CONTEXT-ADAPTIVE FLOW MATCHING FOR ICU TIME-SERIES GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Synthetic ICU time-series data can support clinical modeling, but data that help train a forecaster may still distort temporal dependence, spectral structure, or physiological relationships. We propose OU-WSP FM, a flow-matching framework for conditional ICU data generation with a structured endpoint prior. Its Ornstein–Uhlenbeck-inspired autoregressive component models local mean-reverting dynamics, while its Woodbury-spectral component extrapolates the remaining context residual using a low-rank Fourier basis. Together, these components incorporate context-dependent temporal structure into the source distribution without dense covariance inversion. Experiments on MIMIC-III and MIMIC-IV evaluate forecasting utility, statistical fidelity, and blood-pressure ordering, with ablations of spectral strength, channel sharing, and network architecture. The GRU and TCN experiments report composite forecasting errors 22.0%–27.2% lower than Gaussian FM. To distinguish the effects of endpoint design and calibration, we conduct detailed ablation experiments and calibration analyses under context-only imputation. These findings clarify the distinct roles of structured endpoints, calibration, and clinical safeguards in ICU time-series generation.

1 INTRODUCTION

Access to high-quality clinical time-series (CTS) data is constrained by privacy, governance, and uneven cohort representation, even where ICU records are substantial (Johnson et al., 2016; 2023; 2024; Yan et al., 2022; Bing et al., 2022). Synthetic generation addresses parts of this problem through adversarial, variational, diffusion, and flow-based approaches (Choi et al., 2017; Esteban et al., 2017; Yoon et al., 2019; Xu et al., 2020; Desai et al., 2021; Tian et al., 2023; Suh et al., 2024; Deng et al., 2025; Yuan & Qiao, 2024; Kollovieh et al., 2025; Liu et al., 2025). In this work, we focus on conditional CTS generation: given an observed ICU context window, the generator produces synthetic future vital-sign windows. These generated futures are paired with their observed contexts to train downstream forecasting models.

However, synthetic data that are useful for training a forecaster do not necessarily reproduce the characteristics of real ICU recordings. A generator may preserve patterns relevant to prediction while distorting the distributions of individual vital signs, their temporal dependencies, or their spectral structure. This difference is reflected in benchmarks of synthetic EHR generation, which evaluate statistical fidelity and downstream utility as separate aspects of data quality (Yan et al., 2022). For conditional ICU generation, we therefore examine not only whether synthetic futures provide useful training targets, but also whether they preserve the statistical structure and physiological relationships of real vital-sign sequences.

Flow matching learns a continuous transformation from source samples to data samples (Lipman et al., 2023).One way to incorporate temporal structure into conditional generation is to encode it in the source distribution of flow matching. An isotropic Gaussian source contains no explicit temporal dependence, whereas a structured source can incorporate information from the observed history before the learned transformation begins. TSFlow explores this approach through Gaussian-process priors, including Ornstein–Uhlenbeck (OU) kernels, and context conditioning (Kollovieh et al., 2025). Building on this direction, we investigate a source construction that separates local autoregressive dynamics from the temporal patterns they leave unexplained. We first fit an autoregressive model to the observed ICU context, then represent its residual using a small set of Fourier

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.

components and extrapolate them into the future window. This construction combines a local dynamical anchor with a spectral description of the remaining context variation, drawing on spectral approaches to time-series generation (Wilson & Adams, 2013; Crabbé et al., 2024).

In this paper, we propose OU-WSP FM, a conditional flow-matching framework that combines an OU-inspired autoregressive anchor with a context-adaptive Woodbury-spectral residual. The structured endpoint incorporates local dynamics and residual spectral patterns from the observed context without dense covariance inversion. Experiments on MIMIC-III and MIMIC-IV examine whether the generated futures provide useful training data for forecasting models, how closely they reproduce real vital-sign statistics, and whether the generated blood-pressure values are physiologically plausible. We investigate spectral strength, channel sharing, and network architecture through ablation studies. We also compare source distributions and calibration settings under context-only imputation to examine their respective effects on generation quality.

Our contributions are twofold:

- We propose OU-WSP FM for conditional ICU time-series generation, combining a local autoregressive anchor with a context-adaptive low-rank spectral residual. We connect the residual's Fourier regression formulation to a finite-feature Gaussian model, providing a statistical interpretation of the endpoint construction.
- We evaluate OU-WSP FM on MIMIC-III and MIMIC-IV across downstream forecasting utility, statistical fidelity, and blood-pressure consistency. Ablation and calibration experiments characterize the roles of the model components, revealing dataset-dependent effects of spectral correction and tradeoffs between cross-channel consistency and spectral fidelity.

2 Related Work

Synthetic ICU time-series data and EHR generation. Synthetic clinical data generation spans real-valued medical time series, visit-level EHR records, and mixed-type longitudinal patient trajectories. Early recurrent GAN models introduced train-on-synthetic-test-on-real style evaluation (Esteban et al., 2017), while TimeGAN, COT-GAN, TimeVAE, and diffusion-based time-series generators improved temporal embedding, causal transport, latent-variable synthesis, and denoising generation (Yoon et al., 2019; Xu et al., 2020; Desai et al., 2021; Yuan & Qiao, 2024). Recent methods model high-dimensional data, underrepresented populations, privacy-aware synthetic data, target-oriented utility, and mixed-type clinical variables (Theodorou et al., 2023; Bing et al., 2022; Yan et al., 2022; Tian et al., 2023; Suh et al., 2024; Deng et al., 2025; Liu et al., 2026).

Flow matching. Flow matching regresses a velocity field transporting a source to the data distribution (Lipman et al., 2023; Ma et al., 2024). TSFlow uses GP priors, including OU kernels and context-dependent GP regression (Kollovieh et al., 2025). Building on temporally informed source modeling, our construction uses a per-context autoregressive fit followed by residual-frequency selection and Fourier least-squares extrapolation. Sundial studies flow-based probabilistic forecasting (Liu et al., 2025), whereas frequency-domain diffusion changes the representation of the generative process (Crabbé et al., 2024); here spectral features enter the source construction. JET motivates preserving physiological signal structure in EEG generation (Wang et al., 2026).

Measurement and evaluation. Evaluating clinical time series requires assessing both statistical realism and functional utility. While general generative models emphasize distribution and feature-space distances such as FID (Heusel et al., 2017), clinical benchmarks also demand fidelity to physiological dynamics and downstream effectiveness (Yan et al., 2022). Accordingly, we evaluate synthetic ICU sequences across multiple complementary dimensions: downstream forecasting performance using GRU and TCN forecasters (Che et al., 2018; Bai et al., 2018), marginal and temporal autocorrelation distances, spectral alignment, and feature-level metrics (ICU-FID and Context-FID). Finally, this study focuses on continuous physiological vital signs; synthesizing full EHR records with discrete and ordinal variables remains a promising direction via Binomial Flows, score-entropy discrete diffusion, or discrete flow matching (Shenfeld et al., 2026; Lou et al., 2024; Gat et al., 2024).

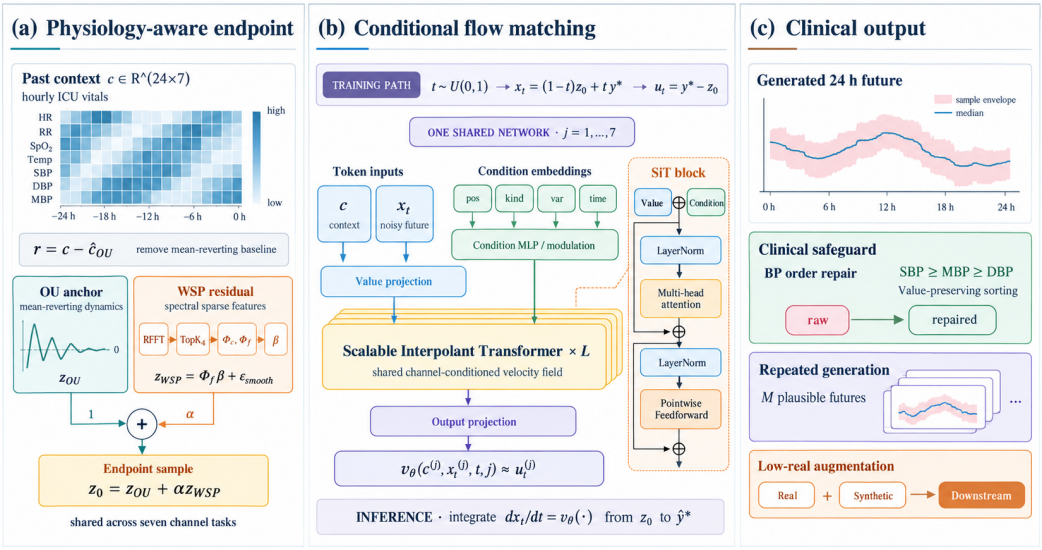


Figure 1: Overview of the proposed OU-WSP Flow Matching model.

3 METHOD

Figure 1 summarizes OU-WSP FM. We generate a future window $y \in \mathbb{R}^{H\times D}$ conditioned on an observed context $c \in \mathbb{R}^{L\times D}$, where L and H are the context and prediction-horizon lengths, and D is the number of clinical channels. We use $D = 7$ and $L = H = 24$. For channel j, the context and target are $c^{(j)} \in \mathbb{R}^L$ and $y^{(j)} \in \mathbb{R}^H$, respectively. The generator synthesizes future values only; its stochasticity comes from sampling the conditional source distribution. Source construction and flow transport operate in normalized coordinates. Preprocessing and normalization are specified by the evaluation protocol (Sec. 4). The source–calibration analysis fits its statistics on the permitted training cohort and imputes each context without using future values (Sec. 4.6).

Conditional flow matching. For each training sample and channel j, draw $z_0^{(j)} \sim q_0(\cdot \mid c^{(j)}, j)$ and $t \sim \mathcal{U}(0,1)$. The linear path, target velocity, and loss are

$$x_t^{(j)} = (1-t)z_0^{(j)} + t y^{(j)}, \qquad u_t^{(j)} = y^{(j)} - z_0^{(j)}, \tag{1}$$

$$\mathcal{L}(\theta) = \mathbb{E}_{(c,y,j),z_0,t}\left\| v_\theta(c^{(j)}, x_t^{(j)}, t, j) - u_t^{(j)} \right\|_2^2. \tag{2}$$

Flow matching is applied to the future window while the observed context remains fixed. A single velocity network is shared across all channels, with a channel embedding identifying j. Each forward pass processes the context and future state of one channel, and the generated channels are stacked into a multivariate future window.

OU-WSP endpoint prior. An isotropic Gaussian source contains no explicit temporal structure. Following informed-source modeling (Kollovieh et al., 2025), we use two components. The first is an OU-inspired Gaussian AR(1) anchor fitted to the observed context. For consecutive subsequences $c_{1:L-1}$ and $c_{2:L}$, we estimate the anchor parameters as

$$\phi = \mathrm{clip}_{[-0.95,0.98]}\left(\frac{\mathrm{Cov}(c_{1:L-1}, c_{2:L})}{\max\{\mathrm{Var}(c_{1:L-1}), 10^{-5}\}}\right), \quad \sigma = \max\{\mathrm{Std}(c_{2:L} - \phi c_{1:L-1}), 0.05\}. \tag{3}$$

The floors are in normalized units. "OU" denotes this stable discrete anchor; a negative fitted coefficient is not an exact continuous-time OU discretization. Starting from the last context value, the OU endpoint is sampled by

$$z_{\mathrm{OU},h}^{(j)} = \phi z_{\mathrm{OU},h-1}^{(j)} + \sigma\epsilon_h; \quad \epsilon_h \sim \mathcal{N}(0,1), \quad z_{\mathrm{OU},0}^{(j)} = c_L^{(j)}. \tag{4}$$

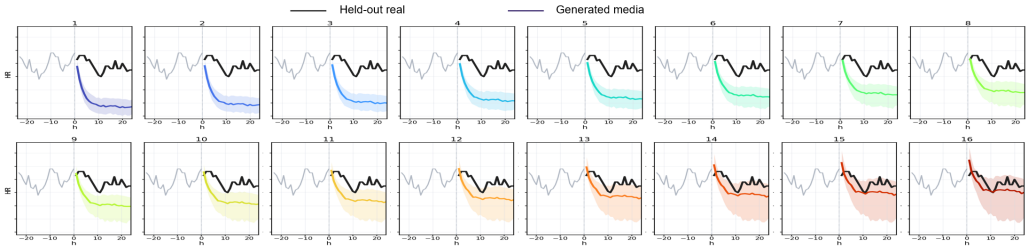


Figure 2: Conditional flow trajectories for one held-out ICU context under the evaluation protocol in Sec. 4. At each displayed Euler integration step, the median and 10–90% envelope are computed across repeated generations. The displayed MAE decreases from 28.18 to 5.00 bpm.

Here $h = 1, \ldots, H$ indexes future hours, whereas $t \in [0, 1]$ indexes the flow path.

The anchor may miss repeated residual patterns. Fit the one-step context path

$$\widehat{c}^{(j)}_{\mathrm{OU},1} = c^{(j)}_1, \quad \widehat{c}^{(j)}_{\mathrm{OU},\ell} = \phi c^{(j)}_{\ell-1} \ (\ell = 2, \ldots, L), \quad r^{(j)} = c^{(j)} - \widehat{c}^{(j)}_{\mathrm{OU}}. \tag{5}$$

Center $r^{(j)}$ and apply the real-valued fast Fourier transform (RFFT), avoiding duplicated positive/negative-frequency information. For eligible non-DC bins, rank

$$P_k = \left|\mathrm{RFFT}(r^{(j)} - \bar{r}^{(j)})_k\right|^2, \qquad \omega_m = 2\pi k_m / L, \quad m = 1, \ldots, Q, \tag{6}$$

where $k_1, \ldots, k_Q$ are the largest-scoring bins. Q is the number of frequencies; P_k is their strength, not their frequency. The real feature vector is

$$\varphi(s) = [1, \sin(\omega_1 s), \cos(\omega_1 s), \ldots, \sin(\omega_Q s), \cos(\omega_Q s)]. \tag{7}$$

Stack it on the context grid to form Φ_c and on the known future time coordinates to form Φ_f. No future *values* enter this construction. Least squares fits the context coefficients, then evaluates the same function on the future grid:

$$\widehat{\beta} \in \arg\min_{\beta} \|\Phi_c \beta - r^{(j)}\|_2^2, \quad z^{(j)}_{\mathrm{WSP-res}} = \Phi_f \widehat{\beta} + \epsilon^{(j)}_{\mathrm{smooth}}. \tag{8}$$

For the core representation, $\epsilon^{(j)}_{\mathrm{smooth}} = \tau_j A_w \eta^{(j)}$, $\eta^{(j)} \sim \mathcal{N}(0, I_H)$: A_w averages neighboring times and τ_j sets the residual-scale noise level. The sampler applies shrinkage, boundary reanchoring, and calibration to this core (Appendix B).

Motivated by Woodbury-style low-rank-plus-diagonal spectral models, WSP uses Fourier least-squares extrapolation without explicit covariance inversion. Before calibration, the OU anchor and spectral residual form the endpoint core:

$$z^{(j)}_0 = z^{(j)}_{\mathrm{OU}} + \alpha z^{(j)}_{\mathrm{WSP-res}}. \tag{9}$$

Zero residual strength removes the spectral term, not the OU anchor or any active source calibration.

Inference and calibration. At inference, an explicit Euler solver transports z_0 through the learned velocity field. For a normalized output $\bar{y}$ and its context, define $\mu_c = \mathrm{Mean}(c)$, $\mu_{\bar{y}} = \mathrm{Mean}(\bar{y})$ and $\sigma_c = \max\{\mathrm{Std}(c), 0.05\}$, $\sigma_{\bar{y}} = \max\{\mathrm{Std}(\bar{y}), 0.05\}$. The scale and boundary corrections are

$$y_{\mathrm{scale}} = \mu_c + \frac{\sigma_c}{\sigma_{\bar{y}}}(\bar{y} - \mu_{\bar{y}}), \quad y_{\mathrm{anchor}} = c_L + (y_{\mathrm{scale}} - y_{\mathrm{scale},1}). \tag{10}$$

The calibration map is

$$\tilde{y} = (1 - \lambda_{\mathrm{cal}})\bar{y} + \lambda_{\mathrm{cal}}\big(\rho y_{\mathrm{scale}} + (1 - \rho) y_{\mathrm{anchor}}\big). \tag{11}$$

It retains part of the transported sample while adjusting context-relative level, variation, and boundary continuity. Generated channels are inverse-normalized and stacked. This post-generation operation is distinct from training-time residual/source calibration.

Velocity network and token conditioning. The additive-conditioning Transformer concatenates the current channel's context and future state, $s = [c_1, \ldots, c_L, x_{t,1}, \ldots, x_{t,H}]$. Token i receives

$$h_{i,j} = \mathrm{ValueProj}(s_i) + e_{\mathrm{pos}}(i) + e_{\mathrm{kind}}(i) + e_{\mathrm{chan}}(j) + e_t. \tag{12}$$

Position, context/future kind, channel identity, and an MLP embedding of flow time are added. Self-attention spans the $L + H$ tokens, while velocities are output only for the H future tokens.

Blood-pressure relation handling. At each future hour $h = 1, \ldots, H$, let SBP_h denote systolic blood pressure, MBP_h denote mean blood pressure, and DBP_h denote diastolic blood pressure. To preserve the physiological ordering of these three quantities, the generated values should satisfy

$$\mathrm{SBP}_h \geq \mathrm{MBP}_h \geq \mathrm{DBP}_h. \tag{13}$$

For $b = (b_s, b_m, b_d)$, descending sorting $S_{\mathrm{BP}}(b) = (b_{(1)}, b_{(2)}, b_{(3)})$ preserves the three-value multiset and reassigns channel order. L2 isotonic projection instead minimizes $\|x - b\|_2^2$ over $x_s \geq x_m \geq x_d$ and may pool values. Filtering discards a full window when any triple violates the relation. The implementation performs isotonic projection, which differs from the "Sorting repair" label in Figure 3; operator attribution is discussed in Appendix B. Both repairs enforce the order by construction, not through learned constraint-following.

4 Experiments

Datasets and protocols. We evaluate on 24-hour context and future windows with seven vital-sign channels from MIMIC-III and MIMIC-IV (Johnson et al., 2016; 2023). The benchmark comprises 4,800-window synthetic training sets and 4,405 / 7,099 held-out real test windows, respectively (sample budgets detailed in Appendix B). We evaluate across two complementary settings: (i) the primary benchmark using standard offline sequence imputation following clinical benchmarking conventions (Yan et al., 2022; Kollovieh et al., 2025); and (ii) a strictly causal protocol in Sec. 4.6 that constrains imputation and statistics exclusively to the observed context window and training split.

Implementation. The generator adopts a six-layer Transformer with hidden dimension 512, eight attention heads, and dropout 0.1. Training uses AdamW for 5,000 steps with batch size 512, learning rate and weight decay 5×10^{-4}, cosine decay, mixed precision, and gradient clipping at 1.0. Generation uses 16 Euler steps, $Q = 4$ RFFT frequencies, and local calibration ($\lambda_{\mathrm{cal}} = 0.5$, $\rho = 0.65$). In the primary benchmark, the default prior configuration is validation-selected as $\alpha = 0$ for MIMIC-III and $\alpha = 0.03$ for MIMIC-IV; Sec. 4.6 provides comprehensive sensitivity analyses across $\alpha \in [0, 0.20]$.

Baselines and metrics. Baselines encompass real-data and augmentation controls; representative generative models (TimeGAN, RGAN, TimeVAE, COT-GAN, Diffusion-TS, and CauKer (Yoon et al., 2019; Esteban et al., 2017; Desai et al., 2021; Xu et al., 2020; Yuan & Qiao, 2024; Xie et al., 2025)); and flow-matching variants. "Wrapper" and "-style" indicate adapted implementations. Gaussian FM uses an uninformative prior $\mathcal{N}(0, I)$, whereas $\alpha = 0$ preserves the continuous autoregressive OU anchor. For synthetic utility, downstream GRU and TCN forecasters are trained on real contexts paired with generated futures, reporting $\mathrm{Total} = \mathrm{MASE} + \mathrm{ND} + \mathrm{NRMSE}$ (mean $\pm$ standard deviation across three trials). Fidelity is assessed via per-channel W1, ACF, log-PSD, ICU-FID, Context-FID (Esteban et al., 2017; Yoon et al., 2019; Heusel et al., 2017), and BP ordering consistency.

4.1 Forecasting Utility

We evaluate the clinical efficacy of synthetic ICU sequences by assessing their downstream predictive utility. Table 1 summarizes downstream forecasting performance across real-data controls, existing generative baselines, and flow-matching formulations.

As shown in Table 1, models trained on synthetic data generated by OU-WSP consistently outperform all competing generative paradigms across both cohorts and architectures. Specifically,

Table 1: Downstream forecasting utility on held-out real test sets (Total = MASE+ND+NRMSE, lower is better). Best performance among synthetic sources is in **bold**.

	MIMIC-III		MIMIC-IV	
Training source	**GRU ↓**	**TCN ↓**	**GRU ↓**	**TCN ↓**
Real-100%	1.4871±0.0007	1.5046±0.0009	1.4528±0.0005	1.4720±0.0033
Real-10%	1.5678±0.0042	1.6610±0.0080	1.5506±0.0037	1.7602±0.0031
Real-10% + Bootstrap	1.5297±0.0054	1.7163±0.0280	1.5011±0.0081	1.6847±0.0325
Real-10% + Jitter	1.5282±0.0062	1.7119±0.0303	1.5000±0.0068	1.6840±0.0316
TimeGAN-PyTorch ref.	2.9639±1.4253	2.9574±1.4098	2.5773±0.8416	2.5890±0.8662
TimeGAN-LSTM wrapper	2.1992±0.0142	85.5821±23.5564	2.1743±0.0034	105.6788±54.6113
RGAN wrapper	2.4738±0.1417	3.9900±0.8585	2.7792±0.2277	7.2659±2.4457
TimeVAE-base wrapper	1.8778±0.0484	2.0972±0.1710	1.8071±0.1219	1.9878±0.0511
COT-GAN-style	1.6813±0.0474	1.7832±0.0768	1.6581±0.0127	1.7581±0.0155
Diffusion-TS-style	2.3857±0.0311	4.1023±1.0430	2.1618±0.1169	3.5303±0.5020
CauKer-style	1.5565±0.0056	1.5940±0.0051	1.5213±0.0038	1.5586±0.0066
Gaussian FM	2.0317±0.0154	1.9928±0.0143	1.9876±0.0128	2.0952±0.0088
OU-WSP	**1.5279**±0.0024	**1.5551**±0.0005	**1.4917**±0.0016	**1.5252**±0.0005

OU-WSP achieves 22.0%–27.2% lower total error than standard Gaussian FM, confirming that replacing uninformative isotropic priors with structured physiological trajectories substantially eases probability flow learning. While adversarial baselines exhibit severe training instability on convolutional architectures (e.g., TimeGAN and RGAN on TCN), OU-WSP maintains tight error margins across all trials.

Importantly, on MIMIC-IV, downstream models trained solely on OU-WSP synthetic trajectories outperform models trained on limited real data (Real-10%) as well as standard heuristics (Bootstrap and Jitter), narrowing the gap to the fully observed Real-100% ceiling. Mechanism analyses in Sec. 4.6 indicate that while the mean-reverting OU anchor contributes the primary stabilization against trajectory divergence, the WSP residual provides fine-grained periodic compensation tailored to cohort-specific spectral densities.

4.2 Statistical Fidelity

Synthetic data can be useful to a forecaster while distorting marginal distributions, temporal autocorrelation, spectral texture, or feature geometry. Table 2 compares generated futures with held-out target distributions under the pipeline-evaluation protocol.

In Table 2, OU-WSP achieves superior fidelity across nearly all metrics, sweeping MIMIC-IV and leading MIMIC-III in W1, ACF, log-PSD, and ICU-FID. While Block-conditioned FM attains a marginally lower MIMIC-III Context-FID (0.373 vs. 0.383), OU-WSP consistently excels in preserving temporal autocorrelation and spectral characteristics over Gaussian FM and external baselines. To decouple intrinsic generative fidelity from post-processing interventions, Sec. 4.6 further benchmarks raw distributions without filtering or repair.

4.3 Physiological Plausibility

Standard statistical metrics can overlook domain-specific clinical constraints. Figure 3 investigates blood-pressure (BP) ordering under three operational policies: raw generation, post-hoc repair, and trajectory filtering, distinguishing sorting and isotonic projection formulations (Sec. 3).

Under raw generation, BP-triple violation rates reach 10.28% on MIMIC-III and 14.53% on MIMIC-IV. While strict trajectory filtering eliminates violations, it retains only 20%–30% of complete sequences because a single hourly violation invalidates the full 24-hour window, severely reducing the effective training budget. In contrast, post-hoc repair reduces violations to zero while maintaining 100% sample retention and competitive downstream utility. Rather than claiming intrinsic constraint adherence by the flow generator itself, these findings highlight deterministic repair as an essential clinical safeguard for downstream deployment. Unrepaired raw violation dynamics across baseline architectures are benchmarked in Sec. 4.6.

Table 2: Statistical fidelity under the pipeline-evaluation protocol. Lower is better. † denotes BP-filtered Gaussian samples; post-processing is not uniformly matched across methods.

Dataset	Source	W1	ACF	log-PSD	ICU-FID	Context-FID
III	Real-100% ref	0.130	0.006	0.025	0.004	0.003
	TimeGAN-PyTorch ref.	12.223	0.465	14.681	5.588	3.096
	TimeGAN-LSTM wrapper	7.931	0.455	10.913	2.101	2.117
	RGAN wrapper	11.458	0.510	4.106	7.931	8.354
	TimeVAE-base wrapper	6.520	0.435	2.808	4.344	4.763
	COT-GAN-style	1.468	0.082	0.355	0.530	0.536
	Diffusion-TS-style	17.806	0.288	1.806	15.217	15.854
	CauKer-style	0.879	0.075	1.206	0.436	0.402
	Block-cond. FM	0.887	0.200	0.369	0.359	**0.373**
	Gaussian FM†	1.830	0.079	0.321	0.446	0.403
	OU-WSP	**0.754**	**0.054**	**0.243**	**0.355**	0.383
IV	Real-100% ref	0.152	0.007	0.021	0.005	0.004
	TimeGAN-PyTorch ref.	11.168	0.476	14.268	3.205	2.265
	TimeGAN-LSTM wrapper	7.964	0.476	10.335	2.133	2.184
	RGAN wrapper	11.749	0.487	5.145	12.994	14.467
	TimeVAE-base wrapper	6.028	0.428	3.015	3.381	3.587
	COT-GAN-style	1.219	0.099	0.368	0.362	0.436
	Diffusion-TS-style	18.033	0.291	1.717	12.203	12.483
	CauKer-style	0.906	0.085	1.254	0.264	0.264
	Block-cond. FM	1.334	0.174	0.279	0.246	0.260
	Gaussian FM†	2.371	0.108	0.352	0.343	0.385
	OU-WSP	**0.786**	**0.071**	**0.243**	**0.184**	**0.238**

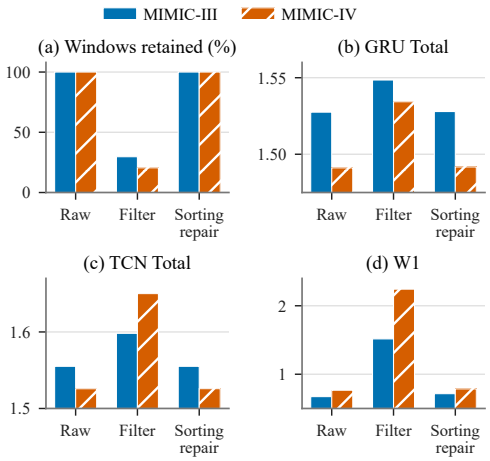


Figure 3: BP-policy tradeoffs.

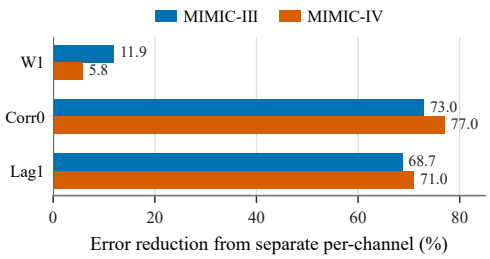


Figure 4: Fidelity-error reductions with channel sharing (pipeline evaluation).

4.4 Ablation Studies

We examine two architectural design choices: the residual spectral weight α, and channel-separate versus shared channel-conditioned generators. These ablations assess how prior parametrization and parameter sharing impact performance across cohorts.

Residual spectral weight. Following Sec. 4, we evaluate sensitivity across the spectral weight α. On MIMIC-III, the anchor-only prior ($\alpha = 0$) yields optimal GRU/TCN totals of $1.5279/1.5551$. On MIMIC-IV, positive residuals ($\alpha > 0$) consistently reduce errors (GRU: $1.4999 \rightarrow 1.4917$; TCN: $1.5308 \rightarrow 1.5252$). Thus, while the OU anchor prevents trajectory divergence, the WSP residual provides targeted frequency refinement for cohorts with pronounced oscillatory dynamics.

Table 3: Architecture diagnostic.

Dataset	Architecture	Params	GRU ↓	TCN ↓
III	Block-conditioned FM	20M	1.5814	1.6061
	AdaLN-Zero	31M	1.5382	1.5734
	SiT-style backbone	20M	1.5279	1.5551
IV	Block-conditioned FM	20M	1.4946	1.5347
	AdaLN-Zero	31M	1.4922	1.5201
	SiT-style backbone	20M	1.4917	1.5252

Channel-separate versus channel-shared generation. We contrast seven independently trained channel-specific generators against a unified generator equipped with learnable channel embeddings. In the channel-separate baseline, models are trained disjointly and stacked post hoc, whereas the shared model processes all channels via unified weights.

As illustrated in Figure 4, channel parameter sharing yields 5.8%–12.0% reductions in W1 error and reduces contemporaneous and lagged correlation mismatches (Corr0/Lag1) by 69%–77%. By projecting distinct vital signs into a unified latent space, channel conditioning provides effective cross-series regularization, substantially improving multi-channel temporal alignment.

Overall, the ablations confirm the OU anchor as a robust inductive bias, complemented by context-adaptive spectral refinement where periodic patterns are present.

4.5 Architecture Optimization

We evaluate architectural alternatives against our default 20M SiT-style backbone under the identical pipeline-evaluation protocol.

Block-conditioned FM. To examine explicit cross-variable dependencies, this variant generates clinically grouped blocks autoregressively: $\mathcal{G}_1 = \{\mathrm{HR}, \mathrm{SBP}, \mathrm{DBP}, \mathrm{MBP}\}$, $\mathcal{G}_2 = \{\mathrm{RR}, \mathrm{SpO_2}\}$, and $\mathcal{G}_3 = \{\mathrm{Temp}\}$. Subsequent blocks condition on previously generated groups:

$$v_\theta^{(g)} = v_\theta(c, \bar{y}^{(<g)}, y_t^{(g)}, t, g). \tag{14}$$

As shown in Table 3, despite introducing explicit conditioning inputs, sequential block generation underperforms parallel channel generation, as intermediate generation errors accumulate across groups.

AdaLN-Zero modulation. We also scale the model to 31M parameters by incorporating adaptive layer normalization (AdaLN-Zero) for flow-step conditioning. While AdaLN-Zero marginally improves MIMIC-IV TCN utility (1.5201 vs. 1.5252), it incurs higher prediction errors on MIMIC-III despite 55% larger model capacity.

Consequently, the 20M SiT-style backbone strikes the most favorable Pareto balance between parameter efficiency and robust forecasting accuracy across diverse patient cohorts.

4.6 Source and Calibration Analysis

We decouple the effects of source structure and output calibration by comparing four configurations: standard Gaussian FM (G), Gaussian with source calibration 0.30 (G-cal), the source-calibrated OU control (O), and full OU-WSP (W). All variants share the identical 20M-parameter backbone, a controlled cohort of 6,000 training windows per dataset, and 5,000 optimization updates with seed 2026. Imputation is strictly confined to observed contexts, and normalization statistics are fitted exclusively on the training cohort. Each generator synthesizes 6,000 future trajectories, evaluated against held-out real futures at output calibration weights 0 and 0.5.

Spectral fidelity. OU-WSP achieves consistently lower log-PSD errors than G and G-cal across both cohorts and calibration settings. On MIMIC-III, it delivers the lowest spectral discrepancy overall: 0.0962 uncalibrated and 0.1545 calibrated, reducing OU-control errors by 30.3% and 15.4%,

Table 4: Boundary-safe, single-seed distributional diagnostic. Distances are lower-is-better.

Source	Output cal.	W1 obs.	ACF	Corr0	Lag1	log-PSD	BP (%)
MIMIC-III							
G	0.0	0.5329	0.0098	0.0992	0.0701	0.1163	12.13
G	0.5	0.5140	0.0098	0.0857	0.0591	0.1863	9.80
G-cal	0.0	0.4398	0.0115	0.0942	0.0644	0.1293	11.62
G-cal	0.5	0.4893	0.0115	0.0834	0.0567	0.1807	9.69
O	0.0	0.4269	0.0319	0.1013	0.0722	0.1380	13.42
O	0.5	0.6805	0.0319	0.0886	0.0594	0.1827	10.53
W	0.0	0.5770	0.0316	0.0969	0.0697	0.0962	12.77
W	0.5	0.7454	0.0316	0.0859	0.0554	0.1545	10.25
MIMIC-IV							
G	0.0	0.4442	0.0084	0.0915	0.0603	0.1779	17.29
G	0.5	0.4346	0.0084	0.0774	0.0468	0.2193	14.77
G-cal	0.0	0.4497	0.0118	0.0882	0.0556	0.1652	17.28
G-cal	0.5	0.4145	0.0118	0.0762	0.0450	0.2052	14.86
O	0.0	0.4381	0.0279	0.0883	0.0541	0.0828	15.74
O	0.5	0.6032	0.0279	0.0759	0.0435	0.1374	13.78
W	0.0	0.6426	0.0276	0.0861	0.0517	0.0928	15.83
W	0.5	0.6128	0.0276	0.0750	0.0428	0.1477	13.94

respectively. On MIMIC-IV, its log-PSD errors (0.0928 and 0.1477) substantially outperform G (0.1779 and 0.2193), with the OU control maintaining comparable spectral alignment. Crucially, the marked advantage over G-cal confirms that source calibration alone cannot replicate the spectral preservation endowed by the WSP residual, demonstrating the distinct benefit of frequency-aware endpoint modeling.

Cross-channel structure. Incorporating the spectral residual systematically reduces contemporaneous (Corr0) and lagged (Lag1) correlation errors relative to O across both cohorts. Under output calibration, OU-WSP achieves the lowest Lag1 discrepancies among all four variants on MIMIC-III (0.0554) and MIMIC-IV (0.0428), alongside the lowest MIMIC-IV Corr0 error (0.0750). This verifies that the benefit of the spectral residual extends beyond univariate frequency matching, actively enhancing multi-channel temporal co-dynamics.

Effect of output calibration. Output calibration consistently diminishes Corr0, Lag1, and raw BP-ordering violations across all sources. For OU-WSP, uncorrected BP violations drop from 12.77% to 10.25% on MIMIC-III and from 15.83% to 13.94% on MIMIC-IV. While local calibration mildly shifts log-PSD and exhibits marginal variations on W1, it provides vital structural regularizations. In summary, source structuring (OU-WSP) and output calibration govern complementary facets: the structured prior drives global spectral and temporal fidelity, while output calibration enforces local boundary continuity and clinical ordering compliance.

5 Conclusion

We presented OU-WSP FM, a conditional flow-matching framework for ICU time-series generation. Its structured endpoint combines a local autoregressive anchor with a context-adaptive low-rank Fourier residual, incorporating temporal information from the observed context without dense covariance inversion. We connect the residual's Fourier regression formulation to a finite-feature Gaussian model, providing a statistical interpretation of the endpoint construction. Experiments on MIMIC-III and MIMIC-IV demonstrate that OU-WSP FM consistently improves forecasting utility, statistical fidelity, and physiological consistency. Relative to Gaussian FM, OU-WSP reduces log-PSD spectral errors by 17%–48% across calibration settings, while yielding the lowest lagged and contemporaneous cross-channel discrepancies. Furthermore, output calibration and deterministic BP projection jointly enhance multi-channel alignment and enforce clinical ordering constraints without discarding samples. Overall, these findings highlight the value of coupling context-adaptive structured sources with explicit clinical safeguards, establishing an interpretable framework evaluated beyond downstream utility alone.

AI USE STATEMENT

We used ChatGPT and Codex to assist with literature search and organization, methodological discussion and experimental design, mathematical derivations and proof drafting, including Lean proof scripts, code development and debugging, preprocessing checks, analysis and interpretation of experimental outputs, and manuscript drafting and revision. The authors take full responsibility for the methodological choices, mathematical claims, code, references, figures, reported results, and final manuscript. The authors reviewed the AI-assisted material and checked the reported numerical values against the experimental records.

ETHICS STATEMENT

This work concerns retrospective ICU vital-sign synthesis, not clinical decision support. Synthetic futures are conditioned on real patient contexts, which must not be treated as freely shareable synthetic records. The study offers no formal privacy or fairness guarantee. Known preprocessing and policy limitations are disclosed; BP ordering alone does not establish physiological validity.

REPRODUCIBILITY STATEMENT

The supplementary material includes numerical tables, result figures, the 16-row source–calibration CSV, a deterministic table-generation script, and metadata linking results to their source records. Appendix B documents preprocessing, calibration, BP policies, and incomplete method-to-policy correspondence. The source–calibration analysis uses one generator seed and does not include downstream evaluation. End-to-end formal verification of the clinical software pipeline is outside the scope of this work.

REFERENCES

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *CoRR*, abs/2303.08797, 2023. doi: 10.48550/arXiv.2303.08797. URL https://arxiv.org/abs/2303.08797.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *CoRR*, abs/1803.01271, 2018. doi: 10.48550/arXiv.1803.01271. URL https://arxiv.org/abs/1803.01271.

Simon Bing, Andrea Dittadi, Stefan Bauer, and Patrick Schwab. Conditional generation of medical time series for extrapolation to underrepresented populations. *CoRR*, abs/2201.08186, 2022. doi: 10.48550/arXiv.2201.08186. URL https://arxiv.org/abs/2201.08186.

Siddharth Biswal, Soumya Ghosh, Jon Duke, Bradley A. Malin, Walter Stewart, and Jimeng Sun. EVA: Generating longitudinal electronic health records using conditional variational autoencoders. *CoRR*, abs/2012.10020, 2020. doi: 10.48550/arXiv.2012.10020. URL https://arxiv.org/abs/2012.10020.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David A. Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8 (1):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 Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems 31*, pp. 6572–6583, 2018. URL https://arxiv.org/abs/1806.07366.

Yixuan Chen, Anshul Thakur, Andrew A. S. Soltan, Yifei Shen, Dongsheng Li, Mingzhi Dong, Li Shang, David A. Clifton, and Yujiang Wang. Mitigating algorithmic unfairness arising from forgetfulness of medical records in clinical artificial intelligence. *Nature Communications*, 2026. doi: 10.1038/s41467-026-72601-7. URL https://doi.org/10.1038/s41467-026-72601-7.

Edward Choi, Siddharth Biswal, Bradley A. Malin, Jon Duke, Walter F. Stewart, and Jimeng Sun. Generating multi-label discrete patient records using generative adversarial networks. In *Proceedings of the 2nd Machine Learning for Healthcare Conference*, volume 68 of *Proceedings of Machine Learning Research*, pp. 286–305, 2017. URL https://proceedings.mlr.press/v68/choi17a.html.

Jonathan Crabbé, Nicolas Huynh, Jan Pawel Stanczuk, and Mihaela Van Der Schaar. Time series diffusion in the frequency domain. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 9407–9438. PMLR, 2024. URL https://proceedings.mlr.press/v235/crabbe24a.html.

Shuai Cui, Wenxuan Chen, Wenjie Du, Jian Lou, Dan Li, and Wenjie Feng. AdaPCLA: Curriculum prior internalization for long-tailed longitudinal EHR generation. In *The Fortieth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=jdm8WK08ec.

Bowen Deng, Chang Xu, Hao Li, Yuhao Huang, Min Hou, and Jiang Bian. TarDiff: Target-oriented diffusion guidance for synthetic electronic health record time series generation. *CoRR*, abs/2504.17613, 2025. doi: 10.48550/arXiv.2504.17613. URL https://arxiv.org/abs/2504.17613.

Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. TimeVAE: A variational auto-encoder for multivariate time series generation. *CoRR*, abs/2111.08095, 2021. doi: 10.48550/arXiv.2111.08095. URL https://arxiv.org/abs/2111.08095.

Cristóbal Esteban, Stephanie L. Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional GANs. *CoRR*, abs/1706.02633, 2017. doi: 10.48550/arXiv.1706.02633. URL https://arxiv.org/abs/1706.02633.

Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky T. Q. Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete flow matching. In *Advances in Neural Information Processing Systems 37*, 2024. URL https://arxiv.org/abs/2407.15595.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems 30*, pp. 6626–6637, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html.

Mahmoud Ibrahim, Bart Elen, Chang Sun, Gökhan Ertaylan, and Michel Dumontier. Enabling granular subgroup-level model evaluations by generating synthetic medical time series. *CoRR*, abs/2510.19728, 2025. doi: 10.48550/arXiv.2510.19728. URL https://arxiv.org/abs/2510.19728.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV (version 3.1). PhysioNet, 2024. URL https://physionet.org/content/mimiciv/3.1/.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016. doi: 10.1038/sdata.2016.35. URL https://doi.org/10.1038/sdata.2016.35.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, 2023. doi: 10.1038/s41597-022-01899-x. URL https://doi.org/10.1038/s41597-022-01899-x.

Hojjat Karami, David Atienza, and Anisoara Ionescu. SynEHRgy: Synthesizing mixed-type structured electronic health records using decoder-only transformers. *CoRR*, abs/2411.13428, 2024. doi: 10.48550/arXiv.2411.13428. URL https://arxiv.org/abs/2411.13428.

Marcel Kollovieh, Marten Lienen, David Lüdke, Leo Schwinn, and Stephan Günnemann. Flow matching with gaussian process priors for probabilistic time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.03024.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems 32*, 2019. URL https://arxiv.org/abs/1904.06991.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Shaonan Liu, Yuichiro Iwashita, Soichiro Nakako, Masakazu Iwamura, and Koichi Kise. CDMT-EHR: A continuous-time diffusion framework for generating mixed-type time-series electronic health records. *CoRR*, abs/2603.23719, 2026. doi: 10.48550/arXiv.2603.23719. URL https://arxiv.org/abs/2603.23719.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: A family of highly capable time series foundation models. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. URL https://arxiv.org/abs/2502.00816.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 2024. URL https://arxiv.org/abs/2310.16834.

You Lu and Bert Huang. Woodbury transformations for deep generative flows. In *Advances in Neural Information Processing Systems 33*, pp. 5802–5811, 2020. URL https://arxiv.org/abs/2002.12229.

Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *Computer Vision – ECCV 2024*, 2024. URL https://arxiv.org/abs/2401.08740.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20*, pp. 1177–1184, 2007. URL https://proceedings.neurips.cc/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html.

Mehdi S. M. Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. In *Advances in Neural Information Processing Systems 31*, 2018. URL https://arxiv.org/abs/1806.00035.

Yair Shenfeld, Ricardo Baptista, and Stefano Peluchetti. Binomial flows: Denoising and flow matching for discrete ordinal data. *CoRR*, abs/2605.00360, 2026. doi: 10.48550/arXiv.2605.00360. URL https://arxiv.org/abs/2605.00360.

Namjoon Suh, Yuning Yang, Din-Yin Hsieh, Qitong Luan, Shirong Xu, Shixiang Zhu, and Guang Cheng. TimeAutoDiff: Combining autoencoder and diffusion model for time series tabular data synthesizing. *CoRR*, abs/2406.16028, 2024. doi: 10.48550/arXiv.2406.16028. URL https://arxiv.org/abs/2406.16028.

Brandon Theodorou, Cao Xiao, and Jimeng Sun. Synthesize extremely high-dimensional longitudinal electronic health records via hierarchical autoregressive language model. *CoRR*, abs/2304.02169, 2023. doi: 10.48550/arXiv.2304.02169. URL https://arxiv.org/abs/2304.02169.

Muhang Tian, Bernie Chen, Allan Guo, Shiyi Jiang, and Anru R. Zhang. Reliable generation of privacy-preserving synthetic electronic health record time series via diffusion models. *CoRR*, abs/2310.15290, 2023. doi: 10.48550/arXiv.2310.15290. URL https://arxiv.org/abs/2310.15290.

Amirsina Torfi and Edward A. Fox. CorGAN: Correlation-capturing convolutional generative adversarial networks for generating synthetic healthcare records. In *Proceedings of the Thirty-Third International Florida Artificial Intelligence Research Society Conference*, 2020. URL https://arxiv.org/abs/2001.09346.

Yifan Wang, Yijia Ma, Wen Li, and Chenyu You. Let EEG models learn EEG. *CoRR*, abs/2605.21280, 2026. doi: 10.48550/arXiv.2605.21280. URL https://arxiv.org/abs/2605.21280.

Zifeng Wang and Jimeng Sun. PromptEHR: Conditional electronic healthcare records generation with prompt learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2022. URL https://arxiv.org/abs/2211.01761.

Andrew Gordon Wilson and Ryan Prescott Adams. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1067–1075, 2013. URL https://proceedings.mlr.press/v28/wilson13.html.

Shifeng Xie, Vasilii Feofanov, Marius Alonso, Ambroise Odonnat, Jianfeng Zhang, Themis Palpanas, and Ievgen Redko. CauKer: Classification time series foundation models can be pretrained on synthetic data only. *CoRR*, abs/2508.02879, 2025. doi: 10.48550/arXiv.2508.02879. URL https://arxiv.org/abs/2508.02879.

Tianlin Xu, Li K. Wenliang, Michael Munn, and Beatrice Acciaio. COT-GAN: Generating sequential data via causal optimal transport. In *Advances in Neural Information Processing Systems 33*, 2020. URL https://arxiv.org/abs/2006.08571.

Chao Yan, Yao Yan, Zhiyu Wan, Ziqi Zhang, Larsson Omberg, Justin Guinney, Sean D. Mooney, and Bradley A. Malin. A multifaceted benchmarking of synthetic electronic health record generation models. *Nature Communications*, 13:7609, 2022. doi: 10.1038/s41467-022-35295-1. URL https://www.nature.com/articles/s41467-022-35295-1.

Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In *Advances in Neural Information Processing Systems 32*, pp. 5509–5519, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/c9efe5f26cd17ba6216bbe2a7d26d490-Abstract.html.

Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=4h1apFjO99.

A Extended Related Work and Proofs

A.1 Extended Related Work

Clinical generators from static records to continuous ICU trajectories. Synthetic clinical-data research covers several distinct objects that should not be conflated. Early methods such as medGAN and CorGAN generate high-dimensional binary or count-valued patient records and emphasize marginal statistics, feature correlations, predictive utility, and disclosure risk (Choi et al., 2017; Torfi & Fox, 2020). Longitudinal record generators subsequently modeled ordered visits and heterogeneous clinical events: EVA uses a conditional variational formulation for encounter sequences, PromptEHR casts controllable multimodal record synthesis as prompted text-to-text generation, HALO uses a hierarchical autoregressive model for high-dimensional codes and visits, and SynEHRgy tokenizes mixed structured records for decoder-only Transformer generation (Biswal et al., 2020; Wang & Sun, 2022; Theodorou et al., 2023; Karami et al., 2024). These systems address diagnosis, medication, procedure, demographic, and visit-level structure. By contrast, recurrent medical time-series generators such as RGAN/RCGAN, TimeGAN, COT-GAN, and TimeVAE target real-valued sequential dynamics, while Diffusion-TS and EHR-specific diffusion systems introduce denoising-based alternatives for general or clinical time series (Esteban et al., 2017; Yoon et al., 2019; Xu et al., 2020; Desai et al., 2021; Yuan & Qiao, 2024; Tian et al., 2023; Suh et al., 2024). OU-WSP FM occupies a narrower conditional setting: it generates aligned, continuous, hourly ICU vital-sign futures from a real patient context rather than synthesizing complete patient records or unconditional whole-window cohorts.

Conditional, utility-oriented, and population-aware clinical synthesis. Recent clinical generators increasingly move beyond unconditional distribution matching. HealthGen conditions on patient characteristics and explicitly models temporal values and missingness to extrapolate underrepresented populations (Bing et al., 2022). TarDiff steers diffusion samples using task-oriented influence guidance, reflecting the observation that high global fidelity does not necessarily imply high downstream value (Deng et al., 2025). AdaPCLA (Cui et al., 2026) studies long-tailed longitudinal EHR generation by using empirical code-frequency priors to guide autoregressive training. The prior-induced logit bias is gradually annealed away, encouraging the model to retain rare-code structure without a persistent external bias at standard inference. It also supports adaptation to target code frequencies without retraining. This differs from OU-WSP FM in both the generated data and the role of the prior: AdaPCLA adjusts discrete-code learning, whereas OU-WSP constructs a continuous source distribution from local autoregressive dynamics and residual Fourier features. Enhanced TimeAutoDiff studies synthetic critical-care trajectories as an instrument for population- and subgroup-level model evaluation, while synthetic-EHR benchmarking work demonstrates that utility, privacy, and fidelity can rank generators differently across use cases (Ibrahim et al., 2025; Yan et al., 2022). CDMT-EHR further addresses numerical and categorical clinical variables within a continuous-time diffusion framework (Liu et al., 2026). These developments motivate the multi-axis protocol used in this paper: generated futures are evaluated as augmentation sources, as conditional distributions relative to held-out real trajectories, and as physiological objects subject to explicit relations. They also motivate our conservative scope statement: gains under low-real forecasting do not establish that synthetic data replace full-real training or that a continuous-vital generator solves mixed-type EHR synthesis.

Flow matching, stochastic interpolants, and Transformer velocity fields. Continuous-time generative modeling builds on neural ordinary differential equations and continuous normalizing flows, which parameterize state derivatives and obtain samples through numerical integration (Chen et al., 2018). Flow Matching replaces simulation-dependent training with regression onto conditional vector fields associated with prescribed probability paths (Lipman et al., 2023); stochastic interpolants provide a broader formulation linking deterministic transport and stochastic diffusion through flexible bridges between endpoint distributions (Albergo et al., 2023). Scalable Interpolant Transformers show how Transformer backbones can parameterize these velocity or score fields at scale (Ma et al., 2024). In time series, TSFlow uses data-dependent Gaussian-process endpoints, whereas Sundial uses a flow-based forecasting objective to generate multiple probable futures (Kollovieh et al., 2025; Liu et al., 2025). JET applies conditional flow matching to raw EEG and argues that continuous physiological generation should preserve spectral structure, stationarity, and signal-level statistics (Wang et al., 2026). OU-WSP FM follows this endpoint-aware and signal-aware line, but

separates the inductive bias into a mean-reverting OU anchor, a context-selected spectral residual, and a shared channel-conditioned velocity network.

Spectral kernels and low-rank endpoint structure. The WSP module is related to finite-feature views of stationary kernels. Random Fourier features approximate shift-invariant kernels by explicit sine/cosine embeddings, replacing dense kernel algebra with a low-dimensional feature representation (Rahimi & Recht, 2007). Spectral-mixture Gaussian-process kernels connect stationary covariance design to a modeled spectral density and support pattern discovery and extrapolation (Wilson & Adams, 2013). Woodbury transformations exploit low-rank structure to make inverse and determinant calculations tractable in generative flows (Lu & Huang, 2020). Our implementation differs from exact Gaussian-process posterior sampling: the observed OU residual is transformed by RFFT, dominant frequencies define a finite Fourier dictionary, coefficients are fitted on the context, and the fitted function is evaluated on the future grid. The proof appendix shows that this path is a flat-prior limit of a finite Fourier Gaussian posterior mean and that the implemented residual and the formal WSP covariance act on the same selected Fourier feature family. The cited kernel literature therefore supplies the mathematical context for the parameterization, while the present derivation specifies what is and is not sampled by the code.

Fidelity, coverage, and downstream evaluation. Evaluation of synthetic data is intrinsically multidimensional. RCGAN popularized train-on-synthetic–test-on-real and reverse protocols for medical time series, and TimeGAN combined discriminative and predictive scores with temporal analyses (Esteban et al., 2017; Yoon et al., 2019). FID compares Gaussian approximations in a learned feature space, while precision–recall formulations separate sample quality from support coverage and expose failure modes that a single scalar distance can hide (Heusel et al., 2017; Sajjadi et al., 2018; Kynkäänniemi et al., 2019). For temporal data, feature encoders and discriminator families must themselves be audited because weak probes can miss long-range or nonlinear differences; this motivates our use of GRU/TCN forecasting utility, deep discriminators, ICU-FID, Context-FID, autocorrelation, cross-correlation, and log-PSD diagnostics (Che et al., 2018; Bai et al., 2018; Nie et al., 2023). Repeated conditional generation additionally evaluates whether a model provides useful forecast distributions rather than isolated point-like samples, an issue emphasized by probabilistic flow-based forecasters such as Sundial (Liu et al., 2025). No one metric is interpreted as proof that generated ICU trajectories are indistinguishable from real trajectories.

Clinical plausibility, privacy, fairness, and typed-data boundaries. MIMIC-III and MIMIC-IV make deidentified ICU data available under credentialed-access governance, but their availability does not remove privacy, representation, or deployment concerns (Johnson et al., 2016; 2023; 2024). Synthetic-health benchmarks consequently combine utility with membership, attribute, and disclosure-risk diagnostics (Yan et al., 2022; Wang & Sun, 2022). Work on clinical record deletion and machine unlearning further shows that privacy interventions can interact with subgroup fairness, motivating subgroup-aware diagnostics rather than universal privacy or fairness claims (Chen et al., 2026). Blood-pressure repair is treated as an explicit safeguard, separately from raw generator quality. Sorting and isotonic projection are mathematically different; the original image labels do not resolve the audited implementation discrepancy (Appendix B). The present state space is continuous; extending it to complete EHR trajectories would require dedicated treatment of ordinal scores, categorical diagnosis/medication/procedure events, and irregular event times. Binomial Flows provide an ordinal discrete construction, while score-entropy discrete diffusion and Discrete Flow Matching provide complementary tools for high-dimensional categorical generation (Shenfeld et al., 2026; Lou et al., 2024; Gat et al., 2024). Such typed modules are a future extension rather than part of the current continuous-vital claim.

Reading convention. This appendix retains the original mathematical development. Context length C here equals L in the main text, and spectral count K equals Q. Its posterior-mean and Fourier-support claims concern the stated regression core and assumptions, not an assertion that every calibrated implementation is an exact Gaussian draw. Mathematical training-coupling identities are not a certificate for a particular finite AdamW checkpoint. The original empirical record and the additional rerun are separate; see Appendix B for provenance and preprocessing qualifications.

A.2 Notation, assumptions, and scope

For one ICU window, the observed context and future of variable $j \in \{1, \dots, D\}$ are

$$c^{(j)} \in \mathbb{R}^C, \qquad y^{(j)} \in \mathbb{R}^H,$$

with $D = 7$ and $C = H = 24$ in the main experiments. The stacked window is

$$\mathbf{c} = [c^{(1)}, \dots, c^{(D)}] \in \mathbb{R}^{C \times D},$$
$$\mathbf{y} = [y^{(1)}, \dots, y^{(D)}] \in \mathbb{R}^{H \times D}.$$

For one channel, the endpoint is

$$z_0^{(j)} \sim q_0(\cdot \mid c^{(j)}, j),$$

and the linear conditional flow-matching path is

$$x_t^{(j)} = (1-t)z_0^{(j)} + ty^{(j)}, \qquad u_t^{(j)} = \frac{dx_t^{(j)}}{dt} = y^{(j)} - z_0^{(j)}. \tag{15}$$

The same channel-conditioned Transformer is reused for all variables:

$$v_\theta(c^{(j)}, x_t^{(j)}, t, j) \in \mathbb{R}^H.$$

The index j enters through a variable embedding, while the Transformer, MLP, and output-head parameters are shared.

Unless explicitly stated otherwise, all random variables are square integrable, all finite-dimensional matrices are real, and all covariance weights are nonnegative. For strict positive definiteness of a WSP covariance we assume a strictly positive diagonal floor $d > 0$. The frequency set selected from a context residual is denoted by $\mathcal{K}(r) = \{\omega_1, \dots, \omega_K\}$, and $p = 2K + 1$ is the resulting real Fourier feature dimension when a constant column is included.

The word *coupling* is used in two precise senses in this appendix. First, *optimization coupling* means that all channel losses update the same parameter vector, so the trained function for one channel depends on training information from the others. Second, *distributional coupling* means that the stacked generated channels can be statistically dependent through correlated patient contexts. Neither statement means that a single channel-conditioned forward call reads numerical values that were not supplied to it.

A.3 Population flow-matching velocity

For fixed channel j, let

$$\mathcal{F}_{j,t} = \sigma(c^{(j)}, x_t^{(j)}, t, j)$$

be the information available to the channel-conditioned network. For any square-integrable measurable function v, define

$$\mathcal{L}_j(v) = \mathbb{E}\left[\|v(c^{(j)}, x_t^{(j)}, t, j) - u_t^{(j)}\|_2^2\right].$$

Proposition 1 (conditional L^2 projection). The Bayes-optimal channel-conditioned velocity is

$$v_j^*(c^{(j)}, x_t^{(j)}, t, j) = \mathbb{E}[u_t^{(j)} \mid \mathcal{F}_{j,t}]. \tag{16}$$

Moreover, for every candidate v,

$$\mathcal{L}_j(v) = \mathbb{E}\left\|u_t^{(j)} - \mathbb{E}[u_t^{(j)} \mid \mathcal{F}_{j,t}]\right\|_2^2 \tag{17}$$
$$+ \mathbb{E}\left\|v - \mathbb{E}[u_t^{(j)} \mid \mathcal{F}_{j,t}]\right\|_2^2. \tag{18}$$

Proof. Write

$$u_t^{(j)} - v = \left(u_t^{(j)} - \mathbb{E}[u_t^{(j)} \mid \mathcal{F}_{j,t}]\right) + \left(\mathbb{E}[u_t^{(j)} \mid \mathcal{F}_{j,t}] - v\right).$$

The second term is $\mathcal{F}_{j,t}$-measurable. By the defining orthogonality property of conditional expectation, its inner product with the first term has zero expectation. Expanding the square gives (18); the second term is nonnegative and is minimized exactly by (16). □

This result explains what the network learns at population optimum: it is not directly fitting the future values, but the conditional mean transport velocity associated with the chosen endpoint and interpolation path.

A.4 SEVEN-VARIABLE COUPLING IN THE CHANNEL-CONDITIONED SHARED NETWORK

Define the channel-conditioned state

$$s_t^{(j)} = (c^{(j)}, x_t^{(j)}, t, j)$$

and the channel loss

$$\mathcal{L}_j(\theta) = \mathbb{E}\left[\|v_\theta(s_t^{(j)}) - u_t^{(j)}\|_2^2\right].$$

If $\pi_j > 0$ and $\sum_{j=1}^{D} \pi_j = 1$, the actual shared-network loss is

$$\mathcal{L}(\theta) = \sum_{j=1}^{D} \pi_j \mathcal{L}_j(\theta). \tag{19}$$

Consequently,

$$\nabla_\theta \mathcal{L}(\theta) = \sum_{j=1}^{D} \pi_j \nabla_\theta \mathcal{L}_j(\theta). \tag{20}$$

Equation (20) already distinguishes the implemented model from seven separately trained generators: all seven losses determine one parameter update.

A.4.1 EXACT CROSS-TASK GRADIENT-FLOW DYNAMICS

Let τ denote optimization time, distinct from the flow interpolation time t, and suppose

$$\frac{d\theta}{d\tau} = -\nabla_\theta \mathcal{L}(\theta).$$

For a query $s^{(j)}$, define

$$J_{j,\theta}(s^{(j)}) = \frac{\partial v_\theta(s^{(j)})}{\partial \theta} \in \mathbb{R}^{H \times |\theta|},$$

and the block tangent kernel

$$\Theta_{jk,\theta}(s^{(j)}, s^{(k)}) = J_{j,\theta}(s^{(j)}) J_{k,\theta}(s^{(k)})^\top \in \mathbb{R}^{H \times H}.$$

Let

$$e_k(s_t^{(k)}) = v_\theta(s_t^{(k)}) - u_t^{(k)}.$$

The chain rule and (20) give the exact identity

$$\frac{d}{d\tau} v_\theta(s^{(j)}) = -2 \sum_{k=1}^{D} \pi_k \mathbb{E}\left[\Theta_{jk,\theta}(s^{(j)}, s_t^{(k)}) e_k(s_t^{(k)})\right]. \tag{21}$$

Thus the error of channel k changes the learned velocity slice for channel j whenever the corresponding expectation in (21) is nonzero. Because the time embedding is a part of every hidden representation, the Jacobians and cross-task tangent blocks may themselves depend on t; the shared-parameter coupling is therefore compatible with a time-varying velocity field.

A.4.2 A FINITE-STEP COUPLING CERTIFICATE THROUGH THE SHARED READOUT

Write the shared final affine layer as

$$v_\theta(s^{(j)}) = W_{\text{out}} h_\vartheta(s^{(j)}) + b_{\text{out}}, \tag{22}$$

where $h_\vartheta(s^{(j)}) \in \mathbb{R}^m$ is the hidden representation, $W_{\text{out}} \in \mathbb{R}^{H \times m}$, and $b_{\text{out}} \in \mathbb{R}^H$ are reused for every channel. Consider one gradient step generated by a channel-k residual

$$e_k = W_{\text{out}} h_k + b_{\text{out}} - u_t^{(k)}, \qquad h_k = h_\vartheta(s_t^{(k)}).$$

The weight gradient is

$$\nabla_{W_{\text{out}}} \|e_k\|_2^2 = 2 e_k h_k^\top,$$

so

$$W_{\text{out}}^{+} = W_{\text{out}} - 2\eta e_k h_k^{\top}.$$

For another channel j, with hidden vector h_j, the exact change caused by this weight update is

$$v_{W_{\text{out}}^{+}}(s^{(j)}) - v_{W_{\text{out}}}(s^{(j)}) = -2\eta e_k \langle h_k, h_j \rangle. \tag{23}$$

This is an exact finite-step identity, not a first-order approximation. If $\eta \neq 0$, $e_k \neq 0$, and $\langle h_k, h_j \rangle \neq 0$, then training on channel k changes the output of channel j.

The shared bias supplies an even more direct common path. Since

$$\nabla_{b_{\text{out}}} \|e_k\|_2^2 = 2e_k,$$

the update $b_{\text{out}}^{+} = b_{\text{out}} - 2\eta e_k$ changes every channel output by $-2\eta e_k$. This proves the existence of a common parameter path independently of any particular attention pattern.

A.4.3 Constructive dense coupling capacity

The token representation in the main method is additive:

$$h_i = \text{ValueProj}(s_i) + e_{\text{pos}}(i) + e_{\text{kind}}(i) + e_{\text{var}}(j) + e_t.$$

The shared position, token-kind, and time embeddings allow the hidden state for channel j to contain a common component. To exhibit one concrete realizable geometry, suppose

$$h_j = q + s_j,$$

where $q \neq 0$, the channel-specific components s_j are mutually orthogonal, and each s_j is orthogonal to q. Then for any $j \neq k$,

$$\langle h_j, h_k \rangle = \|q\|_2^2 > 0.$$

Substituting this into (23) shows that the shared Transformer function class can realize a dense seven-by-seven training coupling graph. This is a capacity statement: the architecture admits such coupling, while the magnitude learned in a particular run is determined by the data and optimization trajectory.

A.4.4 Exact finite-checkpoint path decomposition

The dependence of a finite checkpoint on all seven tasks can be written without taking an infinite-width or infinitesimal-step limit. Let

$$\theta_{n+1} = \theta_n - \eta_n g_n, \qquad g_n = \sum_{k=1}^{D} \pi_k g_{k,n},$$
$$g_{k,n} = \nabla_\theta \mathcal{L}_k(\theta_n).$$

The fundamental theorem of calculus along the parameter segment gives

$$v_{\theta_{n+1}}(s^{(j)}) - v_{\theta_n}(s^{(j)}) = -\eta_n \int_0^1 J_{j,\theta_n - a\eta_n g_n}(s^{(j)}) g_n \, da \tag{24}$$

$$= \sum_{k=1}^{D} \Gamma_{j\leftarrow k}^{(n)}(s^{(j)}), \tag{25}$$

where

$$\Gamma_{j\leftarrow k}^{(n)}(s^{(j)}) = -\eta_n \pi_k \int_0^1 J_{j,\theta_n - a\eta_n g_n}(s^{(j)}) g_{k,n} \, da.$$

Summing (25) over $n = 0, \ldots, N-1$ yields

$$v_{\theta_N}(s^{(j)}) - v_{\theta_0}(s^{(j)}) = \sum_{k=1}^{D} \sum_{n=0}^{N-1} \Gamma_{j\leftarrow k}^{(n)}(s^{(j)}). \tag{26}$$

The term indexed by k is the exact optimization-path contribution of channel k to the final channel-j output. Therefore a finite checkpoint is jointly shaped by all tasks for which the corresponding path contribution is nonzero.

A.4.5 SENSITIVITY OF A CONVERGED SHARED SOLUTION

Introduce channel weights $\rho = (\rho_1, \ldots, \rho_D)$ and define

$$\mathcal{L}(\theta; \rho) = \sum_{k=1}^{D} \rho_k \mathcal{L}_k(\theta).$$

Suppose $\theta^*(\rho)$ is a differentiable branch of strict local minimizers and the Hessian

$$H_\rho = \nabla_\theta^2 \mathcal{L}(\theta^*(\rho); \rho)$$

is invertible. Differentiating the stationarity equation gives

$$\frac{\partial \theta^*}{\partial \rho_k} = -H_\rho^{-1} \nabla_\theta \mathcal{L}_k(\theta^*).$$

Consequently,

$$\frac{\partial v_{\theta^*(\rho)}(s^{(j)})}{\partial \rho_k} = -J_{j,\theta^*}(s^{(j)}) H_\rho^{-1} \nabla_\theta \mathcal{L}_k(\theta^*). \tag{27}$$

A nonzero right-hand side certifies that changing the training weight of channel k changes the converged channel-j velocity. Note that stationarity of the total loss only requires $\sum_k \rho_k \nabla \mathcal{L}_k(\theta^*) = 0$; it does not require every channel gradient to vanish separately.

A.4.6 DEPENDENCE INHERITED THROUGH CORRELATED PATIENT CONTEXTS

Let ξ_j collect the endpoint and solver randomness for channel j, and let the complete channelwise generator be

$$\widehat{y}^{(j)} = T_{\theta,j}(c^{(j)}, \xi_j).$$

Define its conditional mean

$$m_j(c^{(j)}) = \mathbb{E}[\widehat{y}^{(j)} \mid c^{(j)}].$$

For $j \neq k$, the law of total covariance gives

$$\operatorname{Cov}(\widehat{y}^{(j)}, \widehat{y}^{(k)}) = \mathbb{E}\left[\operatorname{Cov}(\widehat{y}^{(j)}, \widehat{y}^{(k)} \mid \mathbf{c})\right] \tag{28}$$

$$+ \operatorname{Cov}(m_j(c^{(j)}), m_k(c^{(k)})). \tag{29}$$

If the per-channel noises are conditionally independent, the first term is zero, but the second can remain nonzero because the seven contexts from one patient are correlated. For example, if

$$\bar{c}^{(j)} = a_j L + \eta_j, \qquad m_j(c^{(j)}) = b_j \bar{c}^{(j)},$$

where L is a shared latent patient state and the η_j are independent of L, then

$$\operatorname{Cov}(m_j(c^{(j)}), m_k(c^{(k)})) = a_j a_k b_j b_k \operatorname{Var}(L).$$

This establishes marginal statistical dependence in the stacked output. It should be distinguished from direct conditional dependence on another channel's unobserved numerical input during one forward call.

Coupling conclusion. The implemented architecture therefore has two rigorous forms of cross-variable coupling:

1. all seven velocity slices share a parameter trajectory, with exact cross-task influences described by (21)–(27);
2. correlated patient contexts can induce cross-channel dependence in the stacked generated distribution through (29).

The result does not assert that one channel-conditioned forward call reads another channel's raw value. It proves shared-parameter transfer and context-mediated statistical coupling, which are the mechanisms present in the current implementation.

A.5 OU ENDPOINT AS A CONDITIONAL MEAN-REVERTING GAUSSIAN ANCHOR

For one channel, the method estimates

$$\phi = \frac{\mathrm{Cov}(c^{(j)}_{1:C-1}, c^{(j)}_{2:C})}{\mathrm{Var}(c^{(j)}_{1:C-1})}, \qquad \sigma = \mathrm{Std}(c^{(j)}_{2:C} - \phi c^{(j)}_{1:C-1}),$$

with numerical stabilization. Starting from $z^{(j)}_{\mathrm{OU},0} = c^{(j)}_C$,

$$z^{(j)}_{\mathrm{OU},h} = \phi z^{(j)}_{\mathrm{OU},h-1} + \sigma\epsilon_h, \qquad \epsilon_h \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1). \tag{30}$$

Unrolling gives

$$z^{(j)}_{\mathrm{OU},h} = \phi^h c^{(j)}_C + \sigma \sum_{a=1}^{h} \phi^{h-a}\epsilon_a. \tag{31}$$

Therefore

$$\mathbb{E}[z^{(j)}_{\mathrm{OU},h} \mid c^{(j)}] = \phi^h c^{(j)}_C, \tag{32}$$

and

$$\mathrm{Cov}(z^{(j)}_{\mathrm{OU},h}, z^{(j)}_{\mathrm{OU},g} \mid c^{(j)}) = \sigma^2 \sum_{a=1}^{\min(h,g)} \phi^{h-a}\phi^{g-a}. \tag{33}$$

Thus the future endpoint is a conditional Gaussian vector. If $|\phi| < 1$, then

$$\mathrm{Var}(z^{(j)}_{\mathrm{OU},h} \mid c^{(j)}) = \sigma^2 \frac{1-\phi^{2h}}{1-\phi^2} \longrightarrow \frac{\sigma^2}{1-\phi^2},$$

and the conditional mean decays toward zero in normalized residual coordinates. This is the discrete mean-reverting behavior used as the physiological anchor.

For comparison, the continuous OU stochastic differential equation

$$dZ_s = -\kappa Z_s\, ds + \eta\, dW_s$$

has exact sampling interval Δ skeleton

$$Z_{n+1} = e^{-\kappa\Delta} Z_n + \zeta_n,$$
$$\zeta_n \sim \mathcal{N}\left(0, \frac{\eta^2}{2\kappa}(1 - e^{-2\kappa\Delta})\right).$$

Hence (30) is an exact OU skeleton when $0 < \phi < 1$, $\phi = e^{-\kappa\Delta}$, and the innovation variance is matched accordingly. If clipping or estimation permits values outside this correspondence, the most precise name is an *OU-inspired Gaussian AR(1) anchor*.

A.6 FROM THE FORMAL SPECTRAL MODEL TO THE IMPLEMENTED WSP CORRECTION

A.6.1 RESIDUALIZATION AS A NONZERO-MEAN MODEL

Let $\widehat{c}^{(j)}_{\mathrm{OU}}$ be the OU fit on the context grid and set

$$r^{(j)} = c^{(j)} - \widehat{c}^{(j)}_{\mathrm{OU}}. \tag{34}$$

This is the exact decomposition

$$c^{(j)} = \widehat{c}^{(j)}_{\mathrm{OU}} + r^{(j)}.$$

A corresponding stochastic model writes

$$C^{(j)}(s) = m^{(j)}_{\mathrm{OU}}(s) + R^{(j)}(s),$$

where the OU module supplies the mean and the zero-mean process $R^{(j)}$ accounts for oscillatory or periodic structure not captured by the mean-reverting component. Spectral analysis of $r^{(j)}$ is thus the finite-sample implementation of residual covariance modeling.

A.6.2 FINITE SPECTRAL MEASURES PRODUCE THE WOODBURY FORM

Consider the finite real stationary kernel

$$k_{\mathcal{K}}(s,t) = \lambda_0 + \sum_{m=1}^{K} \lambda_m \cos(\omega_m(s-t)), \qquad \lambda_m \geq 0. \tag{35}$$

Using

$$\begin{aligned}\cos(\omega_m(s-t)) = \cos(\omega_m s)\cos(\omega_m t) \\ + \sin(\omega_m s)\sin(\omega_m t),\end{aligned}$$

define

$$\begin{aligned}\varphi_{\mathcal{K}}(t) = \big[1,\, \sin(\omega_1 t),\, \cos(\omega_1 t),\, \ldots, \\ \sin(\omega_K t),\, \cos(\omega_K t)\big]^\top \in \mathbb{R}^p,\end{aligned}$$
$$p = 2K + 1.$$

With an appropriate nonnegative diagonal matrix Λ,

$$k_{\mathcal{K}}(s,t) = \varphi_{\mathcal{K}}(s)^\top \Lambda \varphi_{\mathcal{K}}(t).$$

Let

$$\Phi_c = \begin{bmatrix} \varphi_{\mathcal{K}}(1)^\top \\ \vdots \\ \varphi_{\mathcal{K}}(C)^\top \end{bmatrix}, \qquad \Phi_f = \begin{bmatrix} \varphi_{\mathcal{K}}(C+1)^\top \\ \vdots \\ \varphi_{\mathcal{K}}(C+H)^\top \end{bmatrix},$$

and $U_{cf} = [\Phi_c^\top, \Phi_f^\top]^\top$. Adding a diagonal innovation floor $d > 0$ yields

$$K_{cf} = dI_{C+H} + U_{cf}\Lambda U_{cf}^\top. \tag{36}$$

The context, cross, and future blocks are

$$K_{cc} = dI_C + \Phi_c \Lambda \Phi_c^\top, \tag{37}$$
$$K_{cf}^{(c,f)} = \Phi_c \Lambda \Phi_f^\top, \tag{38}$$
$$K_{ff} = dI_H + \Phi_f \Lambda \Phi_f^\top. \tag{39}$$

The non-diagonal component has rank at most p. With the implemented $K = 4$, $p = 9$.

Positive semidefiniteness and strict positive definiteness. For any $x \in \mathbb{R}^{C+H}$,

$$x^\top K_{cf} x = d\|x\|_2^2 + x^\top U_{cf}\Lambda U_{cf}^\top x \tag{40}$$
$$= d\|x\|_2^2 + \|\Lambda^{1/2} U_{cf}^\top x\|_2^2. \tag{41}$$

Thus $K_{cf} \succeq 0$ for $d \geq 0$, $\Lambda \succeq 0$, and $K_{cf} \succ 0$ whenever $d > 0$. No full-column-rank assumption on U_{cf} is required for strict positive definiteness when the diagonal floor is positive.

Woodbury inverse. If $d > 0$ and the active part of Λ is positive definite, the matrix inversion lemma gives

$$\begin{aligned} A &= \Lambda^{-1} + d^{-1} U^\top U, \\ (dI + U\Lambda U^\top)^{-1} &= d^{-1} I - d^{-2} U A^{-1} U^\top. \end{aligned} \tag{42}$$

The expensive solve is therefore moved from the time-grid dimension to the feature dimension p.

A.6.3 WHY RFFT AND TOPK ARE DERIVED CHOICES

Center the residual:

$$\widetilde{r}^{(j)} = r^{(j)} - \bar{r}^{(j)} \mathbf{1}_C.$$

Let F_C be the unitary DFT and

$$\widehat{r} = F_C \widetilde{r}^{(j)}.$$

Parseval's identity gives

$$\|\widetilde{r}^{(j)}\|_2^2 = \sum_{k=0}^{C-1} |\widehat{r}_k|^2.$$

For a conjugate-symmetric support S compatible with a real signal, let P_S be the orthogonal projection onto the associated Fourier modes. Then

$$\|\widetilde{r}^{(j)} - P_S\widetilde{r}^{(j)}\|_2^2 = \sum_{k \notin S} |\widehat{r}_k|^2. \tag{43}$$

Therefore, among supports of a fixed size,

$$\arg\max_{|S|=K} \sum_{k\in S} |\widehat{r}_k|^2 = \arg\min_{|S|=K} \|\widetilde{r}^{(j)} - P_S\widetilde{r}^{(j)}\|_2^2. \tag{44}$$

TopK periodogram selection is thus a best-energy K-term projection in the orthogonal Fourier dictionary under the stated DFT convention.

For a one-sided real FFT, non-DC and non-Nyquist bins represent conjugate pairs and their energy must be counted accordingly. If the implementation uses the standard one-sided energy normalization, the exact statement (44) applies. If it ranks raw one-sided magnitudes without pair correction, the universally safe description is *dominant-energy Fourier support selection*.

The same operation has an empirical-Bayes interpretation. In a stationary Gaussian model, Fourier coefficients are approximately decorrelated and the spectral density controls their variances. The context periodogram is a plug-in estimate of this local power. Selecting its dominant bins is a hard-support empirical estimate of the active finite spectral measure.

A.6.4 Least squares as projection and maximum likelihood

For the selected frequencies, the code computes

$$\widehat{\beta} \in \arg\min_{\beta} \|\Phi_c\beta - r^{(j)}\|_2^2. \tag{45}$$

Every minimizer satisfies the normal equations

$$\Phi_c^\top(\Phi_c\widehat{\beta} - r^{(j)}) = 0. \tag{46}$$

Thus $\Phi_c\widehat{\beta}$ is the orthogonal projection of the residual onto $\mathrm{Col}(\Phi_c)$. If Φ_c has full column rank,

$$\widehat{\beta} = (\Phi_c^\top\Phi_c)^{-1}\Phi_c^\top r^{(j)};$$

otherwise the canonical minimum-norm solution is

$$\widehat{\beta} = \Phi_c^\dagger r^{(j)}.$$

Under the Gaussian observation model

$$r^{(j)} = \Phi_c\beta + \eta, \qquad \eta \sim \mathcal{N}(0, dI_C),$$

least squares is also the maximum-likelihood estimator of β.

A.6.5 The formal WSP posterior and its coefficient-space form

Consider the finite-feature Gaussian model

$$\beta \sim \mathcal{N}(0, \Lambda), \tag{47}$$

$$r^{(j)} = \Phi_c\beta + \eta_c, \qquad \eta_c \sim \mathcal{N}(0, dI_C), \tag{48}$$

$$R_f^{(j)} = \Phi_f\beta + \eta_f, \qquad \eta_f \sim \mathcal{N}(0, dI_H), \tag{49}$$

with mutually independent terms. Completing the square gives

$$\Sigma_{\beta|r} = (\Lambda^{-1} + d^{-1}\Phi_c^\top\Phi_c)^{-1}, \tag{50}$$

$$\mu_{\beta|r} = d^{-1}\Sigma_{\beta|r}\Phi_c^\top r^{(j)} \tag{51}$$

$$= (\Phi_c^\top\Phi_c + d\Lambda^{-1})^{-1}\Phi_c^\top r^{(j)}. \tag{52}$$

The posterior mean is the unique minimizer of

$$\|\Phi_c\beta - r^{(j)}\|_2^2 + d\beta^\top \Lambda^{-1}\beta. \tag{53}$$

The future predictive mean and covariance are

$$\mathbb{E}[R_f^{(j)} \mid r^{(j)}] = \Phi_f \mu_{\beta|r}, \tag{54}$$

$$\mathrm{Cov}(R_f^{(j)} \mid r^{(j)}) = dI_H + \Phi_f \Sigma_{\beta|r}\Phi_f^\top. \tag{55}$$

The same mean in covariance-space notation is

$$\Phi_f \Lambda \Phi_c^\top (dI_C + \Phi_c \Lambda \Phi_c^\top)^{-1} r^{(j)}. \tag{56}$$

Applying (42) to the context covariance shows that (56) is exactly equal to $\Phi_f \mu_{\beta|r}$. The formal GP conditioning and the coefficient-space regularized regression are therefore the same calculation in two coordinate systems.

A.6.6 The implementation as the flat-prior limit

Set $\Lambda = \tau^2 I_p$ and $\gamma = d/\tau^2$. Then

$$\mu_\gamma = (\Phi_c^\top \Phi_c + \gamma I_p)^{-1} \Phi_c^\top r^{(j)}.$$

Let $\Phi_c = Q\Sigma V^\top$ be an SVD. Then

$$\mu_\gamma = V \,\mathrm{diag}\left(\frac{\sigma_\ell}{\sigma_\ell^2 + \gamma}\right) Q^\top r^{(j)}.$$

As $\gamma \downarrow 0$, each nonzero singular component converges to $1/\sigma_\ell$, while a zero singular component remains zero. Hence

$$\lim_{\gamma \downarrow 0} \mu_\gamma = \Phi_c^\dagger r^{(j)}. \tag{57}$$

Therefore the deterministic implementation

$$\Phi_f \widehat{\beta}(r^{(j)})$$

is the flat-prior, zero-ridge limit of the formal WSP posterior predictive mean after the active Fourier support has been selected from the context.

When the selected frequencies are exact DFT-grid bins, the real sine/cosine coefficient vector is related to the complex RFFT coefficients by a fixed invertible real-linear coordinate transform, apart from the standard DC and Nyquist conventions. Thus RFFT support selection and real-basis least squares are two compatible representations of the same finite Fourier projection.

A.7 The implementation and formal covariance share one low-rank Fourier family

Define the continuous selected family

$$\mathcal{S}_\mathcal{K} = \mathrm{span}\{1, \sin(\omega_1 t), \cos(\omega_1 t), \ldots, \sin(\omega_K t), \cos(\omega_K t)\}.$$

Evaluation on the future grid produces

$$E_f(\mathcal{S}_\mathcal{K}) = \mathrm{Col}(\Phi_f).$$

Since the implementation computes

$$z^{(j)}_{\mathrm{WSP-res}} = \Phi_f \widehat{\beta}(r^{(j)}) + \epsilon^{(j)}_{\mathrm{smooth}}, \tag{58}$$

its deterministic component satisfies

$$\Phi_f \widehat{\beta}(r^{(j)}) \in \mathrm{Col}(\Phi_f). \tag{59}$$

Suppose the formal feature matrix differs by a nonsingular coefficient reparameterization,

$$U(r^{(j)}) = \Phi_f S, \qquad S \in \mathbb{R}^{p\times p} \text{ invertible.}$$

This covers column scaling, normalization, and permutation. Then

$$\mathrm{Col}(U(r^{(j)})) = \mathrm{Col}(\Phi_f),$$

and (59) implies

$$\Phi_f \widehat{\beta}(r^{(j)}) \in \mathrm{Col}(U(r^{(j)})). \tag{60}$$

For the formal low-rank covariance action

$$L(r) = U(r)\Lambda(r)U(r)^\top,$$

we have for every vector x,

$$L(r)x = U(r)(\Lambda(r)U(r)^\top x),$$

so

$$\mathrm{Col}(L(r)) \subseteq \mathrm{Col}(U(r)). \tag{61}$$

If every active spectral weight is strictly positive, let $A = U\Lambda^{1/2}$. On the active coordinates, $\Lambda^{1/2}$ is invertible, so $\mathrm{Col}(A) = \mathrm{Col}(U_{\mathrm{act}})$. Moreover, $\ker(AA^\top) = \ker(A^\top)$, because

$$x^\top AA^\top x = \|A^\top x\|_2^2.$$

Taking orthogonal complements gives $\mathrm{Col}(AA^\top) = \mathrm{Col}(A)$. Hence

$$\mathrm{Col}(U\Lambda U^\top) = \mathrm{Col}(U_{\mathrm{act}}). \tag{62}$$

If all selected weights are positive, $U_{\mathrm{act}} = U$; if some are zero, equality holds only for the active columns, while (61) remains valid for the full selected set.

Combining (60) and (62), the implemented future mean, the formal posterior mean, and the active low-rank covariance component all live in the same context-selected Fourier feature family.

Why the implemented covariance is not generally K_{WSP}. Conditional on $r^{(j)}$, the fitted coefficient $\widehat{\beta}(r^{(j)})$ is deterministic. If $\epsilon_{\mathrm{smooth}} = B\xi$, $\xi \sim \mathcal{N}(0, I)$, then

$$z^{(j)}_{\mathrm{WSP-res}} \mid r^{(j)} \sim \mathcal{N}(\Phi_f \widehat{\beta}(r^{(j)}), BB^\top).$$

If the smooth term is deterministic, the conditional covariance is zero. Neither case is generally equal to $dI + U\Lambda U^\top$. An exact formal posterior draw would require

$$\beta_{\mathrm{post}} = \mu_{\beta|r} + L_\beta \xi_\beta, \qquad L_\beta L_\beta^\top = \Sigma_{\beta|r},$$

followed by

$$R_f = \Phi_f \beta_{\mathrm{post}} + \sqrt{d}\,\xi_f.$$

The current implementation replaces coefficient sampling by a point estimate and adds its own smooth perturbation. The strongest accurate statement is therefore:

> The implemented WSP residual is a context-fitted low-rank Fourier extrapolation derived as the weak-prior predictive-mean limit of a finite Woodbury-spectral Gaussian residual model. It shares the active Fourier feature space with the formal low-rank covariance, but it is not generally an exact sample from that covariance.

A.8 Computational interpretation of the Woodbury parameterization

Let $p = 2K + 1$. After selecting the active support, the main operations are

RFFT :	$O(C \log C)$,
TopK :	$O(C \log K)$ or $O(C)$,
coefficient solve :	$O(Cp^2 + p^3)$,
future evaluation :	$O(Hp)$.

The formal posterior mean uses the same $p \times p$ coefficient-space system. By contrast, a naive dense solve on the combined context-future grid scales cubically in the grid dimension. Thus the word *Woodbury* refers to the low-rank coefficient-space structure exposed by (42); the implementation does not need to form or invert the dense matrix K_{cf}.

This argument does not imply that WSP is faster than the pure OU endpoint. OU is the cheaper anchor. The computational claim is that adding a context-adaptive spectral residual remains tractable and avoids a dense GP conditioning path. With $K = 4$, the coefficient dimension is only $p = 9$.

A.9 OU-WSP ENDPOINT, CALIBRATION, AND FLOW PUSHFORWARD

The endpoint used by the main method is

$$z_0^{(j)} = \mathrm{Calib}\left(z_{\mathrm{OU}}^{(j)} + \alpha z_{\mathrm{WSP-res}}^{(j)}; c^{(j)}\right). \tag{63}$$

The case $\alpha = 0$ is the calibrated OU endpoint. Positive α shrinks or amplifies a residual that remains in the selected Fourier family; scaling does not change that subspace. The value of α, the choice $K = 4$, and the exact smoothing operator are controlled modeling choices rather than uniquely forced constants.

Let $T_{\theta,c^{(j)},j}$ denote the exact ODE solution map at time one for

$$\frac{dx_t^{(j)}}{dt} = v_\theta(c^{(j)}, x_t^{(j)}, t, j).$$

If $q_0^{(j)}(\cdot \mid c^{(j)}, j)$ is the endpoint distribution after the deterministic calibration map, then the exact generated distribution is the pushforward

$$\widehat{q}_1^{(j)} = (T_{\theta,c^{(j)},j})_{\#} q_0^{(j)}.$$

The endpoint is therefore a structured source distribution, not a claim that the source itself equals the final ICU future distribution.

The implementation uses explicit Euler steps

$$x_{n+1}^{(j)} = x_n^{(j)} + \Delta t\, v_\theta(c^{(j)}, x_n^{(j)}, t_n, j).$$

Under the standard local Lipschitz and smoothness assumptions on the velocity, Euler integration converges to the ODE solution as $\Delta t \to 0$. The paper's numerical trajectory is consequently an approximation to the mathematical pushforward above.

A.10 BLOOD-PRESSURE SORTING REPAIR

For each generated future hour, let

$$b = (b_s, b_m, b_d) = (\widehat{\mathrm{SBP}}, \widehat{\mathrm{MBP}}, \widehat{\mathrm{DBP}})$$

denote the generated systolic, mean, and diastolic blood-pressure triple. The physiological relation is the closed convex cone

$$\mathcal{C}_{\mathrm{BP}} = \{z \in \mathbb{R}^3 : z_s \geq z_m \geq z_d\}.$$

The raw generator may produce $b \notin \mathcal{C}_{\mathrm{BP}}$. A principled order-restricted regression repair is the Euclidean isotonic projection

$$\Pi_{\mathrm{iso}}(b) = \arg\min_{z \in \mathcal{C}_{\mathrm{BP}}} \|z - b\|_2^2.$$

For the three-channel BP ordering this projection is the decreasing isotonic-regression solution. It can be computed by the pool-adjacent- violators algorithm (PAVA): start with three singleton blocks

$$[b_s], \qquad [b_m], \qquad [b_d],$$

scan adjacent blocks from left to right, and whenever the decreasing order is violated, replace the violating adjacent blocks by their weighted average. The procedure is repeated until the block means satisfy a nonincreasing order. For example, if $b_s < b_m$ and the pooled mean

$$\bar{b}_{sm} = \frac{b_s + b_m}{2}$$

still satisfies $\bar{b}_{sm} \geq b_d$, then

$$\Pi_{\mathrm{iso}}(b) = (\bar{b}_{sm}, \bar{b}_{sm}, b_d).$$

If both adjacent inequalities require pooling, all three entries may be replaced by their common mean. Thus isotonic regression enforces BP ordering by the smallest squared-error value adjustment, but it generally changes the numerical BP values.

The original report described a value-multiset-preserving operation, called BP sorting repair. The following proves properties of that mathematical operation; it does not identify the recovered isotonic code as sorting. Let

$$b_{(1)} \geq b_{(2)} \geq b_{(3)}$$

be the same three generated values sorted in descending order. The sorting repair map is

$$S_{\mathrm{BP}}(b) = (b_{(1)}, b_{(2)}, b_{(3)}),$$

and the repaired triple is assigned as

$$(\widehat{\mathrm{SBP}}, \widehat{\mathrm{MBP}}, \widehat{\mathrm{DBP}}) \leftarrow S_{\mathrm{BP}}(b).$$

By construction,

$$\widehat{\mathrm{SBP}} \geq \widehat{\mathrm{MBP}} \geq \widehat{\mathrm{DBP}}.$$

Sorting repair preserves the multiset of the three generated values and therefore preserves their sum, mean, minimum, and maximum. It is a permutation-based clinical safeguard rather than a Euclidean projection onto $\mathcal{C}_{\mathrm{BP}}$. In general,

$$S_{\mathrm{BP}}(b) \neq \Pi_{\mathrm{iso}}(b).$$

For instance, if $b = (80, 100, 70)$, then sorting repair gives

$$S_{\mathrm{BP}}(b) = (100, 80, 70),$$

whereas decreasing isotonic regression gives

$$\Pi_{\mathrm{iso}}(b) = (90, 90, 70).$$

The former preserves the generated value multiset but reassigns channel order; the latter preserves channel identity but changes values through pooling. This distinction motivates separating raw generation, sorting, isotonic projection, and filtering. The empirical-policy audit is documented below. BP filtering keeps only complete future windows whose every BP triple already belongs to $\mathcal{C}_{\mathrm{BP}}$, while sorting repair keeps the full synthetic budget and enforces the clinical ordering after generation.

B Protocol, Implementation, and Result-Preservation Notes

B.1 Original and additional protocols

All original main-text tables and non-framework result figures are retained numerically or byte-for-byte. Preservation does not remove the original protocol's limitations. The source audit found that stored contexts exactly matched filling the complete 48-hour window; context-only reconstruction changed 226/26/57 train/validation/test windows in MIMIC-III and 519/88/157 in MIMIC-IV. The corresponding full split sizes were 15,413/2,201/4,405 and 24,845/3,549/7,099. Patient-ID intersections were zero, but that does not prevent within-window lookahead. These are full-artifact counts, not measured effects on every historical training subset. Historical generators used filled arrays, whereas the recovered strong evaluator used raw arrays with global-median filling; their input representations must not be conflated. The performance effect has not been quantified by a complete boundary-safe downstream rerun.

The additional protocol reconstructs each context without its future, derives future labels by forward filling, and fits fallback/normalization statistics only on the fixed 6,000 training rows. Raw values and observation masks are retained. Eight fresh generators and 16 train-pair artifacts were completed; the latter each contain 6,000 generated futures. The exported diagnostic field `windows` is 4,405 for MIMIC-III and 6,000 for MIMIC-IV; it is not the original full-cohort size. These are unpaired population comparisons of training-context synthesis to held-out targets. No 256-context $\times$ 8-sample held-out diagnostic, new GRU/TCN utility, or new feature-FID evaluation was completed. The new single-seed scores cannot be pooled with original-protocol results or used to validate the old downstream gains.

B.2 METRIC AND POLICY ATTRIBUTION

The recovered deterministic forecasting evaluator computes Total as MASE+ND+NRMSE. The original table's CRPS+MASE description is corrected without changing any numerical entry. Real-100%, Real-10%, Bootstrap, and Jitter remain historical labels; the 4,800-window synthetic artifact count does not redefine their real-data access. Reference, wrapper, and style labels identify distinct project implementations and do not certify common tuning budgets or official-method replication.

All five values in the original Gaussian fidelity rows match BP-filtered summaries: 610 of 2,400 windows retained for MIMIC-III and 400 of 2,400 for MIMIC-IV. They are marked with a dagger in the retained table. The full operator/export chains for other fidelity rows remain unresolved. Consequently those values are not a uniformly raw, repaired, or sample-count-matched ranking. Original image labels are preserved to avoid silently altering historical evidence.

The recovered `project_descending3` computes L2 isotonic projection: $(80, 100, 70)$ maps to $(90, 90, 70)$, not the descending sort $(100, 80, 70)$. The original BP plot's embedded "Sorting repair" label is retained as an artifact label, not endorsed as implementation verification. Zero violation after either operator is deterministic; filtering additionally changes the accepted population and available sample budget. The new table has no repair or filtering and therefore provides raw cross-source rates without this policy ambiguity.

B.3 CENTRALIZED SPECTRAL SETTINGS AND CALIBRATION

For the original utility row, the reported family configurations are MIMIC-III GRU/TCN: $\alpha = 0$; MIMIC-IV GRU: 0.03; MIMIC-IV TCN: 0.01. These are not a single globally fixed generator configuration, and the retained provenance does not justify inventing a common validation-selection procedure. Other artifact settings cannot be inferred solely from the family name. The additional study fixes positive residual strength to 0.03, with the WSP-series zero-weight path as its control; it does not retroactively change any original setting.

Let $\mathcal{C}_\lambda(g; c)$ denote the calibration map in the main text. The recovered source paths distinguish plain OU, anchor-only OU with source weight 0.25, and WSP-series OU with source weight 0.30. For the latter,

$$z_{\mathrm{res}} = \mathcal{C}_{0.20}(R_{\mathrm{pre}}; r), \qquad z_0 = \mathcal{C}_{0.30}(z_{\mathrm{OU}} + \alpha z_{\mathrm{res}}; c).$$

Thus setting the residual weight to zero does not disable source calibration. The additional Gaussian-cal control uses $\mathcal{C}_{0.30}(\epsilon; c)$, where $\epsilon \sim \mathcal{N}(0, I)$; it is intentionally separate from standard Gaussian FM. Output calibration is a later operation, evaluated at weights 0 and 0.5. Removing only that operation is not a no-calibration training ablation.

The Fourier regression component is $m_r = \Phi_f \hat{\beta}(r)$. In the recovered residual sampler, the pre-calibration wrapper is

$$R_{\mathrm{pre}} = r_L \mathbf{1} + 0.8\big(m_r - \mathrm{median}(r)\mathbf{1}\big) + 0.04 \max\{\mathrm{Std}(r), 0.05\} A_3 \eta, \qquad \eta \sim \mathcal{N}(0, I_H),$$

where A_3 is the length-three zero-padded moving average. This separates the proved Fourier core from reanchoring, shrinkage, stochastic noise, and calibration. Same-feature-space and weak-prior posterior-mean arguments do not imply equality between this full sampler's covariance and K_{WSP}.

B.4 WHAT THE ADDED COMPARISON RESOLVES AND DOES NOT RESOLVE

The added table directly compares raw BP violations across standard Gaussian, calibrated Gaussian, OU control, and WSP, and separates output-calibration settings at a fixed source. It shows decreases in Corr0, Lag1, and BP violation for all eight source/dataset pairs, increases in log-PSD for all eight, and four improvements/four deteriorations in W1. This establishes a descriptive calibration tradeoff in the supplied batch, not an overall winner. Improvements in aggregate cross-channel correlations need not indicate patient-level joint conditioning. The study still lacks a directly implemented TSFlow comparator, multi-seed uncertainty, clean downstream utility, source-only/global-context ablations, and aggregate conditional forecast evaluation on unseen contexts.