Capturing Actionable Dynamics with Structured Latent Ordinary Differential Equations (Supplementary material)

Paidamoyo Chapfuwa[1] **Sherri Rose**[1] **Lawrence Carin**[2] **Edward Meeds**[3] **Ricardo Henao**[4]

[1]Stanford University, USA
[2]KAUST, Saudi Arabia
[3]Microsoft Research, Cambridge, UK
[4]Duke University, USA

A ADDITIONAL RESULTS

Figure 1 and Figures 3-6 provide all qualitative visualizations of the posterior predictive distributions across all methods on SYNTHETIC BIOLOGY and HUMAN VIRAL CHALLENGE datasets. Note that for fair comparisons, Hierarchical-ODE preserves the data generating graphical model of Roeder et al. (2019) but deviate in dynamics and emission functions, resulting in significantly worse performance than reported in Roeder et al. (2019). Additionally, we present results from held-out device posterior predictive distribution and controlled generated observations from novel device $\boldsymbol{g} = R33\text{-}S32$ in Figure 2. See Table 2 for CARDIOVASCULAR SYSTEM quantitative results.

B EXPERIMENTAL SETUP

Below we provide details of the neural-network architectures, selected hyper-parameters and pseudo-code for the proposed SL-ODE algorithm.

B.1 NEURAL-NETWORK ARCHITECTURES

In all experiments, SL-ODE (proposed), GOKU-Net, Latent-ODE, and Hierarchical-ODE share the ODE $f(\cdot)$, emission $m(\cdot)$, and encoder (maps observations $\boldsymbol{y}(t)$ to latent $\boldsymbol{z}$) functions, detailed below. In general, we specify two-layer multilayer perceptrons (MLPs) with 25 hidden units and Rectified Linear Unit (ReLU) as activation functions. Additionally, we implement 2-layer MLPs for the *system input-specific* distributions:

- Prior distribution $p_{\boldsymbol{\psi}}(\boldsymbol{z_u}|\boldsymbol{u})$ used in SL-ODE and Hierarchical-ODE.
- Variational distribution $q_{\boldsymbol{\varphi}}(\boldsymbol{u}|\boldsymbol{z_u})$ used in SL-ODE and GOKU-Net.

Encoder Following Roeder et al. (2019), we apply a $1D$ CNN to observations $\boldsymbol{y}(t) \rightarrow$ average pooling $\rightarrow$ two-layer MLPs $\rightarrow$ latent variable $\boldsymbol{z}$ described with mean $\boldsymbol{\mu}$ and variance $\text{diag}(\boldsymbol{\sigma}^2)$. Note that the Hierarchical ODE model has an additional 2-layer MLP mapping system inputs to an input-specific latent variable.

Black-box Dynamics We leverage the *adjoint solver* Chen et al. (2018) to simulate the state-time matrix X where the dynamics $f_{\boldsymbol{\theta}}(\cdot)$ are 2-layer MLPs with *Sigmoid* output-layer activations. Following Roeder et al. (2019), we specify dynamics as

$$\frac{d\boldsymbol{x}}{dt} = f_1(\boldsymbol{x}, \boldsymbol{z}, t; \theta) - \boldsymbol{x} \odot f_2(\boldsymbol{x}, \boldsymbol{z}, t; \theta)\,,$$

where $\odot$ is the Hadamard product. Further, we initialize the initial state $\boldsymbol{x}_0$ as $\boldsymbol{z} \rightarrow$ 2-layer MLPs with *Sigmoid* output activation $\rightarrow \boldsymbol{x}_0$.

Emission We map the states X to the observations Y with a 1-layer linear MLP. For all baseline methods, the emission function outputs observation means $\boldsymbol{m}(t)$ and variances $\boldsymbol{\epsilon(t)}$. In contrast, our proposed approach (SL-ODE), outputs the median $\boldsymbol{m}(t)$, upper- $\boldsymbol{u}(t)$, and lower- $\boldsymbol{l}(t)$ quantiles according to the specified τ.

B.2 HYPER-PARAMETER SELECTION

We use the Adam optimizer (Kingma and Ba, 2015) with the following hyper-parameters: first moment 0.9, second moment 0.99, and epsilon 1×10^{-8}. We train all models using one NVIDIA P100 GPU with 16GB memory. See Table 1 for data-specific hyper-parameters. We split the CARDIOVASCULAR SYSTEM data into training, validation, and test sets as 80%, 10%, and 10% partitions, respectively. Further, we use the validation set for early stopping and learning model hyper-parameters. However, for the SYNTHETIC BIOLOGY and HUMAN VIRAL CHALLENGE datasets, we perform k-fold cross-validation due to the small sample sizes.

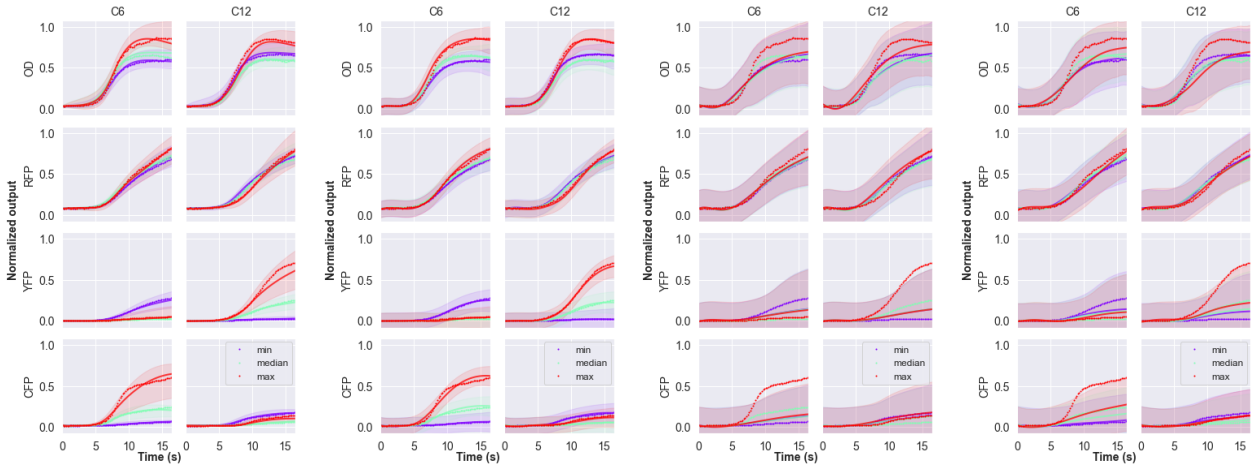
Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

Table 1: Summary of data-specific hyper-parameters.

Hyper-parameter	SYNTHETIC BIOLOGY	CARDIOVASCULAR SYSTEM	HUMAN VIRAL CHALLENGE
Mini-batch size	36	128	28
Learning rate	3×10^{-4}	1×10^{-3}	1×10^{-3}
States dimension (D)	8	5	5

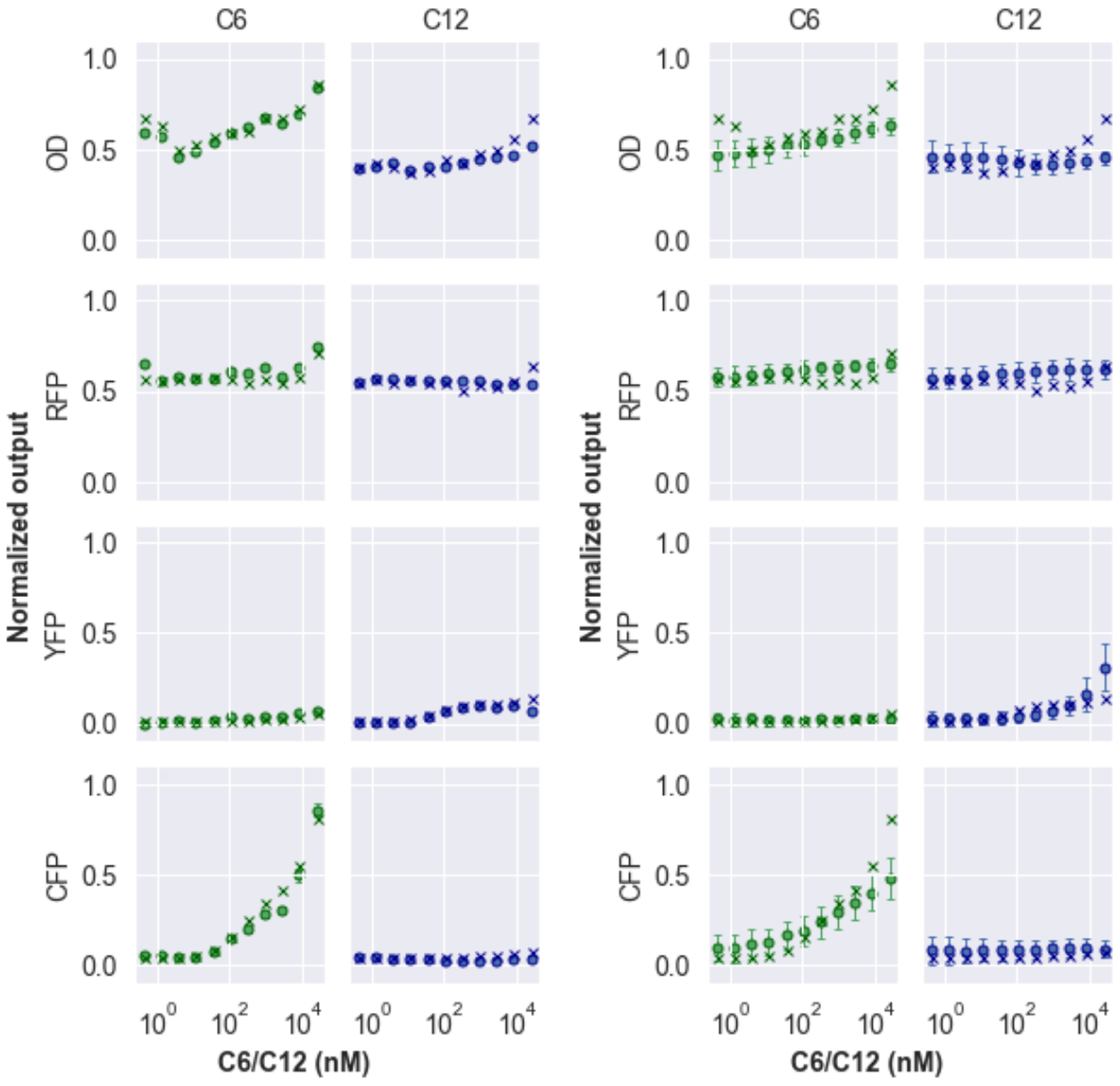
Table 2: Performance comparisons for CARDIOVASCULAR SYSTEM on test data. System inputs $\boldsymbol{u}$ are interpretable patient states. We report methods without system input inference or controlled prior generation mechanisms as NA.

Method	$\boldsymbol{u}$ Accuracy (%) ↑	L_1 error (posterior, prior) ↓	ELBO ↑
Latent-ODE	NA	(6.95, NA)	9.12
GOKU-Net	**100**	(5.06, NA)	324.81
Hierarchical-ODE	NA	(4.25, 4.42)	374.94
SL-ODE-Gaussian (ablation)	**100**	(0.66, 0.67)	561.29
SL-ODE (proposed)	**100**	(**0.56**, **0.57**)	**752.23**



(a) SL-ODE (b) GOKU-Net (c) Latent-ODE (d) Hierarchical-ODE

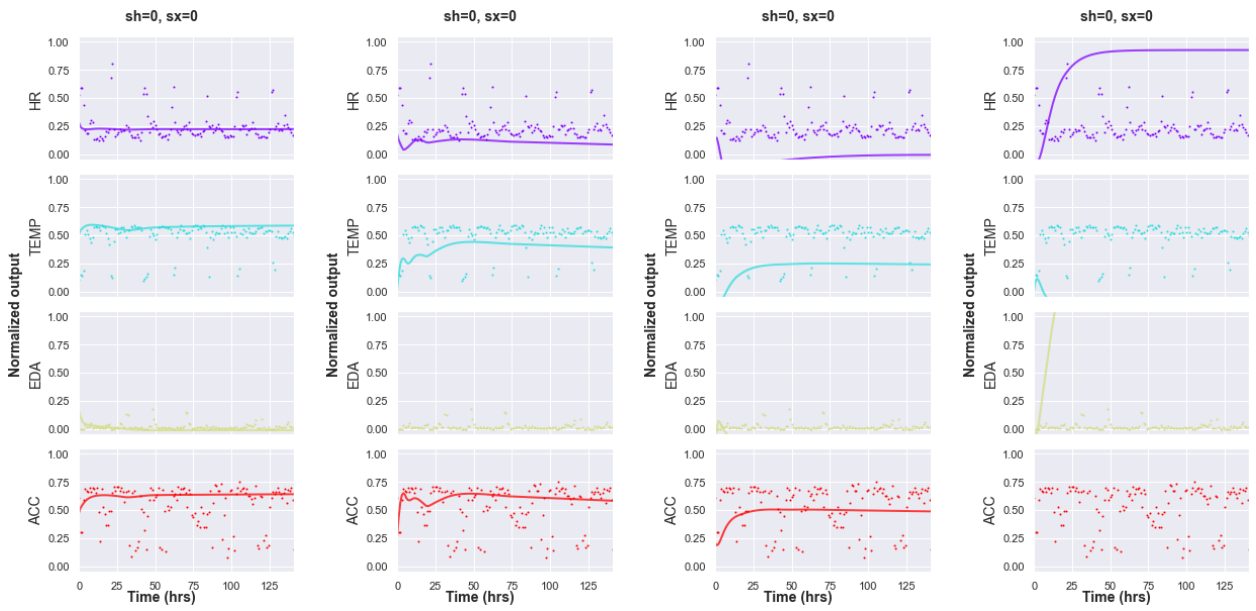
Figure 1: Posterior predictive distribution on SYNTHETIC BIOLOGY data via 4-fold cross-validation *multiple device* inference task for (a) proposed SL-ODE, (b) GOKU-Net, (c) Latent-ODE, and (d) Hierarchical-ODE models. For clarity, we plot ground truth (dotted) time-series against median predictions (solid) across three $\boldsymbol{c} = [C_6, C_{12}]$ treatments (minimum, median, and maximum), *e.g.*, when C_6= minimum, output is averaged across all C_{12}. Shaded areas indicate the predicted 95% confidence interval (CI).



(a) SL-ODE: Posterior

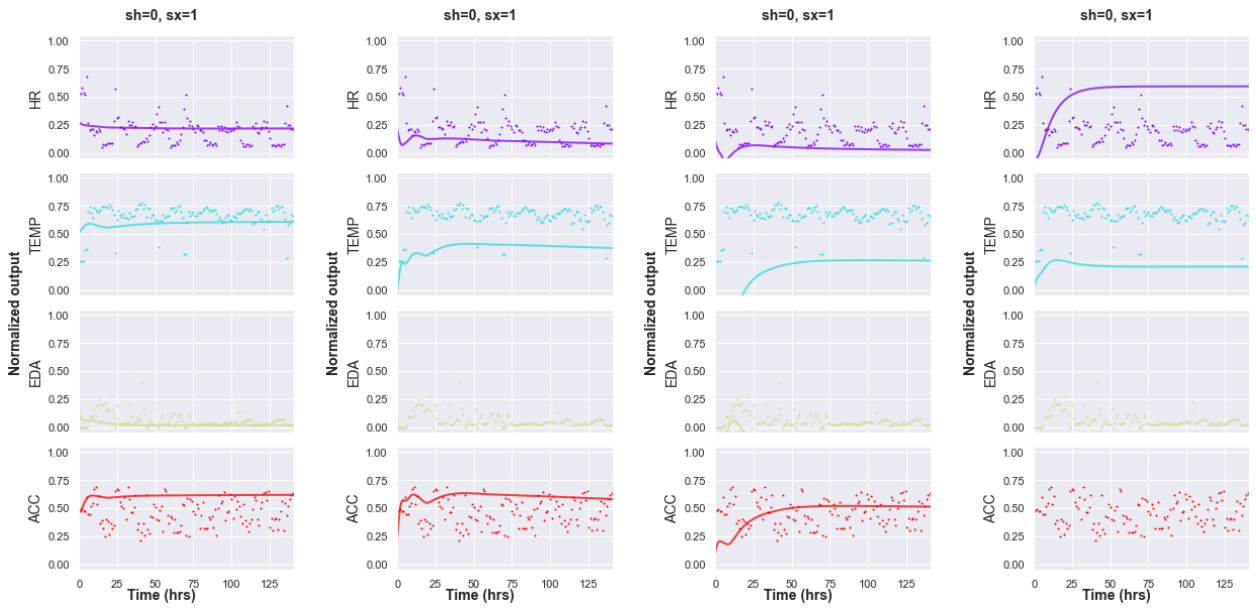
(b) SL-ODE: Prior

Figure 2: SL-ODE SYNTHETIC BIOLOGY *held-out device* ($\boldsymbol{g} = R33\text{-}S32$) task. Ground truth *vs.* (a) posterior predictive distribution and (b) *controlled* generated observations given system inputs $\boldsymbol{u} = [\boldsymbol{g}, \boldsymbol{c}]$ according to assumed prior distribution. We plot the median (circles) with 95% CI against ground truth observations (crosses) averaged (200 $\boldsymbol{z}$ samples) across all observations at the final time-point sweeping all $\boldsymbol{c} = [C_6, C_{12}]$ treatments.



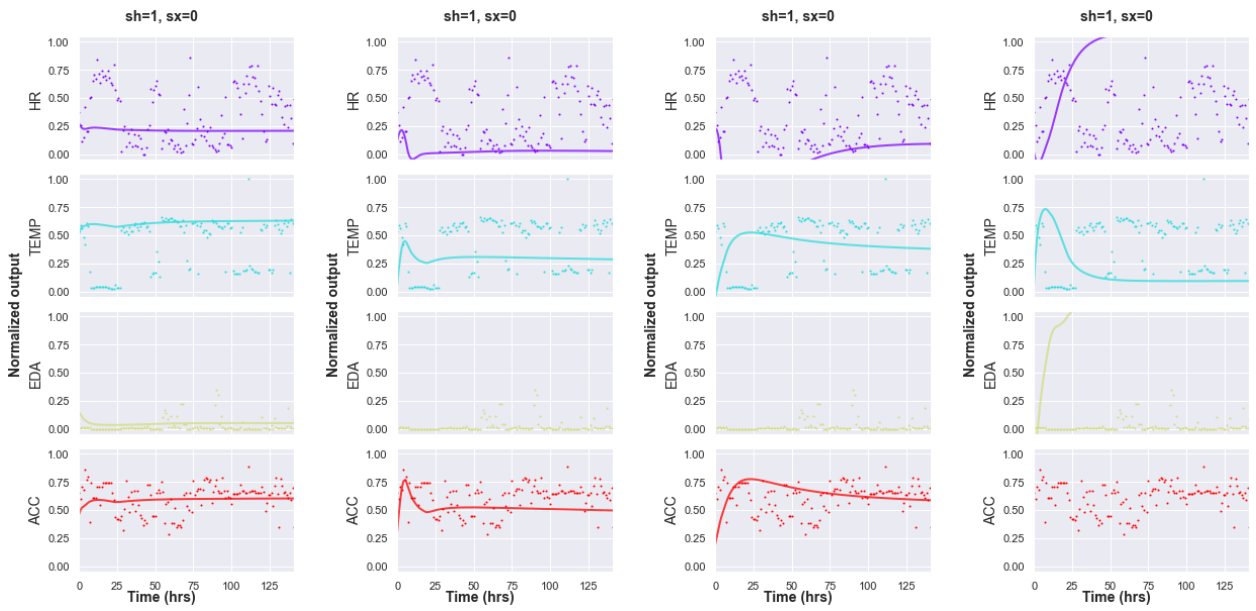
(a) SL-ODE (b) GOKU-Net (c) Latent-ODE (d) Hierarchical-ODE

Figure 3: Posterior predictive distribution on HUMAN VIRAL CHALLENGE for randomly selected test patient showing one of the four combination binary outcomes $\boldsymbol{u}$ for viral shedding (sh=0) and symptoms (sx=0) onset (a) proposed SL-ODE, (b) GOKU-Net, (c) Latent-ODE, and (d) Hierarchical-ODE models. For clarity, we plot ground truth (dotted) time-series against median predictions (solid). We do not show error bars since they are too large due to noisy data.



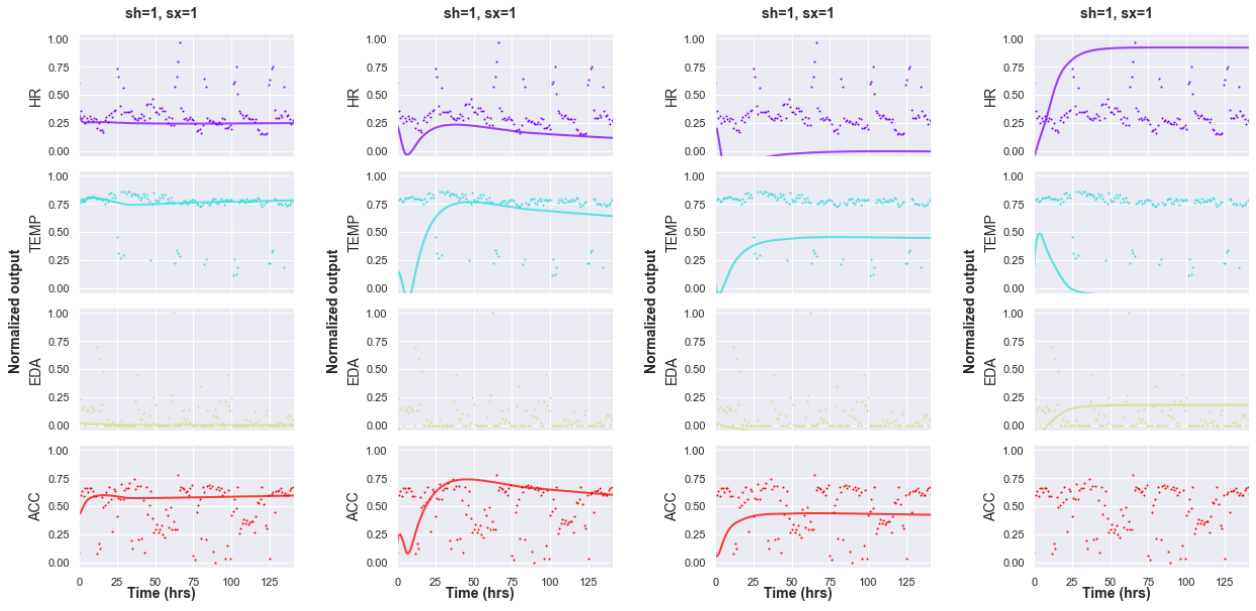
(a) SL-ODE (b) GOKU-Net (c) Latent-ODE (d) Hierarchical-ODE

Figure 4: Posterior predictive distribution on HUMAN VIRAL CHALLENGE for randomly selected test patient showing one of the four combination binary outcomes $\boldsymbol{u}$ for viral shedding (sh=0) and symptoms (sx=1) onset (a) proposed SL-ODE, (b) GOKU-Net, (c) Latent-ODE, and (d) Hierarchical-ODE models. For clarity, we plot ground truth (dotted) time-series against median predictions (solid). We do not show error bars since they are too large due to noisy data.



(a) SL-ODE (b) GOKU-Net (c) Latent-ODE (d) Hierarchical-ODE

Figure 5: Posterior predictive distribution on HUMAN VIRAL CHALLENGE for randomly selected test patient showing one of the four combination binary outcomes $\boldsymbol{u}$ for viral shedding (sh=1) and symptoms (sx=0) onset (a) proposed SL-ODE, (b) GOKU-Net, (c) Latent-ODE, and (d) Hierarchical-ODE models. For clarity, we plot ground truth (dotted) time-series against median predictions (solid). We do not show error bars since they are too large due to noisy data.



(a) SL-ODE (b) GOKU-Net (c) Latent-ODE (d) Hierarchical-ODE

Figure 6: Posterior predictive distribution on HUMAN VIRAL CHALLENGE for randomly selected test patient showing one of the four combination binary outcomes $\boldsymbol{u}$ for viral shedding (sh=1) and symptoms (sx=1) onset (a) proposed SL-ODE, (b) GOKU-Net, (c) Latent-ODE, and (d) Hierarchical-ODE models. For clarity, we plot ground truth (dotted) time-series against median predictions (solid). We do not show error bars since they are too large due to noisy data.

References

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

Geoffrey Roeder, Paul K Grant, Andrew Phillips, Neil Dalchau, and Edwards Meeds. Efficient amortised Bayesian inference for hierarchical and nonlinear dynamical systems. In *ICML*, 2019.