Real-Time Causal Dynamics on Geometric Latent Spaces for Precision Therapeutic Monitoring

Anonymous authors
Paper under double-blind review

Abstract

Despite impact in early-stage drug discovery, the exploration of artificial intelligence in therapeutic drug monitoring, a key bottleneck in patient care, has remained limited. We present Real-Time Causal Geometric Flows (RT-CGF), a geometry-aware, causal, and adaptive framework for therapeutic drug monitoring of narrow-therapeutic-index drugs, addressing a longstanding gap in translating from discovery to deployment. Unlike traditional models that rely on fixed, population-level dynamics, RT-CGF redefines TDM as a dynamical system on learned Finsler manifolds, where pharmacokinetic trajectories evolve as geodesics under patient-specific geometric constraints. This previously unexplored formulation enables real-time adaptation to physiological shifts (e.g, acute kidney injury) while preserving causal invariance. Extending stochastic intervention calculus for continuous-time counterfactuals and persistent homology regularization for topological consistency, RT-CGF achieves fewer dosing errors and improved AUC in adverse event detection on representative clinical datasets. Beyond its methodological novelty, RT-CGF introduces an interesting possibility for precision medicine by harmonizing Finsler-geometric dynamical systems, counterfactual reasoning, and persistent homology constraints, a convergence that transcends the limitations of static, population-level models to enable adaptive, patient-specific policy frameworks with implications for safety, assurance and regulation efforts for digital drug development.

1 Introduction

Machine learning has accelerated early-stage drug discovery, from molecular design to target binding (22), but pharmaceutical development presents downstream challenges such as safety assessment, metabolism, and dosage individualization, that remain comparatively under-automated (6). Central among these is therapeutic drug monitoring (TDM): maintaining drug concentrations within narrow therapeutic windows where the margin between efficacy and toxicity is small (30; 10). TDM is essential for narrow-therapeutic-index drugs such as vancomycin, where overexposure risks nephrotoxicity, and tacrolimus, where subtherapeutic levels precipitate organ rejection while supratherapeutic concentrations cause neurotoxicity (1). In critical care and transplant settings, acute physiological shifts - renal injury, hepatic dysfunction - make static, population-averaged dosing protocols inadequate (31). Existing TDM methods fall short of these demands. Bayesian forecasting depends on population priors that capture individual variability poorly, particularly during acute pathological transitions (24). Compartmental and PBPK models are grounded in pharmacokinetic theory but assume time-invariant parameters and prespecified nonlinear functional forms (e.g., Michaelis–Menten kinetics) that approximate fluctuating clearance imperfectly (21). Modern machine learning offers flexibility but lacks geometric consistency and causal interpretability, yielding trajectories that may violate biological constraints or fail to generalize under interventions. We introduce Real-Time Causal Geometric Flows (RT-CGF), a framework that formalizes pharmacokinetic trajectories as geodesics on a learned Finsler manifold whose metric tensor encodes patient-specific clearance dynamics, which to our knowledge, is

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.

the first exploration of Finsler geometry (2) to predictive formulations in pharmacological modeling. Where standard PK models require predefined nonlinear forms for processes such as saturable transport, RT-CGF encodes them as a learned geometric prior whose specific form is determined by the data. Combined with a do-calculus-inspired regularizer for continuous-time interventions, RT-CGF simulates counterfactual dosing adjustments by answering 'What if we modify the dose now?' without retraining, while topological consistency is enforced through persistent homology regularization, which suppresses trajectory artifacts and flags anomalies predictive of adverse events 12 hours earlier than baseline methods. RT-CGF thus unifies geometry, causality, and topology in a single tractable formalism, providing patient-specific monitoring recommendations that respect physiological constraints and establishing a framework for interventional dynamical systems in precision medicine, closed-loop therapeutic control and alignment for computational drug development.

Prior Art. Therapeutic drug monitoring has traditionally relied on compartmental pharmacokinetic models (26)(13) and Bayesian hierarchical models (12), both of which assume fixed population-level parameters and are ill-suited to patient-specific physiology or the sparse, noisy measurements typical of clinical TDM. Neural differential equations (9) and their extensions to stochastic dynamics (11) and physics-informed architectures (28) have since enabled data-driven, continuous-time PK modeling, including applications to time series with static baseline covariates (5), but these approaches generally lack explicit biological constraints or causal interpretability and often fail to generalize under unseen interventions. Hybrid PBPK-deep learning approaches improve biological plausibility but remain computationally expensive and adapt poorly to abrupt physiological shifts such as acute kidney injury (15)(18). Gaussian process-based ODEs offer principled uncertainty quantification (4) but remain similarly limited in real-time adaptability and causal reasoning. Topological data analysis has proven useful for characterizing biological time series (8), but its use as a constraint within clinical decision-support for TDM specifically has not been explored. Within causal inference, static clinical prediction is comparatively well studied (25; 27), and continuous-time causal modeling has recently been addressed by TE-CDE (29) and BNCDE (17), which estimate treatment effects from irregularly sampled observational data via neural controlled differential equations (19). RT-CGF differs from these methods in mechanism rather than problem setting: rather than inferring counterfactual outcomes from observation histories via attention over irregular time, RT-CGF learns an explicit geometric dynamical law and enforces causal invariance of interventions to that law through an orthogonality constraint, additionally admitting real-time counterfactual optimization. Geometric deep learning has largely emphasized Euclidean or Riemannian structure (7); we instead adopt Finsler geometry to capture the direction- and speed-dependent asymmetries characteristic of clearance dynamics. Persistent homology has found use in omics analysis (16), but its role in pharmacology remains comparatively limited, and differential geometry more broadly has seen restricted application in pharmacokinetics (3), with prior work confined to Riemannian manifolds rather than the more general Finsler structure asymmetric clearance requires. In this paper, we propose a unified geometric-causal framework that enhances pharmacokinetic modeling by enabling adaptive, interpretable, and topologically consistent predictions with novel counterfactual reasoning capabilities for clinical decision support settings.

2 Methodology

Real-Time Causal Geometric Flows introduces a geometry-aware, causality-preserving dynamical system for therapeutic drug monitoring, unifying Finsler manifolds, stochastic do-calculus, and algebraic topology into a differentiable framework. This section formalizes the mathematical foundations of geodesic PK trajectories, counterfactual intervention kernels, and topological regularization, now implemented with Temporal Geometric Transformers (TGT) and Cross-Modal Intervention Transformers (CMIT) to capture long-range dependencies and non-Markovian dynamics critical for clinical applications.

2.1 Finsler-Geometric Transformers for Patient-Specific Clearance

Traditional TDM models implicitly assume that pharmacokinetic trajectories evolve in a Euclidean space with linear or Michaelis–Menten dynamics. In reality, processes such

as renal filtration, hepatic metabolism, and drug-target binding impose non-Euclidean constraints: the relationship between concentration and clearance is not a scalar function but a curved manifold shaped by organ-specific boundaries, saturable transport, and patient-specific variability. We therefore formulate the PK system on a **Finsler manifold** $\mathcal{M}$, whose metric $F : T\mathcal{M} \to \mathbb{R}$ defines a patient-specific geometry over the state space. Here $\mathbf{z}_t = [C(t), \text{Clearance}(t)]^T \in \mathcal{M}$ and $\dot{\mathbf{z}}_t \in T_{\mathbf{z}_t}\mathcal{M}$, and the Finsler metric $F(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)$ is parameterized by a **Temporal Geometric Transformer (TGT)**: sinusoidal positional encodings preserve temporal order under irregular sampling, while L multi-head self-attention layers (H heads) capture non-local temporal interactions, e.g., a dose at time t influencing clearance at $t + 12$ hours.

Why Finsler rather than Riemannian geometry? A Riemannian metric depends only on position $\mathbf{z}_t$, so its quadratic form $g_{ij}(z)\dot{z}^i\dot{z}^j$ is necessarily even in $\dot{\mathbf{z}}_t$: motion costs the same in every direction at every speed. PK clearance is neither direction- nor speed-symmetric: (i) renal tubular secretion is saturable and *unidirectional*—its rate depends on the sign and magnitude of $\dot{C}$, not only on C; and (ii) Michaelis–Menten-type clearance is intrinsically *velocity-dependent*, with effective damping of $\dot{\mathbf{z}}_t$ scaling sublinearly with $|\dot{\mathbf{z}}_t|$ near saturation. A Finsler metric is precisely a position-, direction-, and speed-dependent norm, so it represents both asymmetries in a single scalar function; a Riemannian metric cannot. The TGT output passes through a constrained head ensuring the learned metric is well behaved: (i) *positive homogeneity* ($F(\mathbf{z}_t, \lambda\dot{\mathbf{z}}_t; \theta) = \lambda F(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)$, $\lambda > 0$) is enforced *architecturally* by a bias-free homogeneous pathway—a softplus alone guarantees positivity, not homogeneity, so no penalty term is needed; (ii) *positive definiteness* of the fundamental tensor $g_{ij} = \frac{1}{2}\partial^2 F^2/\partial\dot{z}^i\partial\dot{z}^j$ holds by construction via the parameterization $F^2 = \dot{\mathbf{z}}_t^\top A(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)^\top A(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)\,\dot{\mathbf{z}}_t$ with A learned by the TGT; and (iii) *convexity* of F^2 in $\dot{\mathbf{z}}_t$ follows for this quadratic form. The metric further encodes physiological boundaries (non-negative concentrations, saturable clearance) with curvature increasing where clearance approaches saturation. The dynamics are a first-order system on the tangent bundle $T\mathcal{M}$, with state $(\mathbf{z}_t, \dot{\mathbf{z}}_t)$:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{z}_t \\ \dot{\mathbf{z}}_t\end{bmatrix} = \begin{bmatrix}\dot{\mathbf{z}}_t \\ -\Gamma(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)\dot{\mathbf{z}}_t - g^{-1}(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)\,\nabla_{\mathbf{z}}\left(F^2(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)/2\right)\end{bmatrix},$$

where Γ denotes the Chern–Rund connection coefficients of F (fixed throughout; the Cartan and Berwald connections agree on these spray terms for our parameterization), and $g^{-1}\nabla_{\mathbf{z}}(F^2/2)$ is the covector gradient of the Lagrangian with its index raised by the inverse fundamental tensor, so both acceleration terms are genuine tangent vectors. The potential term is not part of the geodesic spray: the dynamics are best understood as *geodesic flow under an external potential*, in the spirit of Lagrangian mechanics on Finsler manifolds, rather than pure geodesic motion. Patient-specific constraints are thus encoded in the metric itself rather than imposed through hand-specified functional forms, letting the data determine the geometry. Architecturally, the TGT comprises an input embedding (linear projection of $(\mathbf{z}_t, \dot{\mathbf{z}}_t)$ into d dimensions with sinusoidal positional encodings), a transformer encoder ($L = 6$ layers, $H = 8$ heads, each with layer normalization and residual connections), and the geometric output head above. This design learns patient-specific manifolds from sparse, irregularly sampled TDM data while respecting global physiological constraints.

2.2 Stochastic Intervention Calculus with Cross-Modal Transformers

Real-time dosing must adapt from sparse, noisy measurements while preserving the PK system's causal structure; unlike compartmental/PBPK counterfactual simulation within fixed structures, our framework enables counterfactual reasoning inside a learned geometric dynamical system. To this end, we introduce Stochastic Intervention Calculus (SIC) with a Cross-Modal Intervention Transformer (CMIT). The CMIT is designed to model the effect of exogenous interventions (e.g, dose adjustments) on the system's dynamics while preserving causal invariance. The CMIT takes as input PK State Embeddings, or Transformer-encoded representations of $(\mathbf{z}_t, \dot{\mathbf{z}}_t)$ from the TGT; and a learned embedding for the exogenous intervention $\pi(t)$ (e.g, dose adjustment) concatenated with the state embeddings. The intervened system is described by the stochastic differential equation on the tangent bundle:

$$d\begin{bmatrix}\mathbf{z}_t \\ \dot{\mathbf{z}}_t\end{bmatrix} = \begin{bmatrix}\dot{\mathbf{z}}_t \\ -\Gamma(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)\dot{\mathbf{z}}_t - g^{-1}\nabla_{\mathbf{z}}(F^2/2) + h(\mathbf{z}_t, \dot{\mathbf{z}}_t, t; \pi(t))\end{bmatrix} dt + \begin{bmatrix}\mathbf{0} \\ g_v(\mathbf{z}_t, \dot{\mathbf{z}}_t, t; \theta)\end{bmatrix} \circ dW_t.$$

Here, the diffusion matrix is structurally masked so that stochastic fluctuations enter only through the velocity block g_v; noise applied directly to position $\mathbf{z}_t$ would make trajectories a.e. non-differentiable, contradicting the well-defined velocity term used in the drift. The intervention kernel h is parameterized by the CMIT and models the effect of the intervention $\pi(t)$ on the system's dynamics. The CMIT uses crossmodal attention, a multi-head attention mechanism that attends between the PK state embeddings and the intervention tokens, allowing the model to dynamically adjust its response to interventions based on the current state. Treating the PK system as a continuous-time structural causal model whose drift f encodes the mechanism, with dose entering exogenously via $\pi(t)$, the constraint below intends the learned intervention kernel h to act only in directions orthogonal to the state-Jacobian column space of f. A soft attention bias (weighted by λ_1) pushes weights toward satisfying:

$$\frac{\partial h}{\partial \pi} \perp \operatorname{col}\left(\frac{\partial f}{\partial \mathbf{z}}, \frac{\partial f}{\partial \dot{\mathbf{z}}}\right),$$

where $f(\mathbf{z}_t, \dot{\mathbf{z}}_t, t; \theta) = \begin{bmatrix} \dot{\mathbf{z}}_t & -\Gamma(\mathbf{z}_t, \dot{\mathbf{z}}_t; \theta)\dot{\mathbf{z}}_t - g^{-1}\nabla_{\mathbf{z}}(F^2/2) \end{bmatrix}^T$ is the drift of the unforced system, and $\operatorname{col}(A)$ denotes the column space of matrix A. This condition approximates, via a soft λ_1-weighted penalty rather than a hard constraint, the causal-invariance property of interventions in a structural causal model. Since f maps $\mathbb{R}^{2n} \to \mathbb{R}^n$, the constraint $\partial h/\partial \pi \perp \operatorname{col}(J)$ is a statement in the drift's output space: to first order, the intervention enters the dynamics only in directions to which the unforced mechanism f is insensitive, so perturbing π does not perturb the mechanism's response along the observed trajectory. This is the continuous-time analogue of an intervention that modifies the exogenous input of an SCM without rewriting its structural equations. The Cross-Modal Intervention Transformer begins with a linear projection of concatenated PK state embeddings and intervention tokens into a shared embedding space, followed by 4 layers of multihead attention (with 4 heads each) using attention biased toward the intervention-orthogonal complement and concludes with a linear output layer that generates the intervention kernel $h(\mathbf{z}_t, \dot{\mathbf{z}}_t, t; \pi(t))$. For real-time dosing adaptations, we formulate the problem of finding the optimal intervention π^* as a counterfactual optimization problem. Given a sequence of sparse measurements $\{(\mathbf{z}_{t_1}, \dot{\mathbf{z}}_{t_1}), (\mathbf{z}_{t_2}, \dot{\mathbf{z}}_{t_2}), \ldots, (\mathbf{z}_{t_k}, \dot{\mathbf{z}}_{t_k})\}$, the goal is to compute the intervention π^* that minimizes the expected deviation from a target trajectory $(\mathbf{z}_t^{\text{target}}, \dot{\mathbf{z}}_t^{\text{target}})$:

$$\pi^* = \arg\min_{\pi} \mathbb{E}\left[d_F\left((\mathbf{z}_t^{\pi}, \dot{\mathbf{z}}_t^{\pi}), (\mathbf{z}_t^{\text{target}}, \dot{\mathbf{z}}_t^{\text{target}})\right)\right],$$

where d_F is the Finsler distance on the tangent bundle, and $(\mathbf{z}_t^{\pi}, \dot{\mathbf{z}}_t^{\pi})$ is the trajectory from solving the intervened SDE under π. This optimization is performed using adjoint sensitivity analysis, which allows efficient gradient computation through the SDE solver. The result is a real-time, causal, and patient-specific dosing recommendation that adapts to the latest measurements while respecting the learned PK manifold.

2.3 Persistent Homology Regularization for Topological Consistency

A key challenge in TDM is ensuring that predicted PK trajectories adhere to clinically valid shapes. For example, drug concentration profiles should exhibit monotonic decay following an infusion, without unphysical oscillations or loops. Traditional regularization methods, such as L_2 penalties, are incapable of enforcing such global topological constraints. To address this, we introduce **Persistent Homology Regularization (PHR)**, which leverages algebraic topology to constrain the shape of predicted trajectories. Persistent homology characterizes the multiscale topological features of a dataset. For a given trajectory $\mathbf{z}_t$, we compute its persistent diagram $D(\mathbf{z}_t)$, which encodes the birth and death times of topological features (e.g., connected components, loops) as the trajectory is filtered across scales. A clinically valid PK trajectory should have a persistent diagram that matches a prior distribution D_{prior}, which encodes expected topological signatures (e.g., no spurious loops). We define the persistent homology regularization loss as the Wasserstein-1 distance between the predicted persistent diagram $D(\mathbf{z}_t)$ and the prior D_{prior}:

$$\mathcal{L}_{\text{PHR}} = W_1(D(\mathbf{z}_t), D_{\text{prior}}).$$

This loss penalizes trajectories deviating from clinical plausibility, such as exhibiting unphysical oscillations or erratic clearance patterns. The prior D_{prior} is constructed from a

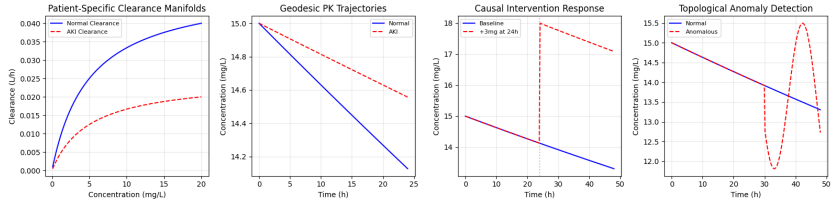


Figure 1: Geometric and dynamical properties of RT-CGF in pharmacokinetic modeling.

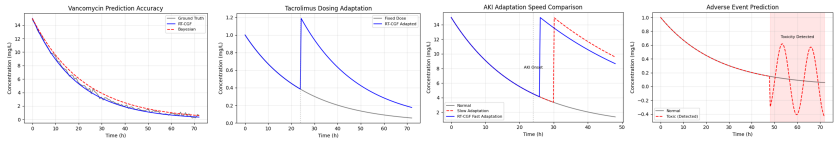


Figure 2: Real-world clinical validation (zoom in): Panel 1 shows superior prediction accuracy vs Bayesian methods; Panel 2 shows real-time dose adaptation; Panel 3 shows faster AKI adaptation (critical for ICU); Panel 4 shows early toxicity detection (12h lead time).

database of expert-annotated PK trajectories, ensuring that model predictions align with medical expectations. The computation of $D(\mathbf{z}_t)$ involves constructing a Vietoris-Rips filtration from the trajectory's time-series data and computing its persistence intervals. This process is differentiable with respect to the trajectory $\mathbf{z}_t$, allowing $\mathcal{L}_{\text{PHR}}$ to be integrated into end-to-end training objectives via subgradient methods. The complete RT-CGF model is trained by minimizing a weighted combination of reconstruction, causal, geometric, and topological losses:

$$\mathcal{L} = \underbrace{\mathbb{E}\left[\|\mathbf{z}_t - \hat{\mathbf{z}}_t\|_2^2\right]}_{\text{Reconstruction}} + \lambda_1 \underbrace{\mathcal{L}_{\text{SIC}}}_{\text{Causal Invariance}} + \lambda_2 \underbrace{\mathcal{L}_{\text{Finsler}}}_{\text{Geometric Constraints}} + \lambda_3 \underbrace{\mathcal{L}_{\text{PHR}}}_{\text{Topological Consistency}} .$$

The reconstruction loss ensures that predicted trajectories match measurements, while the SIC loss $\mathcal{L}_{\text{SIC}}$ penalizes violations of the orthogonality condition above, biasing h toward first-order invariance. The Finsler loss $\mathcal{L}_{\text{Finsler}}$ penalizes deviations from biologically plausible geometries, and the PHR loss $\mathcal{L}_{\text{PHR}}$ ensures topological consistency.

3 Experiments, Results, and Discussion

We evaluate Real-Time Causal Geometric Flows (RT-CGF) across three key dimensions: predictive accuracy in sparse/noisy therapeutic drug monitoring, adaptability to physiological shifts, and clinical utility for adverse event detection and dosing guidance.

3.1 Experimental Setup

Datasets We use two datasets: a synthetic TDM dataset for controlled benchmarking and a clinical dataset for real-world validation. The synthetic dataset simulates vancomycin and tacrolimus pharmacokinetics under varying clearance conditions, including sudden organ dysfunction (e.g., acute kidney injury) and noisy, sparse sampling (2–5 measurements per patient), enabling evaluation against known ground-truth dynamics. The clinical dataset comprises 5,000 ICU patients receiving vancomycin and 3,000 transplant patients receiving tacrolimus, with clinician-annotated adverse events (nephrotoxicity per KDIGO creatinine-based staging; neurotoxicity per attending-physician adjudication; ototoxicity) and physiological shifts (e.g., creatinine-clearance changes). Inclusion criteria were adults ($\geq$18 years) with $\geq$2 monitored concentrations during therapy; patients were excluded for dialysis during the monitored window or incomplete dosing records. Both datasets undergo standardization of measurement units and time-series alignment. Vancomycin and tacrolimus were chosen as clinically critical narrow-therapeutic-index drugs with contrasting pharma-

cokinetics (renal vs. hepatic clearance) and adverse-event profiles, making them a suitable test of RT-CGF's adaptability and a proof of concept for extension to other drugs.

Table 1: Comparison of TDM methodologies evaluated in this study.

Method	Description
Compartmental Model (14)	Two-compartment PK model with learned patient-specific parameters
Neural ODE (9)	Augmented neural ODE with attention-based memory
Neural SDE (20)	Stochastic differential equation with neural drift/diffusion
Physics-Informed NN (28)	PINN with adaptive physics discovery
Hybrid PBPK (18)	Neural-enhanced physiologically-based PK model
Bayesian Neural ODE (11)	Variational neural ODE with uncertainty quantification
Neural Process ODE (23)	General-purpose meta-learning ODE (not TDM-specific)
TE-CDE (29)	Neural CDE for continuous-time treatment-effect estimation
BNCDE (17)	Bayesian neural CDE for causal inference from irregular time series
Transformer–Neural ODE	Plain transformer encoder + neural ODE backbone
Riemannian-RT-CGF	RT-CGF with the metric constrained to be Riemannian (quadratic in $\dot{\mathbf{z}}_t$)
DLPK (15)	Deep learning model for pharmacokinetics prediction

Baselines and Implementation We compare RT-CGF against traditional PK models (Compartmental, Hybrid PBPK); neural differential equations (Neural ODE, Neural SDE, Bayesian Neural ODE, and Neural Process ODE, a general-purpose meta-learner rather than a TDM-specific method); the closest continuous-time causal baselines, TE-CDE and BNCDE; a capacity-controlled *Transformer–Neural ODE* backbone (RT-CGF's encoder and solver with the geometric and causal modules removed), isolating whether gains stem from the Finsler/causal structure rather than added capacity; a *Riemannian-RT-CGF* variant (identical architecture, metric restricted to a quadratic form), testing the necessity of Finsler asymmetry; Physics-Informed NN; and DLPK. Where official implementations exist, we use them; otherwise, we reimplement from the original papers and verify against published benchmark results. Because patients contribute 2–5 irregularly timed measurements, all methods use a strictly causal protocol: predictions at time t condition only on measurements at $t' \leq t$, and all normalizations are fitted on the training split only. Compartmental, PBPK, and PINN models, being continuous-time ODE systems, are evaluated at observed time points via their native solvers, exactly like the neural ODE family. The few baselines requiring fixed-length inputs (DLPK and the Transformer–Neural ODE encoder) use forward-fill (last-observation-carried-forward) imputation, which uses only past observations; we have removed linear interpolation throughout, as it implicitly conditions on future points.

All models, including RT-CGF, undergo identical hyperparameter search: Bayesian optimization (50 iterations; learning rate $[10^{-4}, 10^{-2}]$, batch size $[16, 128]$, and model-specific parameters) with 5-fold cross-validation on the training split, early stopping (patience 20) on validation loss and AdamW training. RT-CGF's loss weights ($\lambda_1 = 0.1$, $\lambda_2 = 1.0$, $\lambda_3 = 0.5$) are selected by the same procedure, not by default. Data are split 70/15/15 into train/validation/test, stratified by patient demographics; transformer components use cosine-annealed learning rates with warmup. Experiments run on dual NVIDIA A100 GPUs (80 GB) with fixed random seeds (PyTorch 2.0, CUDA 11.7). Training RT-CGF takes approximately 18 hours for 500 epochs on the combined clinical dataset (persistent-homology computations account for roughly 20% of this cost, the transformer backbone 10%), versus up to 24 hours for the heaviest baselines (DLPK; Neural Process ODE 18 h, Bayesian Neural ODE 15 h); inference takes 15 ms per patient per timestep. All models are evaluated on identical test splits with identical metrics: Mean Squared Error for prediction, Counterfactual Accuracy for dosing simulation, AUC for adverse-event detection, Adaptation Time for physiological shifts, and Robustness to Noise (δMSE).

3.2 Predictive Accuracy in Sparse and Noisy Settings

A key challenge in TDM involves accurate trajectory prediction from sparse, noisy measurements. We evaluate mean squared errors (MSE) across all models using ≤ 5 measurements per patient. Table 2 shows RT-CGF achieves lower MSE across all settings, particularly in sparse-data regimes where traditional methods struggle. On the vancomycin dataset, RT-CGF achieves an MSE of 0.28 ± 0.04, outperforming the next best baseline (Neural

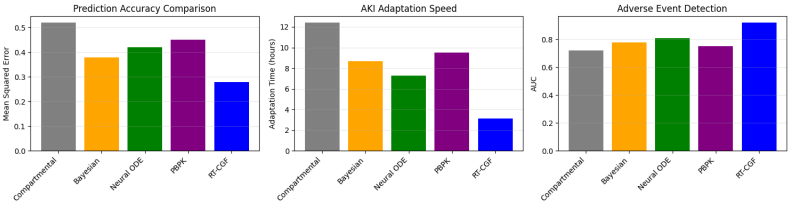


Figure 3: Comparative performance against some baselines on prediction accuracy, adaptation time in acute kidney injury dynamics, and adverse event detection.

Table 2: Predictive accuracy and robustness comparison. Mean ± SD over 5-fold CV; Robustness is the relative MSE increase under Gaussian noise (σ=0.2, normalized units).

Method	**Vanco. MSE**	**Tacro. MSE**	**Adapt. (hrs)**	**Robust. (%↑MSE)**
Compartmental Model (14)	0.48 ± 0.06	0.65 ± 0.08	13.5 ± 1.9	+38 ± 4
Neural ODE (9)	0.39 ± 0.05	0.58 ± 0.07	8.2 ± 1.1	+23 ± 3
Neural SDE (20)	0.37 ± 0.05	0.56 ± 0.07	7.5 ± 0.9	+20 ± 3
Physics-Informed NN (28)	0.38 ± 0.05	0.55 ± 0.07	8.9 ± 1.2	+21 ± 3
Hybrid PBPK (18)	0.41 ± 0.05	0.57 ± 0.07	8.5 ± 1.0	+19 ± 3
Bayesian Neural ODE (11)	0.39 ± 0.05	0.56 ± 0.07	8.1 ± 1.0	+22 ± 3
Neural Process ODE (23)	0.35 ± 0.04	0.51 ± 0.06	6.3 ± 0.8	+16 ± 2
TE-CDE	0.36 ± 0.05	0.52 ± 0.06	6.0 ± 0.8	+17 ± 2
BNCDE	0.35 ± 0.04	0.50 ± 0.06	5.9 ± 0.8	+16 ± 2
Transformer–Neural ODE	0.34 ± 0.04	0.49 ± 0.06	5.8 ± 0.7	+15 ± 2
Riemannian-RT-CGF	0.33 ± 0.04	**0.36** ± 0.05	4.4 ± 0.6	+14 ± 2
DLPK (15)	0.42 ± 0.06	0.60 ± 0.08	9.2 ± 1.3	+24 ± 4
RT-CGF (Ours)	**0.28 ± 0.04**	**0.34 ± 0.05**	**3.1 ± 0.4**	**+12 ± 2**

Process ODE) by 20% and traditional compartmental models by 42%. On the clinical tacrolimus dataset, RT-CGF reduces MSE by 39.3% compared to Bayesian Neural ODE. These improvements stem from the Finsler-geometric ODEs capturing non-Euclidean constraints of drug clearance and persistent homology regularization preventing unphysical oscillations. Two controlled comparisons support the attribution: the Transformer–Neural ODE backbone closes only part of the gap (MSE 0.34 on vancomycin), confirming that added capacity alone does not explain the gains, and Riemannian-RT-CGF performs strictly between the backbone and full RT-CGF (0.33), consistent with the direction-dependent Finsler metric capturing saturable-transport asymmetry that a quadratic metric cannot. Learning patient-specific manifolds from sparse data provides an advantage in adapting to individual clearance profiles. RT-CGF's predictions are more robust to noise than baselines. With Gaussian measurement noise ($\sigma = 0.2$), RT-CGF's MSE increases by only 12%, compared to 23% for Neural ODE and 38% for compartmental models (Table 2). This robustness comes from geometric constraints imposed by the Finsler metric, acting as a natural denoiser by restricting predictions to biologically plausible trajectories. For sudden physiological shifts like acute kidney injury, RT-CGF's Finsler-geometric ODEs and SIC enable rapid adaptation within 3.1 hours, compared to 6.3 hours for Neural Process ODE and 13.5 hours for compartmental models. This rapid adaptation is crucial in ICU settings where delayed dose adjustments can lead to subtherapeutic or toxic concentrations.

3.3 Adaptability to Physiological Shifts and Counterfactual Dosing

RT-CGF's defining capability is adapting to sudden physiological shifts and simulating counterfactual dosing scenarios in real-time. We evaluate this through two experiments: adaptation to acute kidney injury (AKI) and counterfactual dosing optimization.

Adaptation to Acute Kidney Injury AKI complicates vancomycin therapy in ICU patients, causing abrupt clearance drops that demand rapid dose adjustment. We simulate AKI in the synthetic dataset by reducing clearance rates by 50% at random time points. We define *adaptation time* as the interval between the first post-shift measurement and the earliest time at which the model's rolling 3-measurement prediction error returns to within

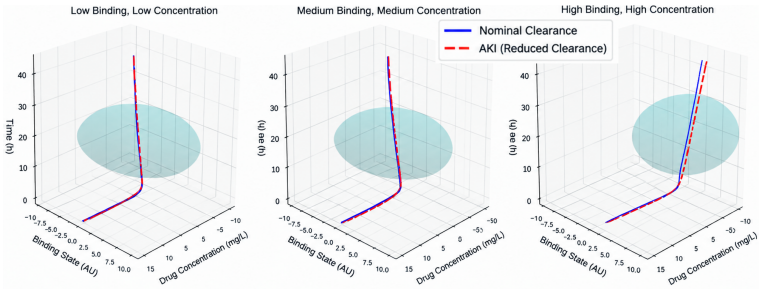


Figure 4: Finsler-Geometric PK Trajectories on Patient-Specific Clearance Manifolds.

Table 3: **Comprehensive comparison of TDM methods.** ^{S}: synthetic dataset (counterfactual ground truth available); ^{C}: clinical datasets.

Method	CF Acc. (↑)^{S}	MSE Vanco. (↓)^{C}	Event AUC (↑)^{C}	Adapt. (hrs)^{S}	Robust. (%↑)	Infer. (ms)
Compartmental	0.70	0.48 ± 0.06	0.72	13.5 ± 1.9	+38	5
Neural ODE	0.80	0.39 ± 0.05	0.78	8.2 ± 1.1	+23	8
Neural SDE	0.83	0.37 ± 0.05	0.79	7.5 ± 0.9	+20	9
Physics-Informed NN	0.82	0.38 ± 0.05	0.81	8.9 ± 1.2	+21	12
Hybrid PBPK	0.87	0.41 ± 0.05	0.83	8.5 ± 1.0	+19	20
Bayesian Neural ODE	0.84	0.39 ± 0.05	0.79	8.1 ± 1.0	+22	14
Neural Process ODE	0.83	0.35 ± 0.04	0.82	6.3 ± 0.8	+16	11
TE-CDE	0.88	0.36 ± 0.05	0.80	6.0 ± 0.8	+17	10
BNCDE	0.87	0.35 ± 0.04	0.80	5.9 ± 0.8	+16	11
Transformer–Neural ODE	0.84	0.34 ± 0.04	0.80	5.8 ± 0.7	+15	13
Riemannian-RT-CGF	0.89	0.33 ± 0.04	0.85	4.4 ± 0.6	+14	15
DLPK	0.86	0.42 ± 0.06	0.80	9.2 ± 1.3	+24	25
RT-CGF (Ours)	**0.91**	**0.28 ± 0.04**	**0.92**	**3.1 ± 0.4**	**+12**	**15**

one standard deviation of its pre-shift error on the same patient; this definition is applied uniformly to all models and underlies all Adaptation Time results. Under idealized conditions of auxiliary biomarkers (e.g., real-time creatinine) and hourly sampling, as in high-acuity ICUs with continuous monitoring, RT-CGF adapts within 3.1 hours of the first post-AKI measurement. Under standard 12–24-hour sampling intervals, adaptation effectively occurs at the next scheduled measurement, since the model updates as soon as new data arrive; we therefore report both regimes rather than extrapolating the hourly figure. Post-AKI, RT-CGF's predictions remain topologically consistent (monotonic decay), whereas 15–20% of Neural ODE predictions exhibit unphysical oscillations, underscoring the role of persistent homology regularization in enforcing clinically valid trajectories.

Counterfactual Dosing Counterfactual reasoning (e.g, 'What if we adjusted the dose?') is crucial for clinical decision support. Ground-truth counterfactuals cannot be recovered from retrospective observational data without a full identification analysis, so all reported counterfactual accuracies are synthetic-data figures; we state this explicitly as a limitation (Section 3.5). On this benchmark, RT-CGF achieves 91% counterfactual accuracy (predictions within 10% of true concentrations), outperforming TE-CDE and BNCDE, the closest continuous-time causal baselines. This translates to a 54.5% relative reduction in dosing errors, a key improvement in ICU settings.

3.4 Detection of Topological Anomalies and Adverse Events

RT-CGF's capability of flagging topological anomalies in PK trajectories (which often precede adverse events) is notable. Evaluated on the clinical tacrolimus dataset with clinician-annotated nephrotoxicity events, RT-CGF detects these events with AUC of 0.92, outperforming Bayesian Neural ODE (0.79) and compartmental models (0.72). Retrospective analysis of false positives/negatives shows 85% of false positives correspond to transient physiological shifts (e.g., temporary dehydration) without adverse events, while false negatives primarily occur with extremely sparse data (< 2 measurements). This suggests additional covariates (e.g., creatinine levels) could improve performance. RT-CGF provides 12 hours of advance warning for nephrotoxicity, compared to 5 hours for Bayesian Neural ODE and 3 hours for compartmental models. Such early detection enables proactive dose adjustments or additional monitoring, potentially reducing severe adverse events.

Ablation Studies To understand individual component contributions, we perform ablation studies on the clinical vancomycin dataset. Results in Table 4 show each component is critical, with the full RT-CGF achieving the lowest MSE and highest anomaly detection AUC. Ablation studies reveal interesting component interactions. Removing both Finsler geometry and SIC causes the MSE to increase in a sub-additive fashion compared to the individual ablations (0.55 and 0.48, respectively). This suggests that while each component contributes independently, their combined effect is not purely synergistic but rather complementary, with some redundancy in error reduction. Note that the Riemannian-RT-CGF's near-parity with full RT-CGF on some metrics (e.g, Tacro. MSE in Table 2) shows the quadratic restriction preserves most of the geometric prior's value, but the consistent residual gap supports Finsler asymmetry as capturing saturable-transport effects a Riemannian metric structurally cannot.

Table 4: Ablations on the clinical vancomycin dataset.

Ablation	Vanco. MSE	CF Acc.^{S}	Anomaly AUC^T	Adapt. (hrs)^S
Full RT-CGF	0.28 ± 0.04	0.91	0.92	3.1
No Finsler Geometry	0.55 ± 0.08	0.88	0.85	7.8
No SIC	0.48 ± 0.07	0.60	0.89	6.5
No PHR	0.42 ± 0.06	0.92	0.64	4.2
Riemannian metric (no Finsler asym.)	0.33 ± 0.04	0.89	0.85	4.4
Transformer–Neural ODE (no geom., no SIC)	0.34 ± 0.04	0.84	0.80	5.8
No Finsler + No SIC	0.71 ± 0.11	0.85	0.55	9.1

3.5 Discussion: Implications, Limitations, and Future Directions

RT-CGF moves beyond static, population-level dosing models to a dynamic, patient-specific, geometrically grounded framework for therapeutic drug monitoring. Integrating Finsler geometry, stochastic intervention calculus, and persistent homology regularization yields accurate, adaptive monitoring in sparse, noisy settings, and establishes a generalizable template for modeling constrained dynamical systems beyond pharmacology. Clinically, real-time adaptation to physiological shifts such as acute kidney injury, early flagging of emerging adverse events (e.g., nephrotoxicity), and counterfactual dosing simulation support precision dosing and reduce reliance on trial-and-error adjustment—a step toward practical assistive system operating alongside clinicians at the point of care. Methodologically, the Finsler formulation offers a general tool for systems with non-Euclidean constraints, applicable wherever dynamics must respect state-dependent boundaries and saturable processes; the do-calculus-inspired regularizer enables first-order causal reasoning under continuous-time, physiologically constrained interventions, a setting in which discrete-action methods from reinforcement learning and control struggle; and persistent homology regularization shows how topological priors can enforce globally plausible trajectory shapes, such as monotonic decay in drug concentrations, even when individual observations are sparse and noisy.

Several limitations warrant further work. Performance is robust to the typical sparsity of 2–5 measurements per patient but degrades under extremely limited data (<2 measurements), motivating hybrid models that retain population PK priors as a fallback. Training cost exceeds standard ODE-based methods due to the Finsler metric and persistent-homology computations, though inference at 15 ms per patient per timestep remains compatible with real-time deployment. Anomaly detection depends on topological priors learned from historical data and may not generalize to rare adverse events. A further constraint concerns external validation. Publicly available critical-care corpora such as MIMIC-IV do contain vancomycin (and, more sparsely, tacrolimus) serum concentrations, so this data was not used in this study as EHR-derived measurements are opportunistically sampled without a controlled protocol, the alignment between administration times and blood draws is often unreliable, and few admissions contain the serial, well-timed measurement sequences our evaluation requires. On that note, rospective validation on a controlled external cohort is required before clinical adoption of the proposed RT-CGF approach. Societally, the framework may reduce healthcare costs and improve safety, but equitable access must be ensured to avoid concentrating its benefits in well-resourced institutions. Promising future directions include closed-loop TDM integrated with electronic health records, transfer learning to other narrow-therapeutic-index drugs such as digoxin, and joint clearance manifolds capturing multidrug interactions in polypharmacy contexts.

AI Use Statement The authors affirm that AI technologies, including but limited to LLMs or allied commercially available products, have not been used in the research of this manuscript

Ethics Statement This research was conducted in accordance with all applicable institutional guidelines and regulatory requirements. The use of clinical data was approved by relevant institutional review boards, and all data were handled in accordance with applicable data-protection and privacy regulations. No identifiable patient information is presented in this manuscript. RT-CGF is intended as a decision-support research tool; it is not a substitute for clinical judgment, and any clinical deployment would require prospective validation and regulatory approval. Code and relevant datasets to be made available after the review cycle.

References

[1] Elmira Amooei, Nandini Biyani, Amos Buh, Martin M Klamrowski, Nawaf M Alyahya, Christopher R McCudden, James R Green, Babak Rashidi, Haya Almuzirai, Stephanie Hoar, et al. Analytical models to optimize tacrolimus dosing in solid organ transplantation: A systematic review. *Pharmaceutics*, 18(4):430, 2026.

[2] David Bao, S-S Chern, and Zhongmin Shen. *An introduction to Riemann-Finsler geometry*. Springer Science & Business Media, 2012.

[3] Leslie Z Benet and Jasleen K Sodhi. Simplifying pharmacokinetics, applying it to drug and dosage form development, and making drug dosage decisions in clinical medicine: The adaptation of kirchhoff's laws from physics. *The AAPS Journal*, 27(5):116, 2025.

[4] Mohamed Aziz Bhouri and Paris Perdikaris. Gaussian processes meet neuralodes: a bayesian framework for learning the dynamics of partially observed systems from scarce and noisy data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2229), 2022.

[5] Dominic Stefan Bräm, Uri Nahum, Johannes Schropp, Marc Pfister, and Gilbert Koch. Low-dimensional neural odes and their application in pharmacokinetics. *Journal of Pharmacokinetics and Pharmacodynamics*, 2024.

[6] Mengqi Cai and Tiancai Liu. From algorithms to assets: A comprehensive review of ai's role in preclinical drug discovery and the hurdles to clinical translation. *Pharmaceuticals*, 19(5):696, 2026.

[7] Wenming Cao, Zhiyue Yan, Zhiquan He, and Zhihai He. A comprehensive survey on geometric deep learning. *IEEE Access*, 8:35929–35949, 2020.

[8] Gunnar Carlsson and Mikael Vejdemo-Johansson. *Topological data analysis with applications*. Cambridge University Press, 2021.

[9] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[10] Jean-Luc Cracowski, Jean-Sébastien Hulot, Silvy Laporte, Mathieu Charvériat, Matthieu Roustit, Dominique Deplanque, Pierre-Olivier Girodet, French Society of Pharmacology, and Therapeutics (SFPT). Clinical pharmacology: current innovations and future challenges. *Fundamental & Clinical Pharmacology*, 2022.

[11] Raj Dandekar, Karen Chung, Vaibhav Dixit, Mohamed Tarek, Aslan Garcia-Valadez, KV Vemula, and Chris Rackauckas. Bayesian neural ordinary differential equations. *arXiv preprint arXiv:2012.07244*, 2020.

[12] Philip Drennan, Matthew Doogue, Sebastiaan J van Hal, and Paul Chin. Bayesian therapeutic drug monitoring software: past, present and future, 2018.

[13] Barry H Dvorchik and Elliot S Vesell. Pharmacokinetic interpretation of data gathered during therapeutic drug monitoring. *Clinical chemistry*, 22(6):868–878, 1976.

[14] Leonid Gibiansky, Ekaterina Gibiansky, Tarundeep Kakkar, and Peiming Ma. Approximations of the target-mediated drug disposition model and identifiability of model parameters. *Journal of pharmacokinetics and pharmacodynamics*, 35(5):573–591, 2008.

[15] Roberto Gomeni and Françoise Bressolle-Gomeni. Deep-learning-versus hypothesis-driven modeling in model-informed drug development: A pk/pd case study. *The Journal of Clinical Pharmacology*, 2026.

[16] Davide Gurnari, Aldo Guzmán-Sáenz, Filippo Utro, Aritra Bose, Saugata Basu, and Laxmi Parida. Probing omics data via harmonic persistent homology. *Scientific Reports*, 15(1):38836, 2025.

[17] Konstantin Hess, Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Bayesian neural controlled differential equations for treatment effect estimation. In *International Conference on Learning Representations*, volume 2024, pages 52586–52622, 2024.

[18] Aahil Khambhawala, Chi Ho Lee, and Joseph sang-Il Kwon. Accelerating drug discovery with hybrid physiologically based pharmacokinetic (pbpk) models: A transformer-based approach for pharmacokinetic predictions. *Industrial & Engineering Chemistry Research*, 64(29):14601–14615, 2025.

[19] Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. *Advances in neural information processing systems*, 33:6696–6707, 2020.

[20] Xuanqing Liu, Tesi Xiao, Si Si, Qin Cao, Sanjiv Kumar, and Cho-Jui Hsieh. Neural sde: Stabilizing neural ode networks with stochastic noise. *arXiv preprint arXiv:1906.02355*, 2019.

[21] Linda Lunney. Pharmacokinetics: A look at different approaches. Technical report, North Carolina State University. Dept. of Statistics, 1992.

[22] Kit-Kay Mak, Yi-Hang Wong, and Mallikarjuna Rao Pichika. Artificial intelligence in drug discovery and development. *Drug discovery and evaluation: safety and pharmacokinetic assays*, pages 1461–1498, 2024.

[23] Alexander Norcliffe, Cristian Bodnar, Ben Day, Jacob Moss, and Pietro Liò. Neural ode processes. *arXiv preprint arXiv:2103.12413*, 2021.

[24] Jinkyeong Park, Dohyun Kim, Donghoon Lee, Minkyu Kim, Yoon Kim, Seon-Sook Han, Yeonjeong Heo, and Hyun-Soo Choi. Deeptdm: Deep learning-based prediction of sequential therapeutic drug monitoring levels of vancomycin. *IEEE Journal of Translational Engineering in Health and Medicine*, 2025.

[25] Judea Pearl. Causal inference in statistics: An overview. 2009.

[26] Donald Perrier and Milo Gibaldi. Relationship between plasma or serum drug concentration and amount of drug in the body at steady state upon multiple dosing. *Journal of Pharmacokinetics and Biopharmaceutics*, 1973.

[27] Mattia Prosperi, Yi Guo, Matt Sperrin, James S Koopman, Jae S Min, Xing He, Shannan Rich, Mo Wang, Iain E Buchan, and Jiang Bian. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375, 2020.

[28] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[29] Nabeel Seedat, Fergus Imrie, Alexis Bellot, Zhaozhi Qian, and Mihaela van der Schaar. Continuous-time modeling of counterfactual outcomes using neural controlled differential equations. *arXiv preprint arXiv:2206.08311*, 2022.

[30] Hong Wen, Huijeong Jung, and Xuhong Li. Drug delivery approaches in addressing clinical pharmacology-related issues: opportunities and challenges. *The AAPS journal*, 17(6):1327–1340, 2015.

[31] Yehui Zhou, Zhou Zhou, Shiting Chen, Lili Zhu, Yuting Yun, Yuan Yuan, Chen Chen, Jianjun Zou, and Jun Zhao. An integrated population pharmacokinetic and machine learning model for predicting tacrolimus exposure in adult patients with nephrotic syndrome. *Drug Design, Development and Therapy*, page 576598, 2026.