Latent Neural-ODEs enable counterfactual precision dosing from sparse data

Benjamin Maurel[1*], Agathe Guilloux[1], Sarah Zohar[1], Moreno Ursino[1], Jean-Baptiste Woillard[2,3]

[1] HeKA, Université Paris Cité, Inria, Inserm, Paris, France [2] P&T, U1248, Université de Limoges, CHU Limoges, France [3] Service de Pharmacologie, CHU Limoges, France

① PROBLEM & CONTEXT › ② MODEL & METHOD › ③ EVIDENCE & RESULTS › ④ APPLICATION

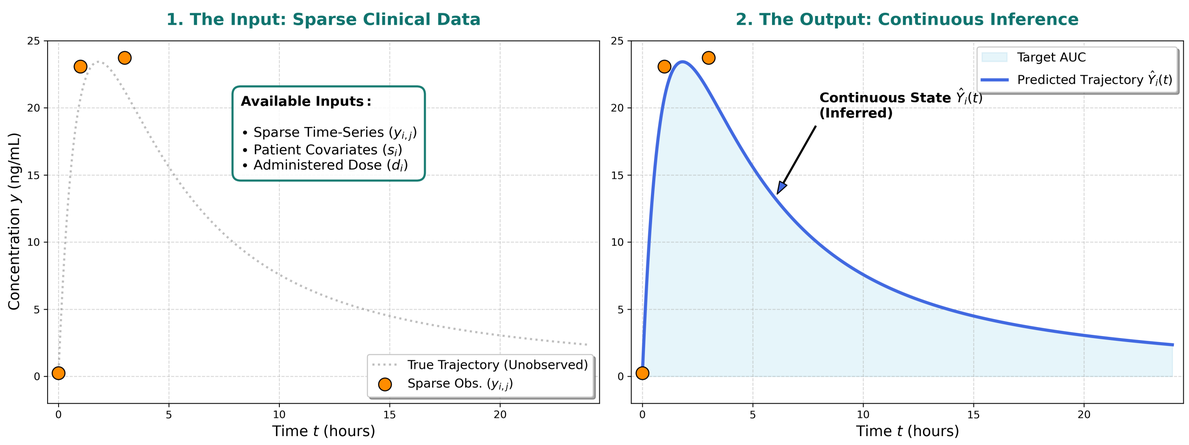
The Challenge of Precision Dosing

Model-informed precision dosing (MIPD) aims to optimize drug therapy by tailoring treatments to individual patients.

This is critical for narrow-therapeutic-index drugs like **tacrolimus** in renal transplantation, where precise continuous exposure control prevents life-threatening rejection or toxicity.

The Data Bottleneck: Determining a patient's Area Under the Curve (AUC) ideally requires 8 to 12 blood draws a day. Clinicians are instead limited to sparse, irregular samples (typically 2 to 3 noisy concentration measurements per patient, e.g., C_0, C_2, C_4).

The Objective: Inferring Continuous AUC



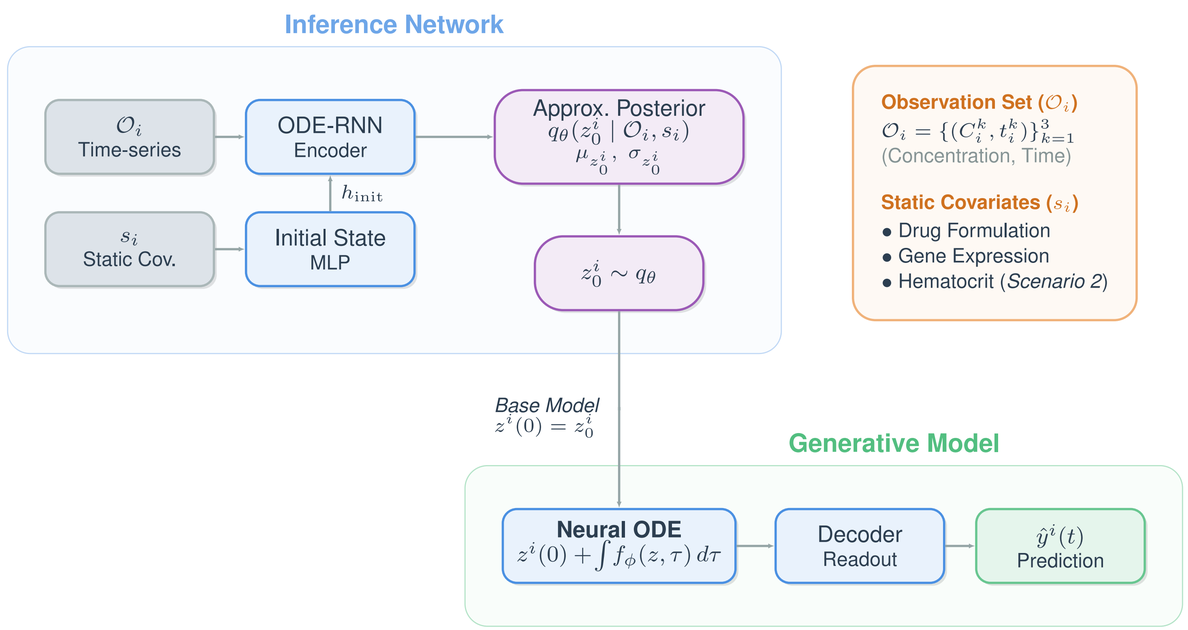
Standard Practice: MAP-BE (NLME)

Traditionally, exposure is estimated using population-derived **NLME models** under rigid assumptions:

- **Rigid Structural ODEs:** Usually pre-defined (e.g., 2-compartment with linear elimination).
- **Fixed Covariate Models:** Assumes specific biological relationships (e.g., Hematocrit on clearance).

The Limitation: If atypical patient biology deviates from these rigid compartmental assumptions, MAP-BE forces data to fit incorrect equations, causing systematic predictive bias.

Replacing Mechanistic Models with Latent ODEs



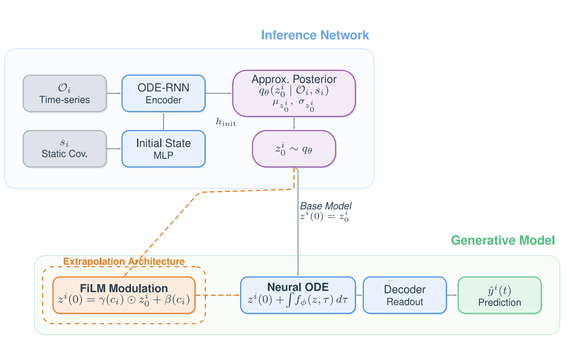
1. Probabilistic Inference Network: Maps sparse observations ($\mathcal{O}_i$, e.g., 3 blood samples) and static covariates (s_i, e.g., age, weight) to a continuous latent distribution $q_\theta(z_0^i|\mathcal{O}_i, s_i)$.

2. Generative Neural-ODE: Governs the continuous dynamical evolution of the patient's state strictly within the latent space, avoiding rigid compartmental constraints before decoding.

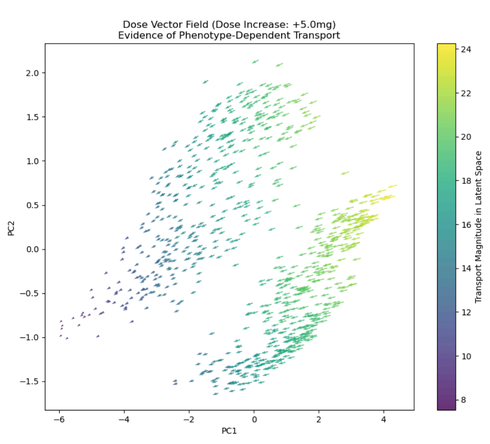
Dosing as Latent Affine Modulation (FiLM)

To avoid artificial arithmetic state-jumps (such as $z_{t^+} = z_{t^-} + \text{Dose}$), we define counterfactual dosing as a transport problem:

$$z_{0,\text{new}} = \gamma(c_i) \odot z_{\text{base}} + \beta(c_i)$$



1. Latent Conditioning: The FiLM layer modulates the base biological state (z_{base}) using scaling $\gamma(c_i)$ and shifting $\beta(c_i)$ networks.



2. Phenotype Preservation: Dose adjustments are expressed as smooth, topology-preserving latent displacements along the curved manifold.

Separation of Dynamics: This formulation cleanly isolates intrinsic patient clearance from administered dose variations, ensuring mathematically robust counterfactual predictions.

Robustness Against Misspecified NLME Models

We compared our generative model against standard clinical estimators (MAP-BE using Monolix-adjacent SAEM priors) across three synthetic data scenarios using a 2-compartment PK model baseline.

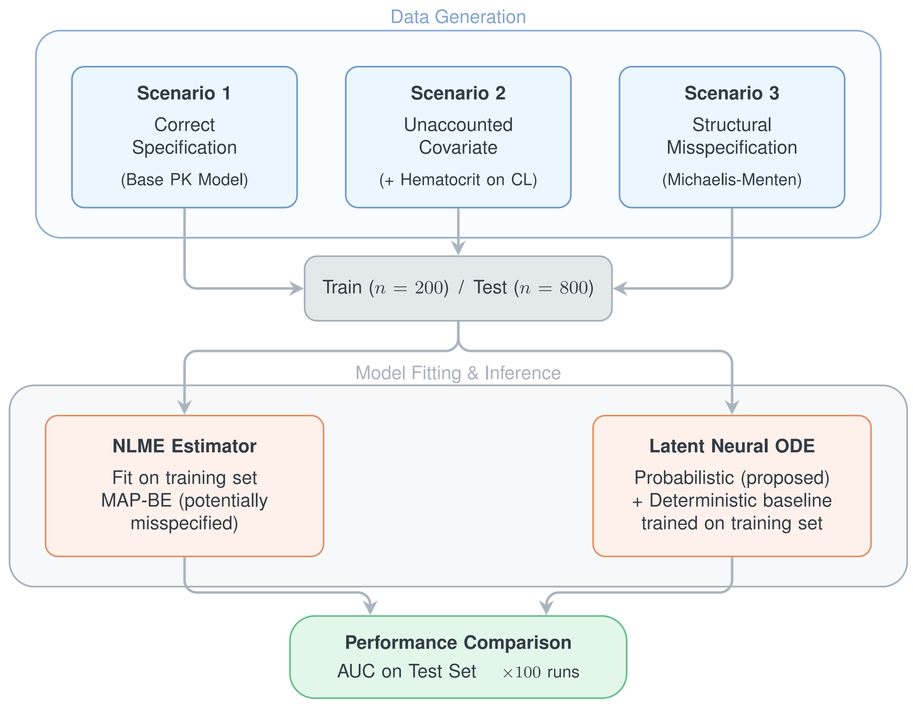


Table 1: Performance over 100 simulation runs

Scenario	Metric	MAP-BE (Clinical Reference)	Latent ODE (Proposed)
1: Correct Spec.	MPE (%)	0.0106 ± 0.0166	0.0123 ± 0.0210
	RMSPE (%)	16.63 ± 1.23	16.72 ± 1.14
2: Unaccounted Cov.	MPE (%)	0.0235 ± 0.0147	**0.0036 ± 0.0252**
	RMSPE (%)	22.96 ± 0.73	**20.42 ± 1.19**
3: Structural Misspec.	MPE (%)	0.0102 ± 0.0749	0.0039 ± 0.0246
	RMSPE (%)	23.57 ± 8.55	**15.40 ± 0.04**

Physiological Validation of the Latent Space

Why a Latent Space? Standard Neural-ODEs operate directly on observable states. Our model instead projects patients into a regularized probabilistic representation (z_0). But is this just an uninterpretable black box?

Recovering Hidden PK Parameters

In Simulation Scenario 2, critical biological covariates—such as Hematocrit (HT) and the elimination rate constant (k_{elim})—were **intentionally withheld** from the model. Yet, applying PCA to the inferred initial latent states (z_0) reveals a highly structured topological organization:

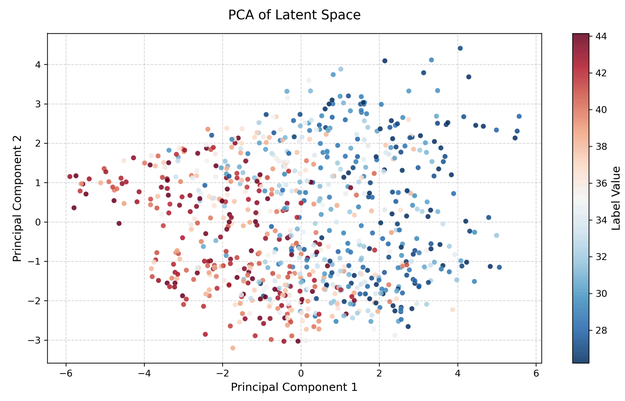


Fig A: Colored by unobserved HT. Shows an autonomous gradient.

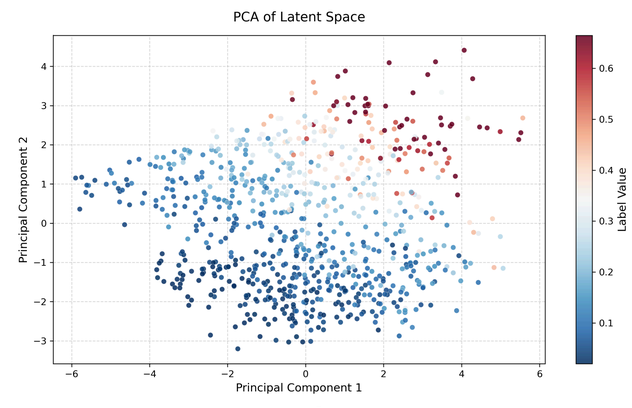


Fig B: Colored by unobserved (k_{elim}). Maps slow to fast metabolizers.

Conclusion: The model infers missing biological heterogeneity strictly from the dynamic shape of sparse time-series inputs. The latent space is a biologically grounded embedding of human physiology, not a black box.

Real-World External Validation (Tacrolimus)

We evaluated the model on a **real-world cohort of renal transplant recipients** receiving Tacrolimus.

The model was trained on an internal cohort ($N = 178$) and evaluated on an entirely unseen **external dataset ($N = 75$)**, comparing predictions against current standard-of-care clinical software (ISBA/it2B and SAEM).

Prediction Method	RMSPE (%) [Precision]	MPE (%) [Bias]
Prob. Latent ODE (Proposed)	**10.45%**	+2.74%
$\text{it2B}_{\text{ISBA}}$ (Clinical Standard)	11.58%	-2.91%
$\text{MAP-BE}_{\text{SAEM}}$	11.48%	-2.28%

Safer at the Extremes

Clinical guidelines define a $\pm 20\%$ relative error as the threshold for safe dosing.

Conclusion & Code

- Continuous-time generative models can perform counterfactual forecasting without heuristic jumps.
- Data-driven structure avoids rigid compartmental assumptions.

Try it yourself with your own Monolix PK model!

Colab Notebook: Scan to run in browser

GitHub: github.com/BenJMaurel/PharmaNODE

Work supported by the DIGPHAT project of PEPR Santé Numérique and funded by a grant from the French government (Grant Reference: ANR-22-PESN-0017).