatives yield similar long-run behaviour, looking deeper into the dynamics and short-term behaviours can be useful.

Set-up: Suppose we have at least two systems to compare and a performance measure of interest. We simulate multiple replications of each system and build up a dataset containing trajectories of the performance measure classified by system.

Methodology

What is a shapelet?

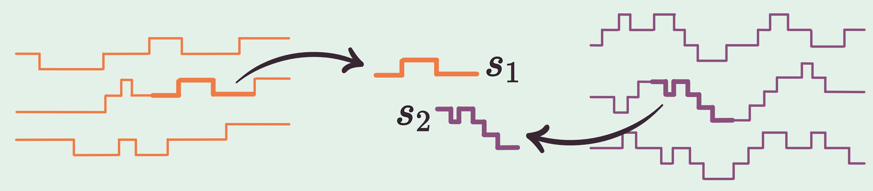
Shapelets were introduced by Ye and Keogh (2011) as a method for time series classification.

A shapelet is a subsequence of a time series which we identify as being representative of its class. This understands that the characteristic behaviour of a class of series often occurs over local patterns rather than on the global structure.

Shapelets offer visual interpretability as to the dynamic behaviour which characterises and discriminates among classes.

Instead of applying shapelets in their original discrete-time setting, we adapt the methodology to continuous-time, piecewise constant simulation trajectories.

$$s : [0, l] \to \mathbb{R},\ l \leq T$$



Finding the best shapelet

Assessing the ability of a shapelet to discriminate among classes relies on a notion of distance between shapelets and trajectories:

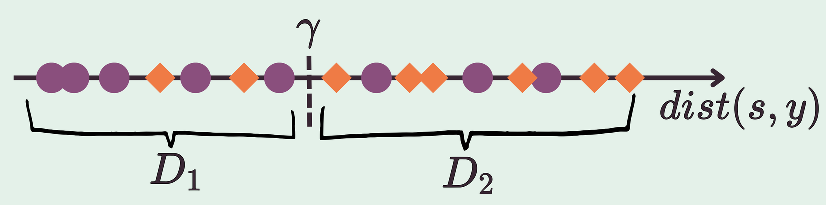
$$dist(s, y) = \min_{t \in [0, T-l], c \in \mathbb{R}} \int_t^{t+l} |s(u-t) + c - y(u)| du.$$

This integral calculates the area between a shapelet, s, of length l, and an equal-length segment of y. The shapelet is allowed free movement horizontally and vertically to minimise this area, so that $dist(s, y)$ measures the closest appearance of the shape of s occurring somewhere in y.

The piecewise constant structure of s and y makes $dist(s, y)$ an easy minimisation.

A shapelet, s, splits a dataset of trajectories, D, via some distance threshold, γ:

$$D_1 = \{y_i \in D \colon dist(s, y_i) \leq \gamma\}$$
$$D_2 = \{y_i \in D \colon dist(s, y_i) > \gamma\}$$



We look for shapelets which can best separate the classes (systems) between D_1 and D_2 using some threshold γ. This is measured by entropy and information gain. The entropy of the dataset D containing n total trajectories, n_c of class $c \in C$, is

$$H(D) = -\sum_{c \in C} \frac{n_c}{n} \log\left(\frac{n_c}{n}\right).$$

The information gain from using the shapelet s with the distance threshold γ is then

$$I(D, s, \gamma) = H(D) - \left(\frac{|D_1|}{n} H(D_1) + \frac{|D_2|}{n} H(D_2)\right).$$

We generate a finite collection of shapelet candidates by extracting segments of meaningful lengths from the trajectories in D. For each candidate s, we can compute

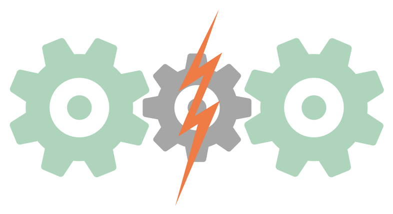
$$\max_{\gamma \in \mathbb{R}} I(D, s, \gamma).$$

The shapelets that maximise this value are considered the most discriminative, and the best able to characterise the dynamic differences among the systems.

Recovery from Disruption

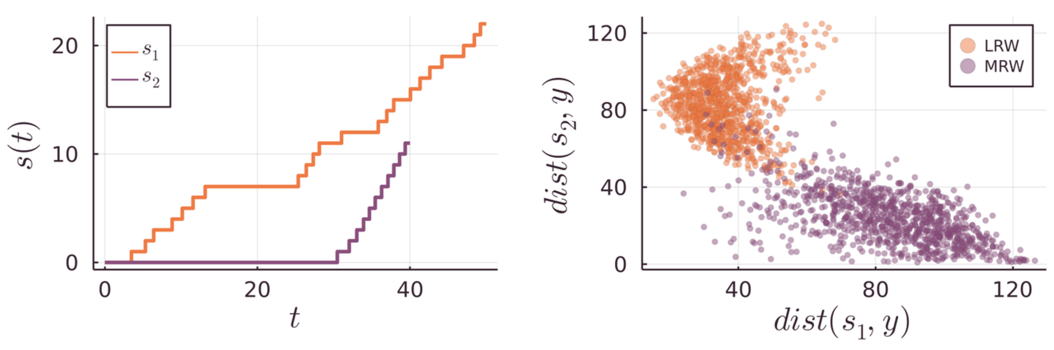
Supply chain and manufacturing systems often experience unexpected disruptions. A system's resilience and recovery from such events is often an important consideration.

We consider a simulation model of a factory, in which jobs follow a processing sequence through a number of stations. The machine at one of the stations occasionally fails, and waiting jobs are held up until the machine is repaired.



We want to understand how the throughput trajectory is impacted by these breakdowns under two system alternatives. The jobs queueing at each station are prioritised by either a 'Least Remaining Work' (LRW) or 'Most Remaining Work' (MRW) rule.

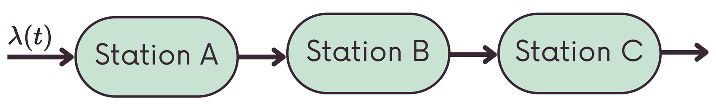
We use week long replications, with injected breakdowns lasting 16 hours. Both systems have similar weekly average throughput. Simulating with common random numbers ensures that the differences between the trajectories of each system reliably reflect the effects of the priority rules.



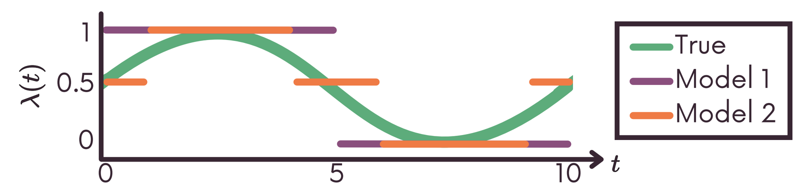
Machine failures cause throughput to be paused for much longer under the MRW rule. However, once it resumes, its rate is consistent. Under LRW, althoughthe throughput restarts more quickly, its rate recovers more gradually.

Dynamic Model Validation

Suppose we have a real-world system and its simulation model. Can we use shapelets to validate the model based on its dynamic behaviour? For a good model, we should struggle to find characteristic shapelets which distinguish it from the true system.

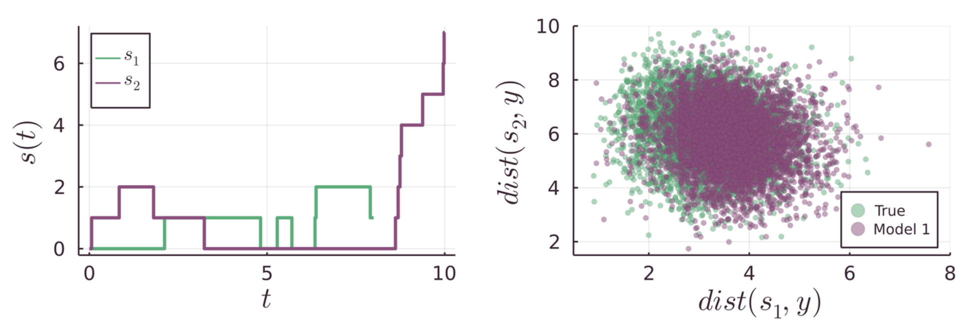


We consider a 3-station tandem queueing model with a single server, infinite queueing space, and exponential service times with rate $\mu = 0.6$ at each station. We assume that the true system experiences Poisson arrivals which follow a sinusoidal rate function with a period of 10 minutes. We consider two modelling attempts:



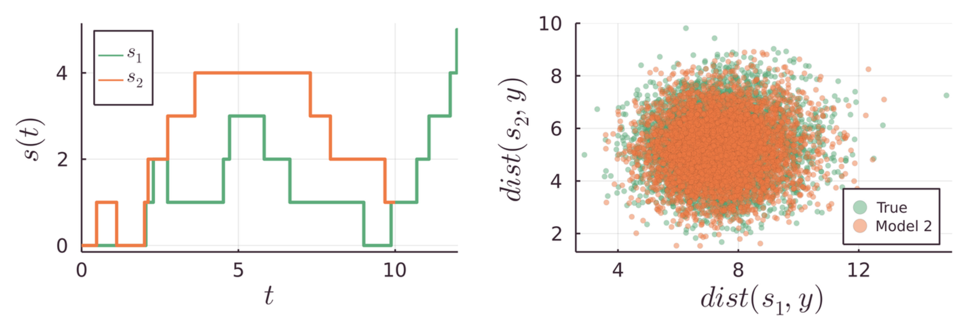
We aim to validate the models based on the dynamic behaviour of their number-in-system trajectories. The mean number-in-system values are all similar. We take replications of length 60 minutes, and look for the shapelets that are best able to discriminate the true system from each of the models.

True system vs Model 1



Model 1's number-in-system behaviour suggests longer periods without arrivals followed by sudden bursts, whereas the true system remains more stable.

True system vs Model 2



The scatter plot of test trajectories suggests no clear difference between the dynamic number-in-system behaviours of the true system and model 2.

Conclusion

Shapelets provide a visually interpretable way to characterise and compare alternative system designs based on their dynamic behaviour.

We show an example of shapelet analysis being used for dynamic model validation, as well as to compare the typical dynamic responses to a specific system event under two design alternatives.

RELATED LITERATURE

Ye and Keogh (2011). "Time Series Shapelets: A Novel Technique that Allows Accurate, Interpretable and Fast Classification", in Data Mining and Knowledge Discovery 22, pg 149 - 182

STOR-i | Lancaster University | UKRI Engineering and Physical Sciences Research Council