PULSE

Probabilistic Uncertainty-Aware Longitudinal Simulation for EHR Trajectories

Robert Lennart Manschke, Angus Roberts, Julia Ive

King’s College London and University College London · NeurIPS 2026

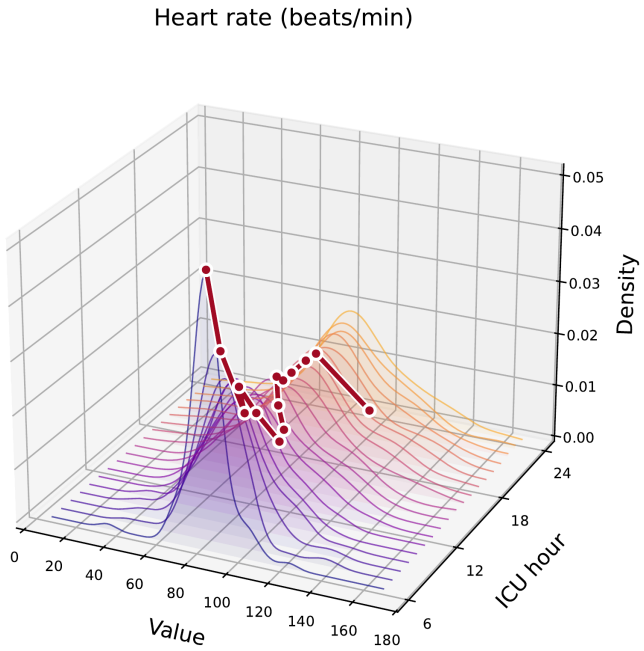
A five-minute walkthrough

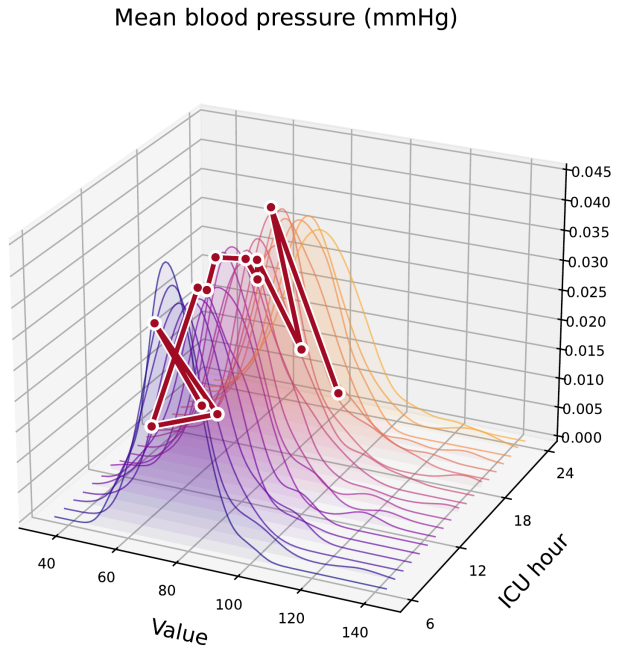
ICU futures are distributions

Patient state changes. Uncertainty changes too.

Each curve shows possible values at one future hour.

The model can represent a different shape and spread for each patient and time.





Public MIMIC-IV demo patient · red points mark recorded measurements

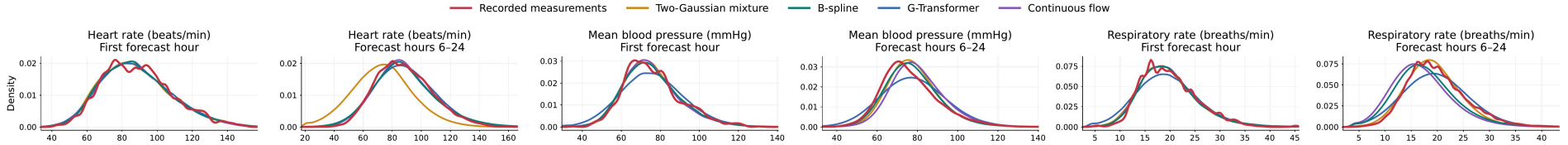
PULSE

Probabilistic Uncertainty-Aware Longitudinal Simulation for EHR Trajectories

Robert Lennart Manschke, Angus Roberts, Julia Ive
DRIVE-Health CDT, King's College London

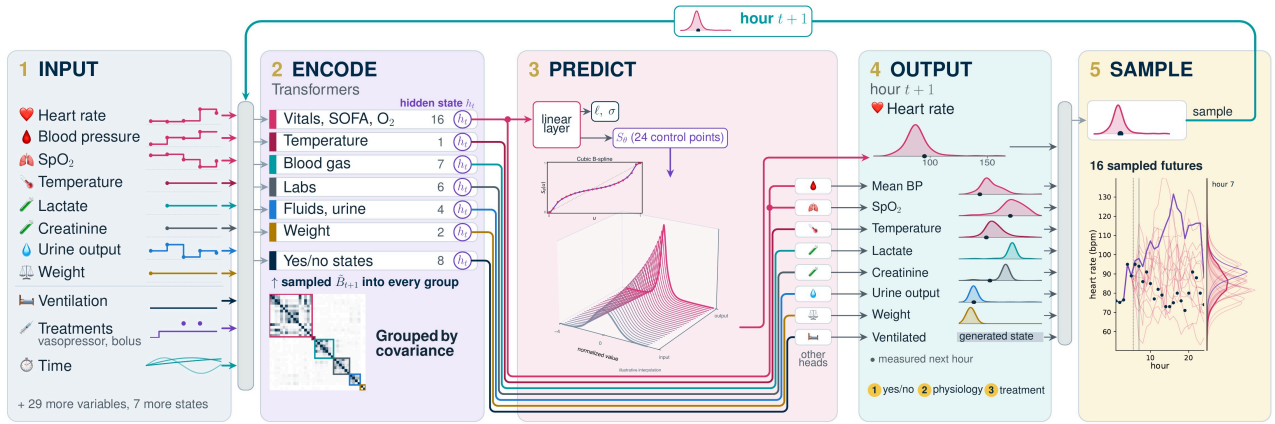
NeurIPS 2026
Main Track

1 Recorded and simulated distributions across patients



Smoothed distributions of recorded measurements and forecasts from hour 6 for the same held-out patient-hours. Right panels pool hours 6–24; left panels show the first generated hour. Only recorded values while in ICU are included; model curves average active paths per patient-hour. G-Transformer uses validation-estimated Gaussian residual noise.

2 How PULSE forecasts the next hour



Each hour, PULSE generates binary patient state B, physiology X and the next treatment A from history H_t:

$$\tilde{B}_{t+1} \sim p_\phi(B_{t+1} \mid H_t)$$
$$\tilde{X}_{t+1} \sim p_\theta(X_{t+1} \mid H_t, \tilde{B}_{t+1})$$
$$\tilde{A}_{t+1} \sim p_\psi(A_{t+1} \mid H_t, \tilde{B}_{t+1}, \tilde{X}_{t+1})$$

A monotone B-spline transforms a Gaussian sample into a patient-dependent distribution. The sampled next state joins the history used to predict later hours. The clinical model's backbone was initialized from a trained continuous flow, then fitted with the spline head.

Examples use a public MIMIC-IV demo patient. The central spline conditions on a sampled binary state; intermediate densities illustrate the transformation. The side curve averages 16 head densities before observation and event handling; faint curves show the individual densities.

3 Forecasting held-out ICU trajectories

MIMIC-IV SEPSIS COHORT / 100 PATHS

Forecasts share a treatment policy and use 37 continuous variables. Vector-normalized RMSE measures trajectory error; CRPS assesses the forecast distribution. Lower is better.

	$k=2$		$k=6$	
Model	VRMSE	CRPS	VRMSE	CRPS
Carry-forward	7.475	.6813	6.686	.5686
Population mean	**6.533**	.6213	6.522	.6366
G-Transformer	6.845	.5191	6.277	.4725
PULSE Gaussian	6.669	.4799	5.841	.4030
PULSE mixture	**6.592**	**.4599**	5.881	**.3935**
PULSE B-spline	**6.556**	.4625	**5.805**	**.3930**
PULSE continuous flow	**6.535**	**.4606**	**5.767**	**.3923**

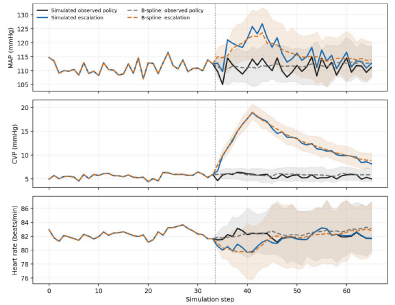
At $k=6$, B-spline minus G-Transformer: VRMSE -0.472 $[-0.584, -0.364]$; CRPS -0.0795 $[-0.0880, -0.0724]$. Brackets give paired 95% patient-bootstrap intervals. Bold marks the best and models whose paired difference from it includes zero; underlining marks the next tier by the same rule. G-Transformer is our reimplementation. Filled values and post-event padding are scored; $n = 892/891$ at $k = 2/6$.

4 Recover a known treatment response

A separate B-spline model was fine-tuned on synthetic cardiovascular trajectories, including escalation regimes.

Across 16 held-out synthetic patient pairs, mean arterial-pressure response was 5.34 mmHg for PULSE and 5.47 mmHg for the simulator; heart-rate response was -0.77 and -0.64 beats/min.

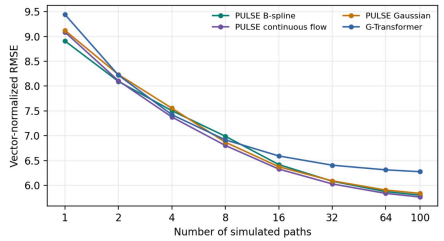
This illustration starts from a public MIMIC-IV demo patient. Dashed lines show model means; bands show the central 90% of 64 paths.



5 Patient-specific uncertainty

The B-spline output lets the shape and spread of each forecast change with patient history. Sampling these distributions produces different possible trajectories.

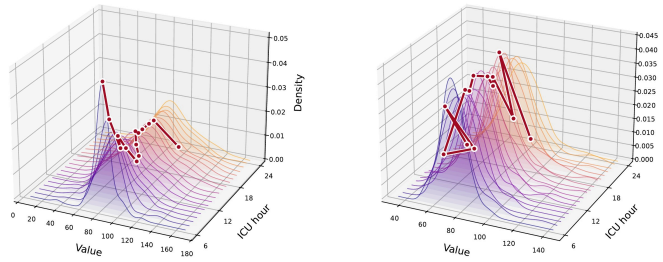
How many trajectories are needed?



Forecast from hour 6 on 891 held-out patients. Each point scores the mean of the first M saved paths using the same patients, policy and scoring rules. One fitted model and one path set are shown per curve.

Example: one public demo patient

Heart rate (beats/min) — Mean blood pressure (mmHg)

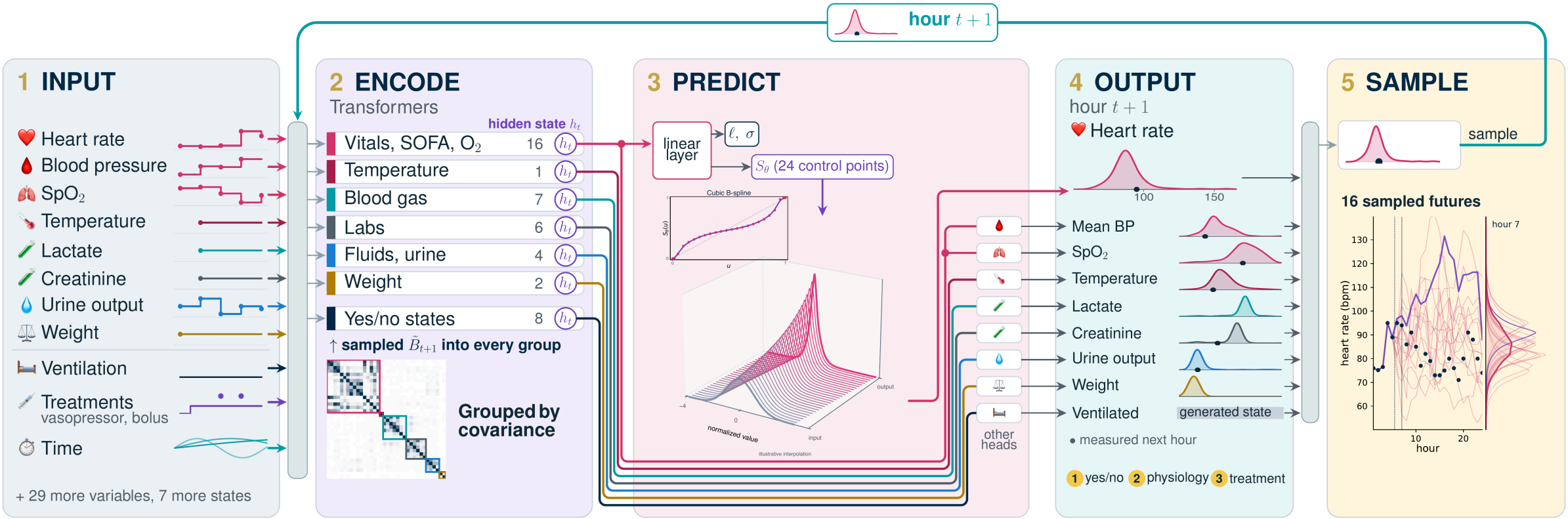


Each curve estimates a future-hour density from up to 100 B-spline paths that remain in the ICU. Red dots mark available demo-patient measurements at their predicted density; lines connect measured hours.

Robert Lennart Manschke
DRIVE-Health CDT | King's College London

This research work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [Grant reference number EP/Y035216/1] Centre for Doctoral Training in Data-Driven Health (DRIVE-Health) at King's College London.

How PULSE generates a patient trajectory



1 Encode history **2 Predict distributions** **3 Sample and repeat**

Inputs and output paths illustrated using public MIMIC-IV demo data

Forecasting held-out ICU trajectories

MIMIC-IV sepsis cohort · 100 paths · lower is better

Model	k = 2		k = 6	
	VRMSE	CRPS	VRMSE	CRPS
Carry-forward	7.475	.6813	6.686	.5686
Population mean	**6.533**	.6213	6.522	.6366
G-Transformer	6.845	.5191	6.277	.4725
PULSE Gaussian	6.669	.4799	5.841	.4030
PULSE mixture	**6.592**	**.4599**	5.881	**.3935**
PULSE B-spline	**6.556**	.4625	**5.805**	**.3930**
PULSE continuous flow	**6.535**	**.4606**	**5.767**	**.3923**

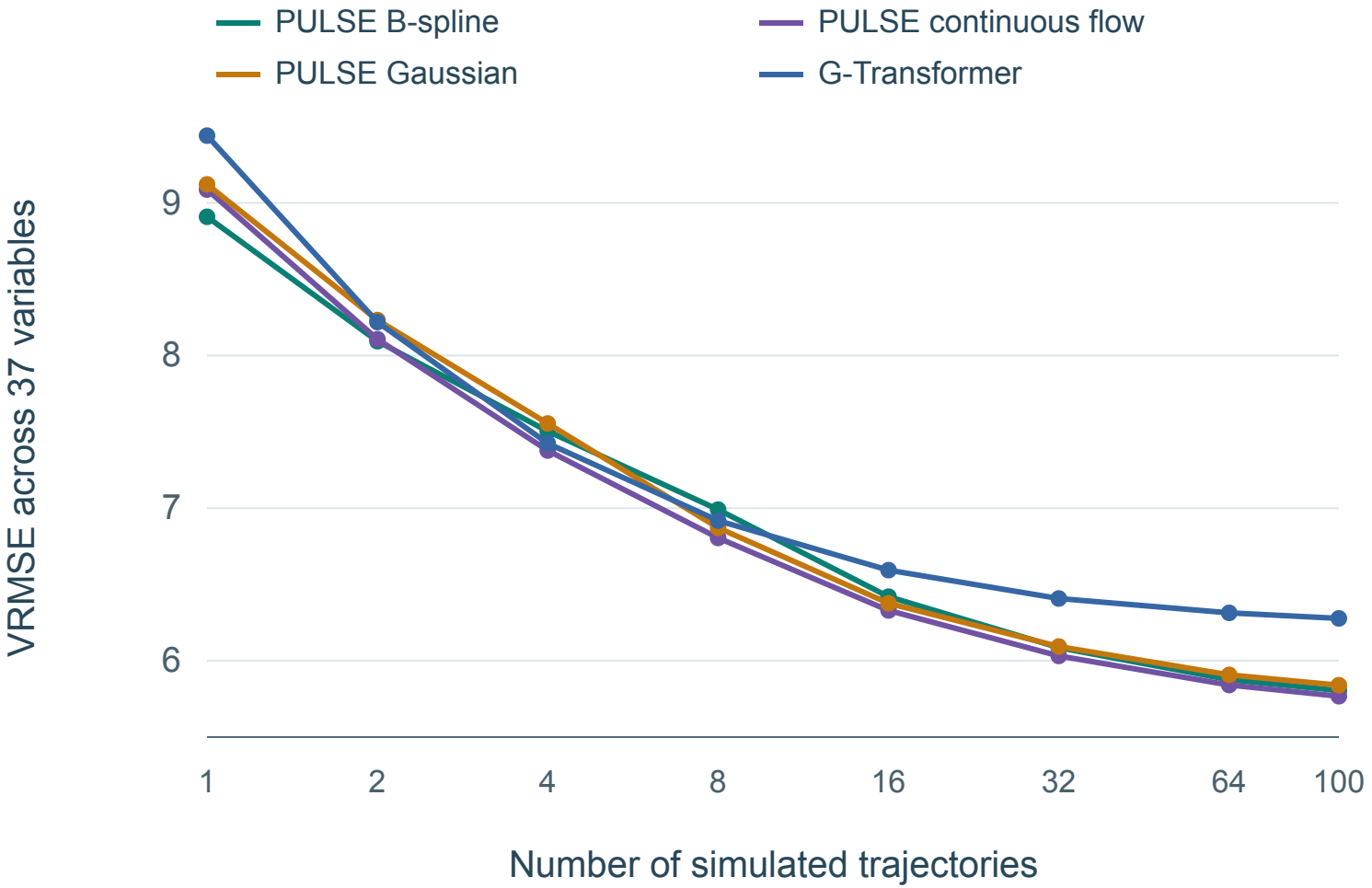
Bold: within the best tier Underlined: within the next tier

B-spline − G-Transformer at k = 6: VRMSE −0.472 [−0.584, −0.364] · CRPS −0.0795 [−0.0880, −0.0724]

Paired 95% patient-bootstrap intervals · G-Transformer reimplementation · filled and post-event padded targets

How many trajectories are needed?

Forecasts from hour 6 · 891 held-out patients · lower is better



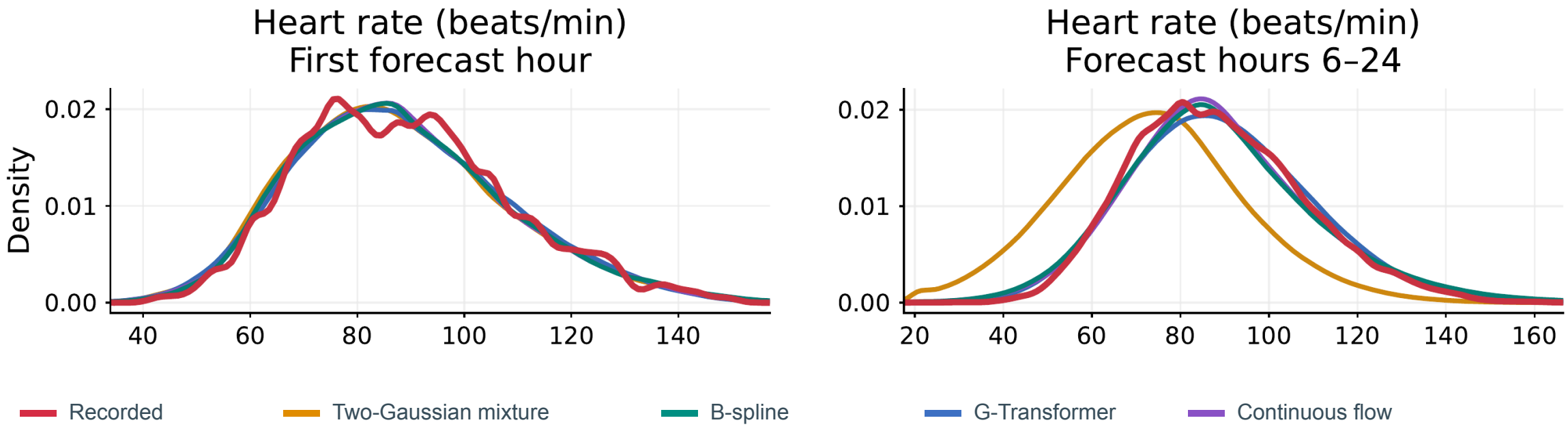
Fewer trajectories, strong forecast

By 32 trajectories, PULSE has lower error than G-Transformer with 100.

Same patients and treatment policy · nested subsets of 100 saved trajectories · one fit per model

Comparing forecast distributions with recorded data

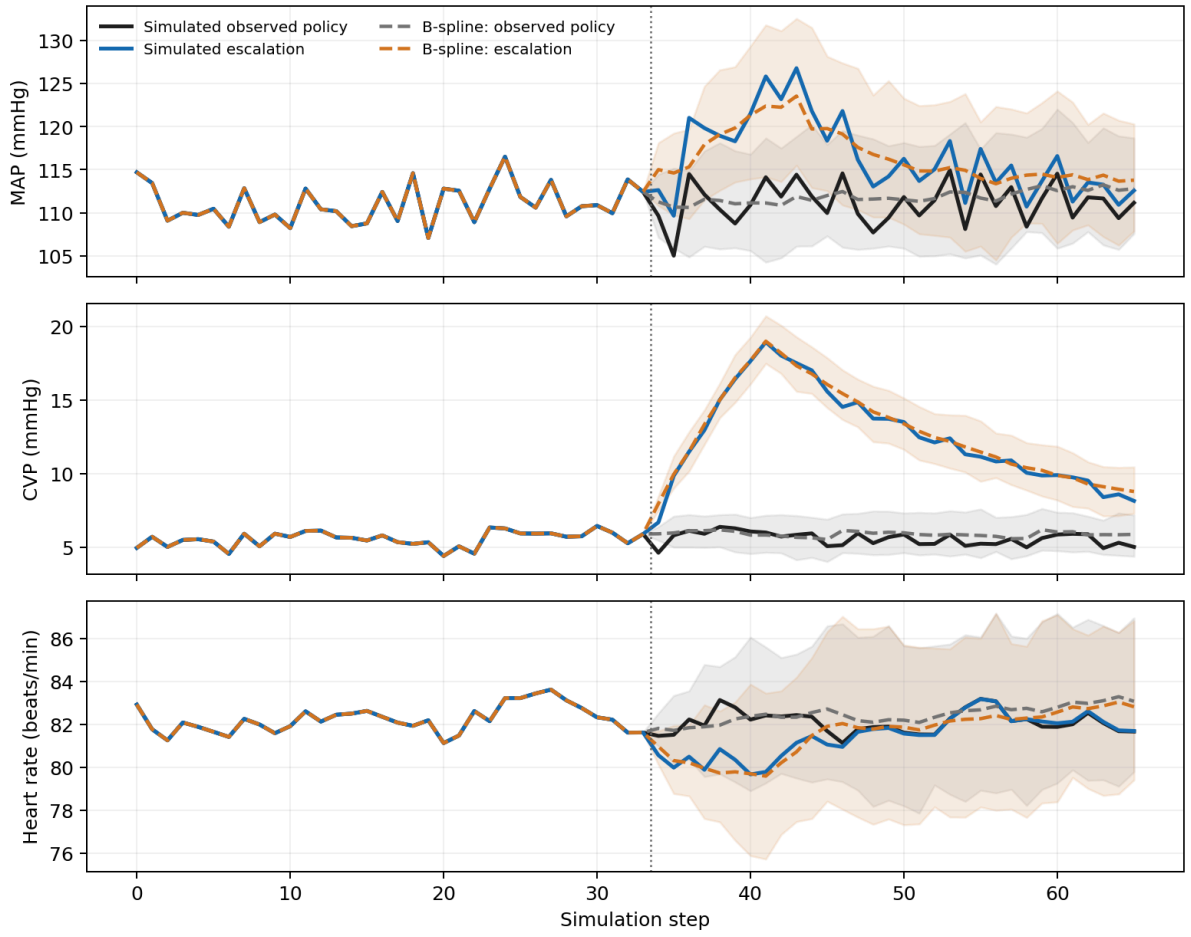
Heart rate



Forecast distributions can drift over repeated simulation.

MIMIC-IV held-out patient-hours · active simulated paths · recorded measurements only

Can PULSE recover a known treatment response?



Known intervention

A synthetic simulator supplies a reference treatment response.

Mean arterial pressure

5.34 vs 5.47 mmHg

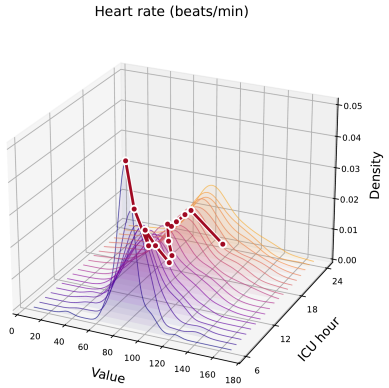
PULSE vs simulator
16 held-out synthetic patient pairs

Separate B-spline fit, fine-tuned on synthetic trajectories.

CVSim-inspired synthetic benchmark · example initialized from a public MIMIC-IV demo patient

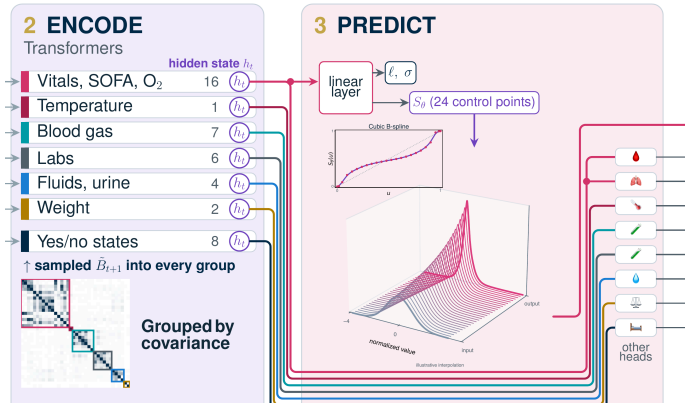
What PULSE contributes

Patient-specific uncertainty



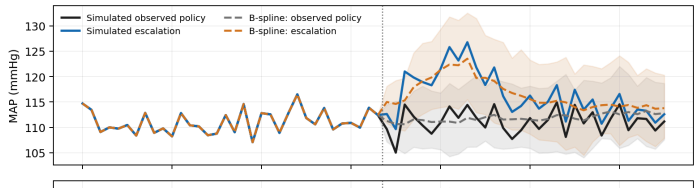
A distribution for each measurement, changing with patient history.

Multivariable trajectories



Covariance groups and sampled states connect the simulation over time.

Treatment-response research



A synthetic benchmark provides a known response to compare against.

Further external and clinical validation is needed.

Code: github.com/manscrober/pulse-artifact