STATE-CONDITIONED SPECTRAL PROPAGATION FOR CLOSED-FORM CONTINUOUS-TIME LIQUID NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Irregular observation sequences exhibit concurrent short-term fluctuations and long-term trends. A key challenge is to equip hidden states with multiple adaptive decay timescales while preserving closed-form propagation. We propose the State-Conditioned Spectral Closed-Form Continuous-Time Network (SC-SCfC), which models hidden-state evolution toward an interval target between consecutive observations as a mixture of multiple exponential decays. While individual state components can blend fast and slow decay rates, restricting decay to fixed coordinate axes limits the model's expressiveness. To overcome this, we further propose a state-conditioned matrix allocation (SC-MA) mechanism that extends SC-SCfC. By decomposing the residual between the current state and its target into vector components in the latent space using positive semi-definite allocation matrices that sum to the identity, each component can decay at its own rate along data-dependent directions. The parameters remain fixed within each interval, yielding closed-form propagation. Experiments across four benchmark families demonstrate competitive performance in classification and remaining-useful-life regression.

1 INTRODUCTION

With the spread of continuous monitoring in healthcare and industry, irregularly sampled sequences have become a common input to sequence learning models (Che et al., 2018; Kidger et al., 2020). Continuous-time neural networks, which parameterize the hidden state by elapsed time, are widely regarded as effective for such data (Chen et al., 2018; Rubanova et al., 2019; Hasani et al., 2021). In clinical monitoring and equipment degradation analysis in particular, these networks extend prediction to arbitrary times between observations, a setting where short-term fluctuations fade quickly while long-term trends persist. However, retaining history over multiple timescales between observations without compromising closed-form and bounded propagation remains challenging.

In practice, richer dynamics between observations usually raise computational cost sharply, as numerical ordinary differential equation (ODE) solvers make the per-interval cost depend on interval length and solver tolerance. Since the residual between the hidden state and the target it approaches carries the retained history (Hasani et al., 2021; 2022), preventing its single-timescale decay along fixed directions under closed-form propagation has become a pressing need.

Continuous-time state propagation offers a feasible way to retain and update this history across irregular observations. Existing methods fall into two categories, the first evolving the hidden state under learned vector fields with numerical ODE solvers, as in ODE-RNN, neural controlled differential equations, and Liquid Time-Constant Networks (Rubanova et al., 2019; Kidger et al., 2020; Hasani et al., 2021). The second updates the state in closed form, through exponential decay in Gated Recurrent Unit with Decay (GRU-D) and the continuous-time LSTM (Che et al., 2018; Mei & Eisner, 2017), time gating in Closed-form Continuous-time Networks (CfC) (Hasani et al., 2022), or linear state-space transitions in continuous recurrent units (CRU) and the simplified state space layer (S5) (Schirmer et al., 2022; Smith et al., 2023). The second category avoids numerical integration and keeps the per-interval cost independent of interval length, making it more promising for multi-timescale modeling between irregular observations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Despite this progress, existing closed-form updates still show clear limitations when history within an interval spans multiple timescales. First, these updates decay each coordinate or eigen-direction at a single rate, forcing fast fluctuations and slow trends onto one timescale, and the logistic gate of CfC even makes the effective decay rate grow with elapsed time (Che et al., 2018; Mei & Eisner, 2017; Hasani et al., 2022; Schirmer et al., 2022). Second, with decay directions tied to coordinate axes or to an eigenbasis fixed after training, existing methods can hardly adapt them to each event at low cost, as even the state-dependent CRU evaluates a matrix exponential at every observation (Schirmer et al., 2022; Smith et al., 2023). To overcome these limitations, an ideal closed-form

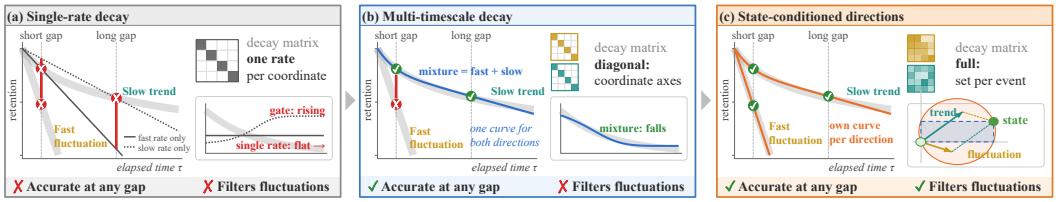


Figure 1. Retention of a slow trend and a fast fluctuation (gray bands) under single-rate decay, multi-timescale decay, and decay along state-conditioned directions.

framework for irregular sequences should meet two design objectives (Figure 1). First, the retained history should decay over multiple timescales within each interval, with an effective decay rate that can decrease over elapsed time under closed-form propagation. Second, the directions along which the residual decays should be generated from the current state and observation without per-interval matrix exponentials or excessive parameter cost.

In this paper, we propose the State-Conditioned Spectral Closed-Form Continuous-Time Network (SC-SCfC), a closed-form framework for irregularly observed sequences. By letting the current state and observation determine an interval target and an adaptive spectrum of decay rates and weights, SC-SCfC decouples the value the state approaches from how the residual decays as a mixture of exponentials. To adapt the decay directions, we further propose a state-conditioned matrix allocation (SC-MA) mechanism that decomposes the residual through positive semi-definite matrices summing to the identity, generated by a shared congruence normalization with a compact dictionary. With these mechanisms, SC-SCfC with SC-MA achieves multi-timescale propagation along context-dependent directions with exact closed-form evaluation, while preserving the state at zero elapsed time and bounding the residual by its initial norm.

The main contributions of this work are summarized as follows.

1. We propose SC-SCfC, whose state-conditioned exponential mixtures generalize single-rate exponential decay and let the effective decay rate decrease over elapsed time, balancing multi-timescale retention with exact closed-form evaluation.
2. We design SC-MA, which allocates the residual through state-conditioned positive semi-definite matrices summing to the identity, removing the fixed-direction bottleneck of closed-form updates without per-interval matrix exponentials, while a compact dictionary keeps its parameter cost practical.
3. Experiments on four benchmarks, covering regression, classification, and clinical risk prediction, show that SC-SCfC with SC-MA achieves competitive performance against continuous-time and recurrent baselines.

2 Related Work

Timescales of retention. GRU-D and the continuous-time LSTM use coordinate-wise decay (Che et al., 2018; Mei & Eisner, 2017), and Time-Aware LSTM discounts short-term memory by a predefined time-gap function (Baytas et al., 2017). CfC retains one head through a logistic time gate whose effective decay rate increases with elapsed time, and its exponential variant assigns each coordinate a single rate per interval (Hasani et al., 2022). Other methods model temporal trajectories (Holt et al., 2022; Chen et al., 2023; Shende et al., 2026) or aggregate observations at multiple granularities (Zhang et al., 2023; Zhong et al., 2025; Zhang et al., 2024; Li et al., 2026). A SurF variant

models point-process intensity as a history-conditioned positive combination of decaying exponentials (Rezaei et al., 2026). Our method instead applies a normalized, state-conditioned exponential mixture to the state-to-target residual, yielding multiple decay lifetimes and a non-increasing effective decay rate within each interval.

Directions of decay. Structured state space models constrain transitions for efficiency (Gu et al., 2022; Smith et al., 2023; Farsang & Grosu, 2025). CRU supports state-conditioned dynamics through matrix exponentials, and its fast variant reduces this cost by sharing an orthogonal eigenbasis across events (Schirmer et al., 2022). Selective and delta-rule models adapt forgetting strength or direction through input-conditioned discrete-step updates (Gu & Dao, 2024; Lahoti et al., 2026; Yang et al., 2025; Kimi Team, 2025). Our method extends the mixture by pairing positive decay rates with state-conditioned positive semi-definite allocation matrices that sum to the identity. These allocations adapt residual decay directions to each event, while propagation requires only scalar exponentials rather than a matrix exponential. The dense and compact allocation generators relate to HyperNetworks (Ha et al., 2017) and CondConv (Yang et al., 2019), respectively.

3 METHODOLOGY

SC-SCfC models the hidden state between irregularly spaced observations as approaching an interval target, while the residual between the state and this target decays according to a spectrum of decay rates and weights. At each event, the current state and observation jointly determine the target and the spectrum. Figure 2 illustrates the shared event recurrence and the closed-form propagation within each interval.

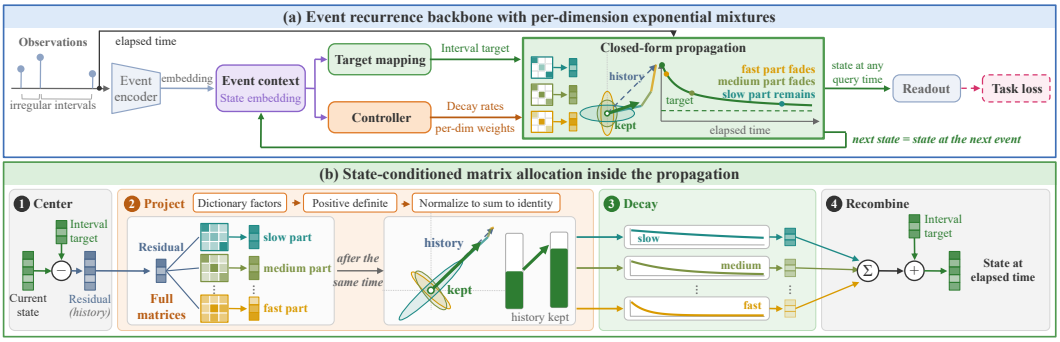


Figure 2. Overview of our framework. (a) Shared event recurrence. (b) Closed-form propagation within an interval, combining the exponential mixture with SC-MA.

3.1 SHARED EVENT RECURRENCE

Consider an event sequence $\mathcal{E} = \{(x_i, t_i)\}_{i=1}^{N}$ with irregularly spaced times $t_1 < \cdots < t_N$. An encoder maps each observation to $z_i = E_\omega(x_i)$, and $h_i = h(t_i^-) \in \mathbb{R}^H$ denotes the hidden state immediately before event i, with h_1 a learned initial state. From the event context (h_i, z_i), a target mapping produces the interval target $h_i^{\text{tar}} = C_\phi(h_i, z_i) \in \mathbb{R}^H$, and a controller produces a spectrum of decay rates and weights that parameterizes the propagation operator $G_i(\tau) \in \mathbb{R}^{H\times H}$. For $1 \le i < N$, the state on the interval of length $\Delta_i = t_{i+1} - t_i$ approaches the target as

$$h(t_i + \tau) = h_i^{\text{tar}} + G_i(\tau)\, e_i^{\text{res}}, \qquad e_i^{\text{res}} = h_i - h_i^{\text{tar}}, \qquad \tau \in [0, \Delta_i], \tag{1}$$

where the initial residual e_i^{res} carries the retained history and τ is the elapsed time. The boundary condition $G_i(0) = I_H$ yields $h(t_i^+) = h_i$, so the observation affects the state through the interval target and the spectrum rather than through a jump. Evaluating Eq. (1) at $\tau = \Delta_i$ gives the next state h_{i+1}, whose context (h_{i+1}, z_{i+1}) then sets the next target and spectrum.

Since h_i^{tar} and the spectrum stay fixed over the interval and, in both forms below, $G_i(\tau)$ is a finite sum of exponentials in τ, the state at any elapsed time admits exact closed-form evaluation without a

numerical ODE solver. Only the H-dimensional h_i is carried across events, regardless of how many decay rates the spectrum contains. A task-specific readout maps the state at events or any query time between them to a prediction, and all modules are trained end to end by backpropagating a task loss through the unrolled recurrence.

3.2 State-Conditioned Exponential Mixtures

To capture multiple timescales within a single interval, SC-SCfC allocates the residual of each coordinate to K spectral atoms, each defined by a weight and a decay rate. For coordinate j at event i, the spectrum and the retention function it induces are

$$\nu_{i,j} = \sum_{k=1}^{K} w_{i,j,k}\delta_{\lambda_{i,j,k}}, \qquad \Gamma_{i,j}(\tau) = \int_0^\infty e^{-\lambda\tau}\,\nu_{i,j}(d\lambda) = \sum_{k=1}^{K} w_{i,j,k}e^{-\lambda_{i,j,k}\tau}, \tag{2}$$

where δ_λ denotes a point mass at rate λ, $w_{i,j,k} \geq 0$, $\sum_{k=1}^{K} w_{i,j,k} = 1$, and $\lambda_{i,j,k} > 0$. The weights determine the fraction of the residual carried by each atom, the rates govern how quickly each fraction decays, and K sets the spectral resolution. With $K = 1$ the mixture reduces to single-rate exponential decay $e^{-\lambda\tau}$. The coordinate-wise operator $G_i^{\mathrm{cw}}(\tau) = \mathrm{Diag}(\Gamma_{i,1}(\tau), \ldots, \Gamma_{i,H}(\tau))$ instantiates G_i in Eq. (1), so coordinate j evolves as $[h(t_i + \tau)]_j = h_{i,j}^{\mathrm{tar}} + \Gamma_{i,j}(\tau)\, e_{i,j}^{\mathrm{res}}$, and normalization gives $\Gamma_{i,j}(0) = 1$ and hence $G_i^{\mathrm{cw}}(0) = I_H$. The diagonal structure constrains only the interval propagation, since the encoder, target mapping, and controller still mix all coordinates of the hidden state at each event.

The controller maps the event context (h_i, z_i) to weight logits $\ell_{i,j,k}^{\mathrm{wt}}$, which give $w_{i,j,k} = \mathrm{softmax}_k(\ell_{i,j,k}^{\mathrm{wt}})$, and to rate logits $\ell_{i,j,k}^{\mathrm{rate}}$, which give rates in a bounded log-scale range,

$$\lambda_{i,j,k} = \exp\left[\log\lambda_{\min} + \sigma\big(\ell_k^{\mathrm{anchor}} + s_\lambda \tanh \ell_{i,j,k}^{\mathrm{rate}}\big)\log\frac{\lambda_{\max}}{\lambda_{\min}}\right], \tag{3}$$

where $0 < \lambda_{\min} < \lambda_{\max}$ define the rate range, σ is the sigmoid function, ℓ_k^{anchor} is a fixed logit that anchors atom k at its own base timescale, and $s_\lambda > 0$ bounds the state-conditioned offset around this anchor.

Decreasing effective decay rate. Although its parameters are fixed within an interval, the mixture has a time-varying effective decay rate. Dropping the event and coordinate indices, this rate is $\rho(\tau) = -\partial_\tau \log\Gamma(\tau) = \sum_k \widetilde{w}_k(\tau)\lambda_k$, a weighted average under the normalized surviving weights $\widetilde{w}_k(\tau) = w_k e^{-\lambda_k\tau}/\Gamma(\tau)$, with $\rho'(\tau) = -\mathrm{Var}_{\widetilde{w}(\tau)}(\lambda) \leq 0$. Fast atoms thus lose relative weight over time and slow atoms dominate, so ρ decreases unless all atoms with positive weight share a single rate. A single exponential matching Γ at interval length Δ needs the rate $\lambda_{\mathrm{eff}}(\Delta) = -\log\Gamma(\Delta)/\Delta$, the average of ρ over $[0, \Delta]$, which decreases with Δ; the mixture realizes this entire family with parameters fixed at the event, before Δ is known.

Bounded state. Since every rate lies in $(\lambda_{\min}, \lambda_{\max})$, the retention decreases monotonically with $e^{-\lambda_{\max}\tau} < \Gamma_{i,j}(\tau) < e^{-\lambda_{\min}\tau}$ for $\tau > 0$, so each coordinate approaches its target monotonically without overshooting. By induction over events, if every target satisfies $\|h_i^{\mathrm{tar}}\|_\infty \leq M$, then $\|h(t)\|_\infty \leq \max\{\|h_1\|_\infty, M\}$ at all modeled times, independently of the sequence length and interval durations.

3.3 State-Conditioned Matrix Allocation

Although the coordinate-wise exponential mixture adapts decay rates and weights, its decay directions remain tied to the coordinate axes. As hidden coordinates carry no privileged meaning, a slow mode spread across several coordinates shares these coordinates with faster modes, and a per-coordinate retention cannot decay the two at different rates. To remove this fixed-direction constraint, SC-MA extends the exponential mixture by replacing the scalar weights $w_{i,j,k}$ with positive semi-definite matrices that sum to the identity and are generated from the event context, so that the residual decays along context-dependent directions.

For each atom k, the event context determines an allocation matrix $P_{i,k} \in \mathbb{R}^{H\times H}$ satisfying $P_{i,k} \succeq 0$ and $\sum_{k=1}^{K} P_{i,k} = I_H$. These matrices split the initial residual into components $e_{i,k}^{\mathrm{res}} = P_{i,k}e_i^{\mathrm{res}}$ that

sum to e_i^{res}. In contrast to the eigenprojectors of a diagonalized transition matrix, the allocation matrices need not be idempotent, mutually orthogonal, or commuting, allowing multiple atoms to share a common direction. Each atom carries a single rate $\lambda_{i,k}$ shared by all directions, obtained from Eq. (3) with a per-atom rate logit $\ell_{i,k}^{\mathrm{rate}}$ in place of $\ell_{i,j,k}^{\mathrm{rate}}$. The resulting propagation operator is

$$G_i^{\mathrm{MA}}(\tau) = \sum_{k=1}^{K} e^{-\lambda_{i,k}\tau} P_{i,k}, \tag{4}$$

and SC-SCfC with SC-MA sets $G_i = G_i^{\mathrm{MA}}$ in Eq. (1), giving the interval state $h(t_i + \tau) = h_i^{\mathrm{tar}} + \sum_{k=1}^{K} e^{-\lambda_{i,k}\tau} e_{i,k}^{\mathrm{res}}$. The matrix $P_{i,k}$ determines which residual directions atom k carries and in what proportion, while $\lambda_{i,k}$ governs their decay rate. Since the allocation matrices sum to the identity, $G_i^{\mathrm{MA}}(0) = I_H$ satisfies the boundary condition, and since they are formed once per event, evaluating Eq. (4) at any elapsed time requires only K scalar exponentials and no matrix exponential.

Directional interpretation. SC-MA admits a direction-wise reading without assuming a common eigenbasis. Dropping the event index, any unit vector v gives $v^\top G^{\mathrm{MA}}(\tau) v = \sum_{k=1}^{K} e^{-\lambda_k \tau} (v^\top P_k v)$ with non-negative weights $v^\top P_k v$ that sum to one. For a residual initially aligned with v, the fraction retained along v is a convex exponential mixture, whose effective decay rate decreases with elapsed time. Taking v as the j-th coordinate axis gives a coordinate-wise mixture with weights $(P_k)_{jj}$ and shared rates λ_k, while the off-diagonal entries of P_k couple coordinates during propagation.

Generating the allocation matrices. To obtain context-dependent positive semi-definite matrices that sum to the identity, the controller generates unconstrained matrix factors, forms positive definite matrices from them, and normalizes these by their sum. Let $\xi_i^{\mathrm{alloc}} \in \mathbb{R}^H$ be a feature computed by the controller from the event context (h_i, z_i), and let $\pi_k > 0$ with $\sum_{k=1}^{K} \pi_k = 1$ be base weights. The dense parameterization uses $B_{i,k} = \sqrt{\pi_k} I_H + (s_B/\sqrt{H})\,\mathrm{MatGen}_k(\xi_i^{\mathrm{alloc}})$, where $\mathrm{MatGen}_k : \mathbb{R}^H \to \mathbb{R}^{H\times H}$ is an affine map whose output is reshaped into an $H \times H$ matrix and $s_B > 0$ scales the factor offset. The matrices are constructed as

$$A_{i,k} = B_{i,k} B_{i,k}^\top + \varepsilon \pi_k I_H, \qquad S_i = \sum_{k=1}^{K} A_{i,k} = L_i L_i^\top, \qquad P_{i,k} = L_i^{-1} A_{i,k} L_i^{-\top}, \tag{5}$$

where $\varepsilon > 0$ and L_i is the Cholesky factor of S_i, which exists because the ridge terms and $\sum_k \pi_k = 1$ give $S_i \succeq \varepsilon I_H$. This shared congruence normalization applies the same $L_i^{-1}(\cdot)L_i^{-\top}$ to every $A_{i,k}$, preserving positive semi-definiteness and mapping $\sum_{k=1}^{K} A_{i,k} = S_i$ to $\sum_{k=1}^{K} P_{i,k} = I_H$. When the offset vanishes, $P_{i,k} = \pi_k I_H$ and SC-MA reduces to an isotropic scalar mixture with weights π_k, from which the controller learns directional deviations.

Compact dictionary parameterization. The dense generator requires $KH^2(H+1)$ parameters, cubic in H. The compact form instead represents each factor offset with a matrix dictionary shared across events: for each atom k, it learns R full matrices $D_{k,r} \in \mathbb{R}^{H\times H}$, whose coefficients $\alpha_{i,k,r}^{\mathrm{dict}}$ are produced by an affine map of ξ_i^{alloc}, giving the factor

$$B_{i,k} = \sqrt{\pi_k} I_H + \frac{s_B}{\sqrt{HR}} \sum_{r=1}^{R} \alpha_{i,k,r}^{\mathrm{dict}} D_{k,r}. \tag{6}$$

These factors then pass through Eq. (5) and the propagation exactly as in the dense form. Since each $D_{k,r}$ is a full matrix, the offset is restricted to an R-dimensional span of matrices rather than to matrices of rank R. The compact generator requires $KR(H^2+H+1)$ parameters, an approximately H/R-fold reduction, while the per-event cost of forming and normalizing the $H \times H$ matrices remains unchanged.

Residual bound. Since every rate lies in $(\lambda_{\min}, \lambda_{\max})$ and the allocation matrices are positive semi-definite and sum to the identity, $e^{-\lambda_{\max}\tau} I_H \preceq G_i^{\mathrm{MA}}(\tau) \preceq e^{-\lambda_{\min}\tau} I_H$, so the residual never exceeds its initial norm, $\|G_i^{\mathrm{MA}}(\tau) e_i^{\mathrm{res}}\|_2 \le e^{-\lambda_{\min}\tau} \|e_i^{\mathrm{res}}\|_2$. This envelope does not render the residual norm monotone when the $P_{i,k}$ do not commute, and non-diagonal allocations no longer confine each coordinate between its starting value and its target. The coordinate-wise exponential mixture thus uses HK rates with axis-aligned allocation and a global ℓ_∞ bound, whereas SC-MA uses K shared rates with context-dependent directions and a per-interval ℓ_2 bound.

4 EXPERIMENTS

We evaluate SC-SCfC and its matrix-allocation extension SC-MA on four sequence tasks.

4.1 EXPERIMENTAL SETUP

The tasks are remaining-useful-life (RUL) regression on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) subsets FD001 to FD004 (Saxena et al., 2008), sequence classification on QuickDraw10, a ten-class subset of QuickDraw (Google, 2017), event classification on Localization Data for Person Activity (PAR) (Vidulin et al., 2010), and hourly sepsis prediction on the PhysioNet/Computing in Cardiology Challenge 2019 (P19) (Reyna et al., 2019; 2020). Together, they test whether a closed-form propagator retains information across irregular intervals when the useful history spans multiple timescales and its decay directions depend on the current context. QuickDraw10 requires one decision that combines short local strokes with the overall shape, PAR varies the interval length between consecutive events, P19 varies the elapsed time at which the state is queried after an observation, and the C-MAPSS subsets vary the number of fault modes and operating conditions. We report MAE, RMSE, and NASA Score for C-MAPSS, sequence accuracy for QuickDraw10, event-position accuracy for PAR, and hourly AUROC for P19.

SC-MA uses the compact dictionary parameterization with $R = 2$. Baselines, from the most recent, are DFA-CfN (Liang et al., 2026), LrcSSM (Farsang & Grosu, 2025), CfC and mixed-memory CfC (MM-CfC) (Hasani et al., 2022), Neural CDE (Kidger et al., 2020), GRU-D (Che et al., 2018), GRU (Cho et al., 2014), and LSTM (Hochreiter & Schmidhuber, 1997). The baseline set differs by task, and each table or figure lists the methods it compares. All results are mean $\pm$ sample standard deviation (SD) over training seeds on a fixed data split.

Table 1. QuickDraw10 test accuracy (%) on 6,982 drawings, as mean $\pm$ sample SD over five training seeds.

Method	SC-MA	GRU	SC-SCfC	Neural CDE	LSTM	DFA-CfN	CfC	LrcSSM	MM-CfC
Acc.	**97.44**	96.50	96.98	95.96	96.65	93.69	95.95	94.37	96.43
(%)	**$\pm$0.34**	$\pm$0.16	$\pm$0.23	$\pm$0.14	$\pm$0.18	$\pm$0.69	$\pm$0.14	$\pm$0.29	$\pm$0.18

4.2 MAIN RESULTS

On QuickDraw10, SC-MA reaches $97.44 \pm 0.34\%$ accuracy (Table 1), exceeding LSTM, the strongest external baseline, by 0.79 percentage points (pp) and SC-SCfC by 0.46 pp. Per class, SC-MA attains a higher mean recall than each of the four strongest external baselines, LSTM, GRU, MM-CfC, and CfC, in all ten classes and than SC-SCfC in eight of them (Figure 3a). On PAR, SC-SCfC and SC-MA both exceed every external baseline in each non-clipped gap group (Table 3), while the two remain within one seed SD of each other. On P19, SC-MA reaches an hourly AUROC of 0.7889 ± 0.0033 (Table 4), exceeding SC-SCfC by 0.0082 and every external baseline except GRU-D (0.7956 ± 0.0059).

On C-MAPSS, SC-SCfC or SC-MA attains the lowest mean RMSE on every subset (Table 2), whereas the lowest NASA Score is obtained by SC-SCfC on FD001 and FD003 and by MM-CfC on FD002 and FD004. NASA Score penalizes each error exponentially, more steeply for late than for early predictions, and sums the penalty over engines (Saxena et al., 2008), so a few engines with large late errors can dominate it; this effect is strongest on FD002 and FD004, whose six operating conditions produce the largest errors for every method. On both subsets, SC-MA has a lower RMSE than MM-CfC but a slightly higher Score, which suggests that its errors are smaller on average but include a few larger late predictions, and the Score differences (55 on FD002 and 32 on FD004) are smaller than the seed SD of either model. On FD003, SC-MA attains the lowest RMSE while SC-SCfC attains the lowest Score, and the larger seed SD of SC-MA in Score (208 versus 66) indicates that this gap (58) comes mainly from a few seeds. Relative to external baselines, the error reductions of SC-MA concentrate at true RUL below 75 cycles on FD002 and FD003, the range most relevant to end-of-life decisions (Figure 4).

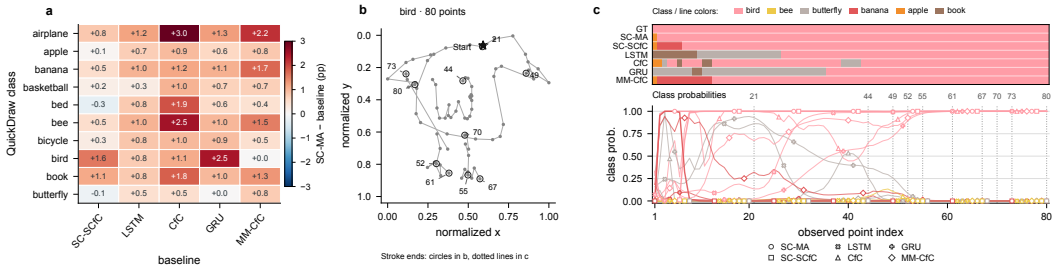


Figure 3. QuickDraw10 per-class recall and one prefix case. (a) Mean class-recall gain of SC-MA over SC-SCfC and the four strongest external baselines in Table 1 (pp) on 6,982 test drawings over five training seeds, with red favoring SC-MA. (b) Normalized trajectory of one held-out validation drawing of the bird class with 80 points. (c) Top-1 prediction of SC-MA, SC-SCfC, LSTM, CfC, GRU, and MM-CfC after each observed point (top) and prefix class probabilities (bottom), with line colors for classes and marker shapes for methods.

Table 2. C-MAPSS final-cycle test results on the fixed split, averaged over training seeds 42–46. Values are mean ± sample SD. MAE and RMSE are in cycles, and NASA Score is summed over test engines. Bold marks the lowest mean per column.

	FD001			FD002		
Method	MAE	RMSE	Score	MAE	RMSE	Score
SC-SCfC	**10.34** ± 0.16	**13.93** ± 0.41	**273** ± 21	**14.95** ± 0.70	**23.19** ± 0.79	4308 ± 1326
SC-MA	10.82 ± 0.84	14.59 ± 1.10	314 ± 53	14.98 ± 0.36	23.37 ± 0.35	4099 ± 284
DFA-CfN	12.01 ± 0.26	16.34 ± 0.34	391 ± 51	17.77 ± 0.27	25.06 ± 0.38	4780 ± 503
LrcSSM	21.51 ± 4.72	28.17 ± 5.82	5704 ± 8687	26.30 ± 7.93	33.81 ± 7.95	69983 ± 128432
MM-CfC	11.57 ± 0.53	15.82 ± 0.52	358 ± 34	15.63 ± 0.28	23.59 ± 0.56	**4044** ± 459
CfC	12.75 ± 0.22	17.86 ± 0.18	828 ± 52	20.34 ± 0.35	29.16 ± 0.35	11608 ± 801
LSTM	11.80 ± 0.43	15.90 ± 0.47	338 ± 27	17.05 ± 1.83	25.00 ± 1.94	5486 ± 2107
	FD003			FD004		
Method	MAE	RMSE	Score	MAE	RMSE	Score
SC-SCfC	11.04 ± 0.21	15.42 ± 0.30	**463** ± 66	17.53 ± 1.11	24.95 ± 0.74	4002 ± 560
SC-MA	**10.72** ± 1.16	**14.92** ± 1.58	521 ± 208	**17.08** ± 0.25	**24.63** ± 0.76	3842 ± 776
DFA-CfN	13.83 ± 0.31	18.24 ± 0.56	1191 ± 304	20.04 ± 0.13	27.43 ± 0.19	5163 ± 133
LrcSSM	23.73 ± 7.67	30.30 ± 9.52	152927 ± 260457	28.31 ± 3.07	36.47 ± 5.08	24529 ± 21823
MM-CfC	13.65 ± 1.58	18.11 ± 1.98	1255 ± 447	17.35 ± 0.45	24.84 ± 0.23	**3810** ± 171
CfC	15.56 ± 1.11	20.66 ± 1.05	2229 ± 639	23.43 ± 0.46	30.89 ± 0.43	7964 ± 401
LSTM	12.87 ± 0.74	17.04 ± 0.75	855 ± 78	18.23 ± 1.33	25.47 ± 1.02	4058 ± 456

4.3 Analysis of Model Components

We examine how decisions form at events, how propagation behaves across interval lengths and elapsed times, and where the state-conditioned allocation of SC-MA helps over SC-SCfC.

Decision formation at events. Each observation changes the state only through the interval target and the spectrum (Eq. 1), and the mixture lets evidence at different timescales persist for different durations. In the bird drawing of Figure 3b, SC-MA predicts bird from the second point onward and retains this prediction through all 80 points (Figure 3c). SC-SCfC briefly predicts banana before settling on bird within the first ten points. The external baselines pass through banana, butterfly, and other classes before settling on bird: MM-CfC after about 12 points, LSTM after about 28, GRU after about 35, and CfC only after more than 40. In this example, GRU and LSTM hold a butterfly prediction for about 20 consecutive points, whereas in SC-MA the evidence of individual strokes decays quickly and does not override the accumulated shape. This behavior is consistent with the intended role of the mixture, in which a slow component accumulates the overall shape while the evidence of each local stroke decays quickly.

Propagation across interval lengths and elapsed times. SC-SCfC and SC-MA both fix the interval target and the spectrum at an event before the next interval length is known, and the closed-form

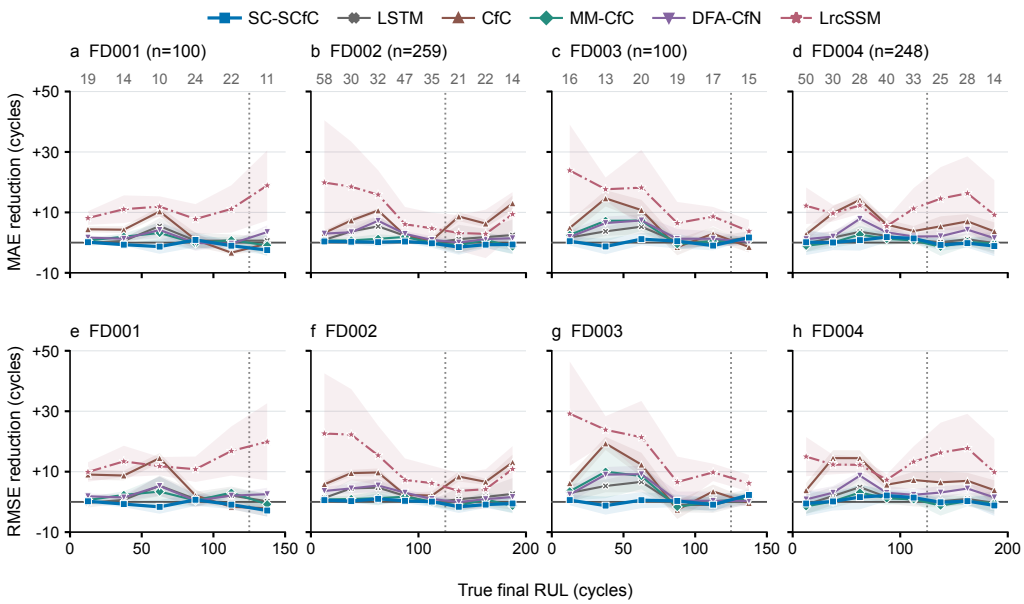


Figure 4. C-MAPSS MAE (top) and RMSE (bottom) reduction, computed as baseline minus SC-MA, with positive values favoring SC-MA. Curves show the mean across five paired training-seed contrasts and ribbons show the sample SD. Metrics use matched engines in 25-cycle true-RUL bins, and counts apply to both rows. FD001/FD003 axes end at 150 cycles because the final two bins are empty; FD002/FD004 extend to 200. The dotted line marks the training-label cap.

operator then gives the state at any elapsed time. On PAR, where the interval length varies, accuracy drops for every method once the gap exceeds 54 ms, whereas the margin of SC-MA over LrcSSM, the strongest external baseline, stays between 1.27 and 1.38 pp across the three non-clipped gap groups (Table 3). On P19, where the query time after an observation varies, SC-MA reaches an AUROC of 0.8107 ± 0.0138 one hour after an observation, compared with 0.7953 ± 0.0164 for LrcSSM (Figure 5a). After two or more hours without an observation, LrcSSM and DFA-CfN exceed both of our models. SC-MA therefore keeps its advantage at moderate elapsed times, and long unobserved gaps remain a limitation of both models.

Effect of state-conditioned matrix allocation. SC-MA differs from SC-SCfC by replacing the coordinate-wise scalar weights with state-conditioned allocation matrices and the HK coordinate-

Table 3. PAR gap-conditioned accuracy (%), as mean $\pm$ sample SD over five training seeds, computed over the 31 post-initial event positions pooled across 1,942 overlapping test windows. The groups contain 23 (0 / clipped), 48,144 ($\sim$27 ms), 10,581 ($\sim$54 ms), and 1,454 (> 54 ms) positions; the sparse 0 / clipped group consists of positions whose preceding gap is zero or clipped. Bold marks the highest mean in each gap group.

Gap	SC-SCfC	SC-MA	LrcSSM	MM-CfC	GRU	LSTM	CfC	DFA-CfN	Neural CDE
0 / clipped	83.70 ±2.01	83.70 ±4.50	**87.83 ±1.94**	84.35 ±3.89	73.91 ±4.35	76.52 ±9.02	80.87 ±9.02	75.65 ±4.96	71.30 ±7.90
$\sim$27 ms	**89.44 ±0.41**	89.42 ±0.40	88.15 ±0.50	87.95 ±0.48	85.18 ±0.26	84.43 ±0.59	86.20 ±0.74	86.42 ±0.57	85.18 ±0.46
$\sim$54 ms	88.97 ±0.34	**89.12 ±0.43**	87.80 ±0.47	87.51 ±0.51	84.79 ±0.25	84.10 ±0.54	86.01 ±0.75	86.17 ±0.61	84.62 ±0.31
> 54 ms	85.81 ±0.80	**85.93 ±0.51**	84.55 ±0.70	83.65 ±0.49	81.71 ±0.37	80.69 ±0.25	82.35 ±1.66	81.90 ±1.02	80.80 ±0.46

Table 4. P19 hourly test AUROC on 6,050 held-out patients over five training seeds.

Method	AUROC
SC-MA	0.7889 ± 0.0033
SC-SCfC	0.7807 ± 0.0040
CfC	0.7366 ± 0.0146
MM-CfC	0.7382 ± 0.0060
DFA-CfN	0.7786 ± 0.0031
GRU-D	0.7956 ± 0.0059
LrcSSM	0.7869 ± 0.0043
GRU	0.7624 ± 0.0071
LSTM	0.7721 ± 0.0028

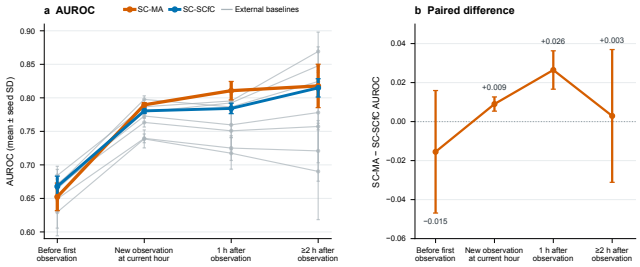


Figure 5. P19 AUROC by observation state over five training seeds. The left panel (a) shows absolute AUROC, with SC-MA and SC-SCfC highlighted and external baselines in grey. The right panel (b) shows the paired difference between SC-MA and SC-SCfC.

wise rates with K shared rates, and the paired comparison of the two models measures the joint effect of these two changes. On C-MAPSS, SC-MA attains the lower mean RMSE of the two models on FD003 and FD004, which contain two fault modes, and SC-SCfC on FD001 and FD002, which contain one; the direction of the difference follows the number of fault modes on all four subsets, although each difference lies within one seed SD. With two fault modes, the directions along which the residual decays can differ across engines, and state-conditioned allocation adapts these directions at each event. With a single fault mode, axis-aligned directions may suffice, and the finer rate resolution of SC-SCfC may become the more useful property. On PAR, about 80% of positions follow a gap of roughly 27 ms, and at this interval length the residual has little time to express a different decay direction, which is consistent with the two models remaining within one seed SD.

On P19, the paired difference in Figure 5b follows the lifetime of the allocated residual. Before the first observation, when the allocation has no observation to condition on, the difference is -0.0155 and lies within one seed SD. It rises to $+0.0090$ at a new observation and peaks at $+0.0265$ one hour later, while the allocated residual is still decaying. After two or more hours, most of the allocated residual has decayed in both models, and the difference falls to $+0.0029$, within one seed SD. Across tasks, the extension helps when the appropriate decay directions vary across sequences and the allocated residual persists long enough within an interval to act on them.

5 Conclusion

We presented SC-SCfC, a closed-form continuous-time framework for irregularly observed sequences. SC-SCfC models hidden state evolution toward an interval target as an adaptive mixture of exponential decays. The state-conditioned matrix allocation mechanism, SC-MA, decomposes the residual into vector components along data-dependent directions, each associated with a decay rate, while preserving closed-form propagation. A compact dictionary parameterization reduces the parameter overhead of generating matrix factors. Experiments demonstrate competitive predictive performance, with the gains of SC-MA over coordinate-wise SC-SCfC varying across tasks and observation conditions. Future work will investigate more efficient parameterizations of the state-conditioned matrix allocation mechanism.

AI use statement

In this work, we used large language models (LLMs) solely as a writing aid to polish the manuscript, i.e., to improve grammar, wording, and readability of text drafted by the authors. We did not use generative AI tools for any task that requires disclosure under the ICLR 2027 policy: LLMs were not used to formulate the research problem or hypotheses, develop the proposed methods or theoretical results, design experiments, implement code, process or generate data, or analyze and interpret results. Tasks such as generating synthetic data and assisting with translation are not applicable to this work. All LLM-suggested edits were reviewed by the authors, who verified that the revised

text preserves the intended technical meaning and introduces no new claims. The authors take full responsibility for the entire content of this paper, including any text refined with the aid of LLMs.

REFERENCES

Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65–74, 2017. doi: 10.1145/3097983.3097997.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*, 8:6085, 2018. doi: 10.1038/s41598-018-24271-9. URL https://doi.org/10.1038/s41598-018-24271-9.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Yuqi Chen, Kan Ren, Yansen Wang, Yuchen Fang, Weiwei Sun, and Dongsheng Li. ContiFormer: Continuous-Time Transformer for Irregular Time Series Modeling. In *Advances in Neural Information Processing Systems*, 2023. URL https://papers.nips.cc/paper/2023/hash/9328208f88ec69420031647e6ff97727-Abstract-Conference.html.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734. Association for Computational Linguistics, 2014. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Mónika Farsang and Radu Grosu. Parallelization of Non-linear State-Space Models: Scaling Up Liquid-Resistance Liquid-Capacitance Networks for Efficient Sequence Modeling. In *Advances in Neural Information Processing Systems*, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/f778e9efdb4392e56332d25647b87c09-Abstract-Conference.html.

Google. The Quick, Draw! Dataset. Dataset, 2017. URL https://github.com/googlecreativelab/quickdraw-dataset.

Albert Gu and Tri Dao. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In *Conference on Language Modeling*, 2024. URL https://arxiv.org/abs/2312.00752.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently Modeling Long Sequences with Structured State Spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

David Ha, Andrew M. Dai, and Quoc V. Le. HyperNetworks. In *International Conference on Learning Representations*, 2017. URL https://research.google/pubs/hypernetworks-2/.

Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. Liquid Time-constant Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. doi: 10.1609/aaai.v35i9.16936. URL https://ojs.aaai.org/index.php/AAAI/article/view/16936.

Ramin Hasani, Mathias Lechner, Alexander Amini, Lucas Liebenwein, Aaron Ray, Max Tschaikowski, Gerald Teschl, and Daniela Rus. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 4:992–1003, 2022. doi: 10.1038/s42256-022-00556-7. URL https://www.nature.com/articles/s42256-022-00556-7.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Samuel Holt, Zhaozhi Qian, and Mihaela van der Schaar. Neural Laplace: Learning diverse classes of differential equations in the Laplace domain. In *Proceedings of the International Conference on Machine Learning*, 2022. URL https://proceedings.mlr.press/v162/holt22a.html.

Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural Controlled Differential Equations for Irregular Time Series. In *Advances in Neural Information Processing Systems*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4a5876b450b45371f6cfe5047ac8cd45-Abstract.html.

Kimi Team. Kimi Linear: An Expressive, Efficient Attention Architecture, 2025. URL https://arxiv.org/abs/2510.26692. Technical report.

Aakash Lahoti, Kevin Y. Li, Berlin Chen, Caitlin Wang, Aviv Bick, J. Zico Kolter, Tri Dao, and Albert Gu. Mamba-3: Improved Sequence Modeling using State Space Principles. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8abd2043b71a074278d5f687947bff9c-Abstract-Conference.html.

Boyuan Li, Zhen Liu, Yicheng Luo, and Qianli Ma. Learning Recursive Multi-Scale Representations for Irregular Multivariate Time Series Forecasting. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/410dacbcfda0388f69b8983ed771c44e-Abstract-Conference.html.

Xingyu Liang, Siyi Zhou, Li'Ao Chen, Liguo Weng, Haifeng Lin, Yuxin Huang, and Min Xia. Rederived Closed-Form Continuous-Time Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 37(9):4381–4395, 2026. doi: 10.1109/TNNLS.2026.3669345. URL https://ieeexplore.ieee.org/document/11424014/.

Hongyuan Mei and Jason Eisner. The neural Hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Matthew Reyna, Chris Josef, Russell Jeter, Supreeth Shashikumar, Benjamin Moody, M. Brandon Westover, Ashish Sharma, Shamim Nemati, and Gari D. Clifford. Early Prediction of Sepsis from Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *PhysioNet*, 2019. doi: 10.13026/v64v-d857. URL https://physionet.org/content/challenge-2019/1.0.0/. Version 1.0.0.

Matthew A. Reyna, Christopher S. Josef, Russell Jeter, Supreeth P. Shashikumar, M. Brandon Westover, Shamim Nemati, Gari D. Clifford, and Ashish Sharma. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, 48(2):210–217, 2020. doi: 10.1097/CCM.0000000000004145. URL https://doi.org/10.1097/CCM.0000000000004145.

Mohammad R. Rezaei, Tejas Balaji, and Rahul G. Krishnan. SurF: A Generative Model for Multivariate Irregular Time Series Forecasting, 2026. URL https://arxiv.org/abs/2605.14069.

Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 International Conference on Prognostics and Health Management*, pp. 1–9. IEEE, 2008. URL https://c3.ndc.nasa.gov/dashlink/static/media/publication/2008_IEEEPHM_CMAPPSDamagePropagation.pdf.

Mona Schirmer, Mazin Eltayeb, Stefan Lessmann, and Maja Rudolph. Modeling irregular time series with continuous recurrent units. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 19388–19405, 2022.

Yashas Shende, Aritra Das, Reva Laxmi Chauhan, Arghya Pathak, and Debayan Gupta. Oscillators Are All You Need: Irregular Time Series Modelling via Damped Harmonic Oscillators with Closed-Form Solutions, 2026. URL https://arxiv.org/abs/2602.12139. Preprint, version 2.

Jimmy T. H. Smith, Andrew Warrington, and Scott W. Linderman. Simplified state space layers for sequence modeling. In *International Conference on Learning Representations*, 2023.

Vedrana Vidulin, Mitja Lustrek, Bostjan Kaluza, Rok Piltaver, and Jana Krivec. Localization Data for Person Activity. UCI Machine Learning Repository, 2010. URL https://archive.ics.uci.edu/dataset/196/localization.

Brandon Yang, Gabriel Bender, Quoc V. Le, and Jiquan Ngiam. CondConv: Conditionally Parameterized Convolutions for Efficient Inference. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/f2201f5191c4e92cc5af043eebfd0946-Abstract.html.

Songlin Yang, Jan Kautz, and Ali Hatamizadeh. Gated Delta Networks: Improving Mamba2 with Delta Rule. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/4904fad153f6434a7bcf04465d4be2cc-Abstract-Conference.html.

Jiawen Zhang, Shun Zheng, Wei Cao, Jiang Bian, and Jia Li. Warpformer: A Multi-scale Modeling Approach for Irregular Clinical Time Series. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023. doi: 10.1145/3580305.3599543. URL https://doi.org/10.1145/3580305.3599543.

Weijia Zhang, Chenlong Yin, Hao Liu, Xiaofang Zhou, and Hui Xiong. Irregular Multivariate Time Series Forecasting: A Transformable Patching Graph Neural Networks Approach. In *Proceedings of the International Conference on Machine Learning*, 2024. URL https://proceedings.mlr.press/v235/zhang24bw.html.

Shuhan Zhong, Weipeng Zhuo, Sizhe Song, Guanyao Li, Zhongyi Yu, and S.-H. Gary Chan. MTM: A Multi-Scale Token Mixing Transformer for Irregular Multivariate Time Series Classification. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025. doi: 10.1145/3711896.3737058. URL https://doi.org/10.1145/3711896.3737058.