We thank all our reviewers for their time and for their kind words. We are delighted to receive such uniformly positive reviews.

Reviewer 1 comments that the experiments could be on larger problems – we agree that this is fair, and this is certainly of interest going forward.

Reviewer 2 comments that some details are unclear. Apologies, to clarify, the loss function was binary cross entropy for the PhysioNet dataset, and the usual cross entropy for CharacterTrajectories and Speech Commands. Training was via SGD in the usual manner. These details were put in the appendix for space; we will add them to the main paper. As you suggest, the model implements a feedforward step analogous to an RNN. An additional reconstruction step analogous to a VAE would be possible if desired, but wasn't something we explored here.

On the number of function evaluations (NFE), leaving this out was an oversight. Running the experiments with `method='dopri5', rtol=1e-4, atol=1e-6` we observe:

	Speech Commands	Sepsis (OI)	Sepsis (No OI)	CharacterTrajectories 30%
Neural CDE	**6009 $\pm$ 494**	2872 $\pm$ 158	3123 $\pm$ 223	**1955 $\pm$ 233**
GRU-ODE	8579 $\pm$ 1355	3175 $\pm$ 194	3304 $\pm$ 178	1974 $\pm$ 110
ODE-RNN	6399 $\pm$ 2	**2472 $\pm$ 127**	**2681 $\pm$ 47**	7242 $\pm$ 6

CharacterTrajectories 50%, 70% are similar to 30% and so are not shown. The high mean and low variance of the NFEs for ODE-RNN on some problems is likely due to the high sampling rate of the data.

Regarding error tolerances, we have observed successful training, in the sense of high classification accuracies, across a range of tolerances ranging from 10^{-3} down to 10^{-8}, even for oscillatory problems such as the Speech Commands dataset. Generally speaking Neural CDEs seem to be slightly harder (but not dramatically harder) to solve than a comparable Neural ODE, no doubt due to the time-varying nature of the problem.

Regarding stability of training, Neural CDEs seem to be unusually stable. We found that they always seem to train well, regardless of the choice of optimiser and learning rate, even when competing methods (GRUs etc.) are prone to fail. We make a brief remark about this in the paper, but frankly this isn't a phenomenen we understand yet.

Reviewer 3 asks about the smoothness prior of splines. Splines can actually represent very oscillatory data without issues. See the figure to the right, which is a cubic spline through one channel of a sample from the Speech Commands dataset (and it is a typical such sample). Integrating this simply requires putting down enough points, and moreover this is not computationally daunting.
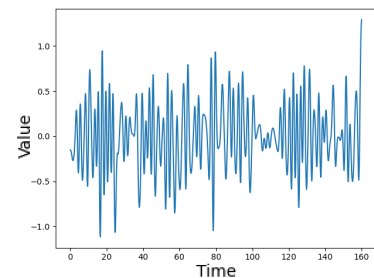
We have indeed experimented with other interpolation schemes. We find that performance differences are minor; trained models have similar classification accuracies. We do however observe that linear interpolation produces a model that requires fewer function evalutions on the forward pass, but increased function evaluations on the backward pass. Furthermore linear interpolation is causal whilst cubic splines are non-causal. This is a topic we are performing follow-up work on presently.

Regarding the correctness of the experiments, the other models received the same information (in particular the observational intensity) that the Neural CDE did – we appreciate that to have done otherwise would have been unfair. The paper does word this a little oddly (essentially we sought to emphasise the difference between X and $\mathrm{d}X/\mathrm{d}t$), so we will ensure that it is clear on this point.

Regarding the choice of only considering classification tasks – there was no real reason for this choice, and we expect Neural CDEs to be able to perform regression, forecasting, etc. as well. We completely agree it would be valuable to consider other problems; were we to do this again (with the benefit of hindsight) then we probably would. Each experiment was chosen to demonstrate a particular type of problem: CharacterTrajectories for varying irregular data; PhysioNet for observational intensity and partially observed data; Speech Commands for regular data.

Reviewer 4 gives a truly delightful review – thank you! We are very happy that our paper is identified as making both significant theoretical and practical contributions. Regarding the brevity of the proof descriptions, we will aim to use part of the additional page to give more detail of these.

On the choice of datasets, these were intended to be standard choices. CharacterTrajectories is drawn from the UEA archive, PhysioNet is a central resource for medical time series, and Speech Commands is now provided through `torchaudio`. We agree that a sample from the datasets could be informative to fix what is going on in a reader's mind; if space allows we will add this to the paper. Thankyou for the suggestions of other speech problems – we're always looking for good ideas for follow up work.