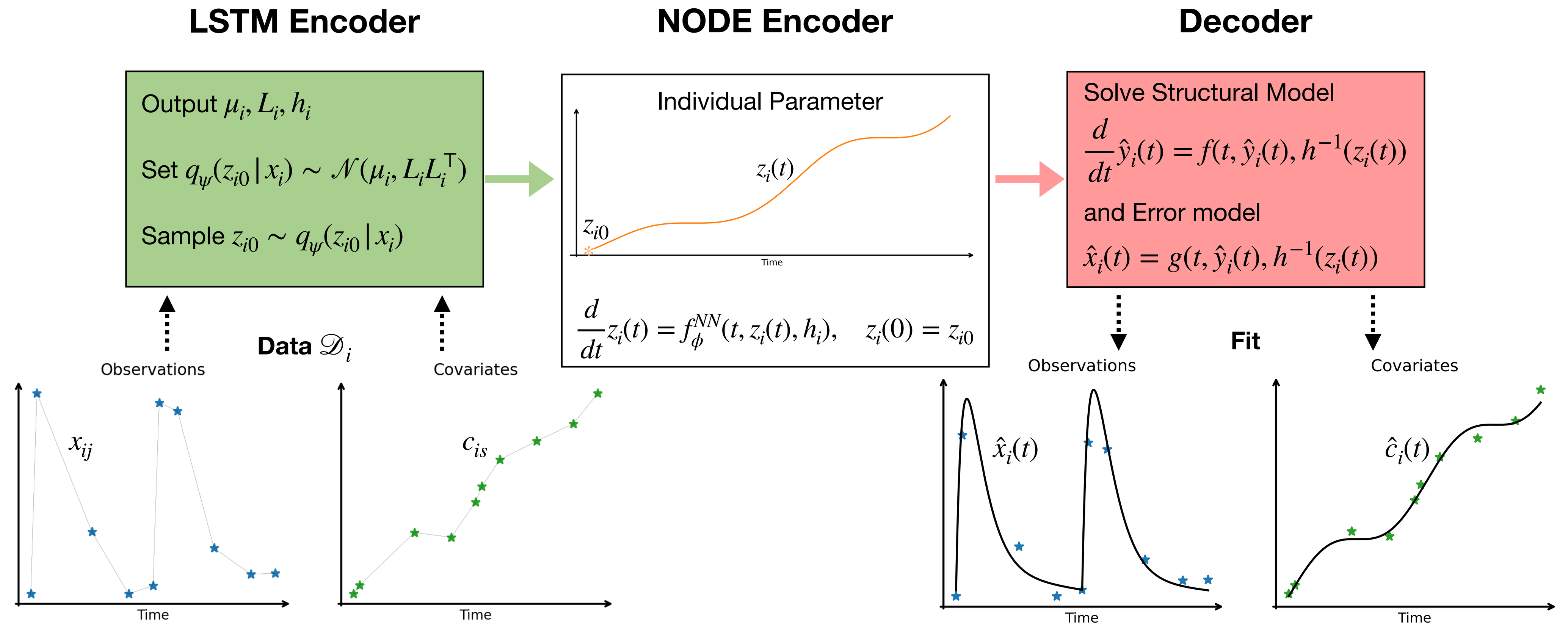
# Using Generative AI Variational Autoencoders with Neural Ordinary Differential Equations to improve NLME Pharmacometric Modeling

Universität Konstanz




UKBB kompetent & menschlich


**J. Schropp[1], F. Bachmann[1], B. Steffens[2], M. Kasper[1], G. Koch[2], J. Rohleff[1]**

(1) Department of Mathematics and Statistics, University of Konstanz, Germany

(2) Pediatric Pharmacology and Pharmacometrics, University of Basel Children's Hospital, Switzerland


## Aim of the project

- Design Variational Autoencoders (**VAEs**) to solve Nonlinear Mixed Effects (**NLME**) Models with time-dependent prior. Combine **VAEs** with Neural Ordinary Differential Equations (**NODEs**) to automatically learn hidden dynamics such as time-varying covariates (**TVCs**)
- Train the VAE via a time-dependent prior adapted Evidence Lower Bound (**ELBO**) to simultaneously capture the NODE dynamics (TVC dynamics) and improve the NLME fit
- Assess the performance with respect to **accuracy** (Log-Likelihood) and **uncertainty** (relative standard error (**RSE**) based on the Fisher Information matrix) and **prediction quality**

## Nonlinear Mixed Effects Models (with time-dependent prior)

Consider a data set of $N$ individuals indexed by $i \in \{1, \ldots, N\}$. For each individual $i$ we consider a TVC $c_i(t)$, $t \geq 0$ that evolves continuously over time. Its individual specific data set is given by

$$\mathcal{D}_i = \{(t_{ij}, x_{ij}); 1 \leq j \leq n_i\} \cup \{(t_{is}^c, c_{is}^{obs}); 0 \leq s \leq n_i^c\}$$

where $t_{ij}$ is the time of the $j$-th observation $x_{ij}$, and $t_{is}^c$ denotes the $s$-th observation of the covariate $c_{is}^{obs}$, modeled as $c_{is}^{obs} \sim \mathcal{N}(c_{is}, \Lambda)$ with $c_{is} = c_i(t_{is}^c)$, and $\Lambda$ is a diagonal covariance matrix.

### Structural Model and Error Model:

For each individual $i$, the structural model reads as

$$\frac{d}{dt} y_i(t) = f(t, y_i(t), \phi_i) \quad \text{for } t \in (0, T], \quad y_i(0) = y_0(\phi_i).$$

The error model satisfies

$$x_{ij} = g(t_{ij}, y_i(t_{ij}), \phi_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, a^2) \quad \text{for } j = 1, \ldots, n_i.$$

### Individual Parameters:

The individual parameters follow the **time-dependent prior distribution**

$$h(\phi_i(t)) = z_i(t), \quad z_i(t) = z_{\text{pop}} + \beta c_i(t) + \eta_i \quad \text{with } \eta_i \sim \mathcal{N}(0, \Omega).$$

Since only discrete measurements of the covariate process are available, we evaluate the prior at

$$z_{is} = z_i(t_{is}^c) = z_{\text{pop}} + \beta c_{is} = z_{\text{pop}} + \beta c_{is}^{obs} + \eta_i + \beta \epsilon_{is}, \quad \epsilon_{is} \sim \mathcal{N}(c_{is}, \Lambda), \quad \eta_i \sim \mathcal{N}(0, \Omega).$$

### Population Fit:

Maximize the marginalized log-likelihood function

$$\mathcal{LL}(x; \theta) = \log p(x) = \int p(x|z) p(z) dz, \text{ with respect to } \theta = (z_{\text{pop}}, \beta, \Omega, \mathbf{\Lambda}, a).$$

## Variational Autoencoder

The VAE-NODE framework extends our previous approach [1]. The encoder is extended to learn TVCs using NODEs. An illustration of the VAE-NODE framework is shown in the figure on top.

### Encoder:

**Part 1 (LSTM)**: For each individual $i$ the encoder processes the individual data set $\mathcal{D}_i$ using a Long-Short-Term Memory (**LSTM**) network with parameters $\Psi$ and returns

$$(\mu_i, L_i, h_i) = \text{EncoderLSTM}_\Psi \mathcal{D}_i \quad \text{for } i = 1, \ldots, N.$$

Here $h_i$ is the final hidden state and the parameters $\mu_i$, $L_i$ are used to sample the posterior

$$q_\Psi(z_{i0}|x_i) = \mathcal{N}(\mu_i, \Sigma_i), \quad \Sigma_i = L_i L_i^T.$$

**Part 2 (NODE)**: Starting from $z_{i0}$, a continuous-time trajectory $z_i(t)$ is generated by the NODE

$$\frac{d}{dt} z_i(t) = f_\Phi^{NN}(t, z_i(t), h_i), \quad z_i(0) = z_{i0},$$

where the right hand side is modeled by the neural network $f_\Phi^{NN}$.

### Decoder:

The decoder applies the time-varying individual parameters $z_i(t)$ within the structural model and generates the predicted trajectories $\hat{y}_i(t)$ and the measurement predictions $\hat{x}_{ij}$.

### Loss Function:

The aim is to maximize the log-likelihood $\mathcal{LL}$ which is done via a TVC adapted ELBO

$$\mathcal{L}_{\Psi,\Phi}^{adapt}(x; \theta) = \sum_{i=1}^{N} \mathbb{E}_{z_{i0} \sim q_\Psi(\cdot|x_i)} \left[ \log p(x_i | z_i(\phi)) - \log q_\Psi(z_{i0}|x_i) + \log p(z_i(\phi)) \right].$$

## References


[1] J. Rohleff, F. Bachmann, U. Nahum, D. Bräm, B. Steffens, M. Pfister, G. Koch, J. Schropp, Redefining Parameter Estimation and Covariate Selection via Variational Autoencoders: One run is all you need. CPT Pharmacometrics Syst. Pharmacol, 2025; 14: 2232-2243.

[2] World Health Organization, https://www.who.int/tools/child-growth-standards/weight-for-age


## Case study

Levothyroxine treatment in newborns and infants with congenital hypothyroidism is investigated. Two versions of a structural model are applied. Model 1 incorporates weight as a TVC, whereas model 2 takes weight as a second output modeled by an expert-based growth model.

### Model equations:

$$\frac{d}{dt} A_B(t) = -k_a A_B(t) + In(t, D), \qquad A_B(0) = 0,$$

$$\frac{d}{dt} A_C(t) = k_a A_B(t) + k_{endo} - k_{el} A_C(t), \quad A_C(0) = k_{endo}/k_{el},$$

where $A_B$ and $A_C$ represent the absorption and central compartments, $In(t, D)$ describes the dosing schedule and $D$ the dose. $FT4$ concentrations are obtained as

$$C^{FT4}(t) = \frac{0.3 A_C(t)}{V^{FT4}(t)},$$

where $V^{FT4}(t)$ is an individual parameter, varying over time as bodyweight increases with age.

### Model 1: Body weight as TVC:

The time-dependent prior of $V^{FT4}$ is captured within the NLME formulation as

$$\log\left(V_i^{FT4}(t)\right) = \log\left(V_{pop}^{FT4}\right) + \beta_V^W \log\left(\frac{W_i(t)}{W_{ref}}\right) + \eta_{V,i}, \quad \eta_{V,i} \sim \mathcal{N}(0, \omega_V^2).$$

### Model 2: Body weight as second output (non-generative, time-independent prior):

Weight dynamics is modeled explicitly using the Leffler function

$$W(t) = \begin{cases} \alpha_1 + \frac{\alpha_2}{180} t & \text{for } 0 \leq t \leq 180 \\ \alpha_1 + \alpha_2 + \frac{\alpha_3}{180}(t - 180) & \text{for } t > 180 \end{cases}$$

and $FT4$ volume is linked to body weight via

$$\log\left(f_{V,i}^{FT4}\right) = \log\left(f_{V,pop}\right) + \eta_{f_V,i}, \quad \eta_{f_V,i} \sim \mathcal{N}(0, \omega_{f_V}^2)$$

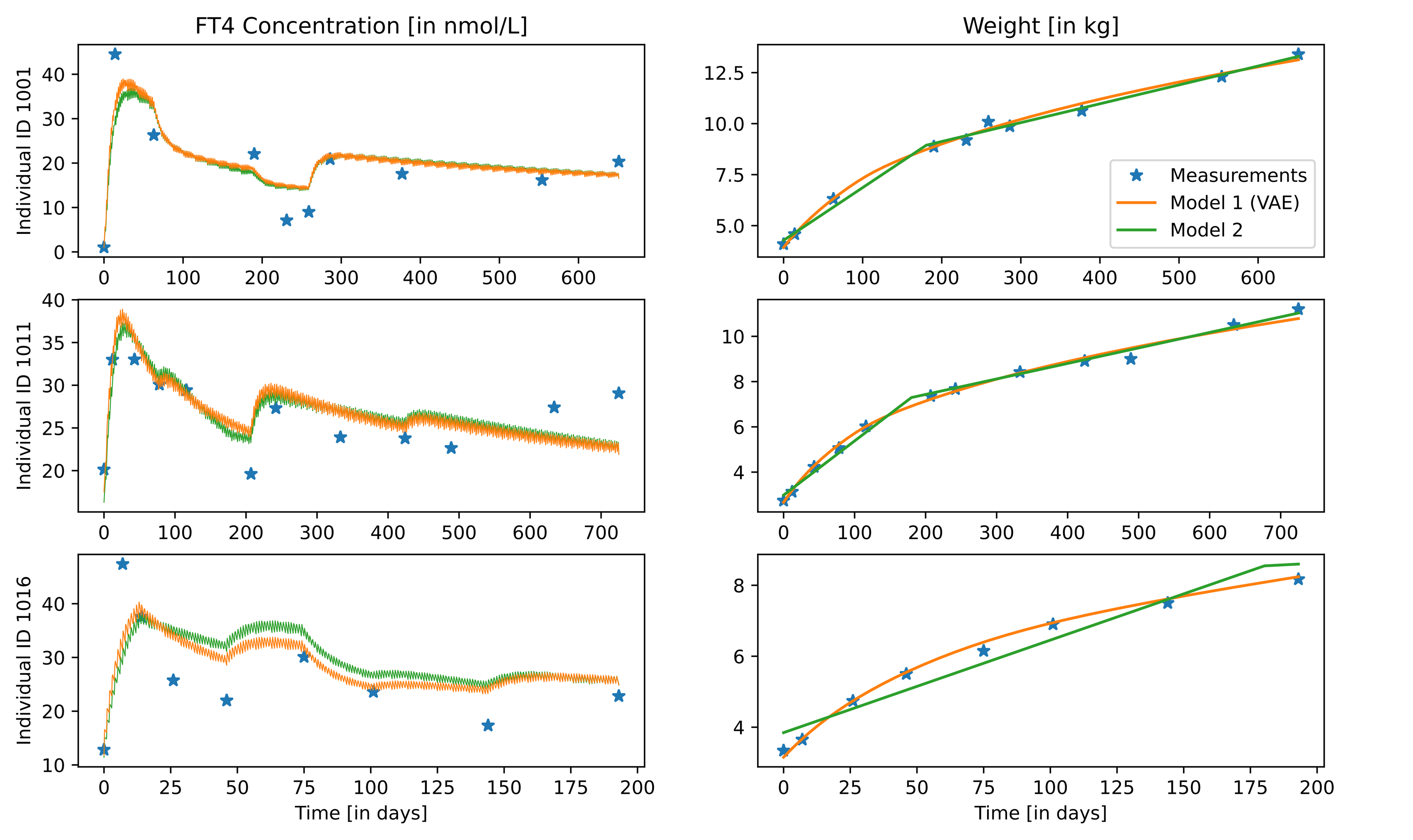
$$\log\left(V_i^{FT4}(t)\right) = \log\left(f_{V,i}\right) + \beta_V^W \log\left(\frac{W_i(t)}{W_{ref}}\right).$$

## Results

Parameter estimates and statistical criteria for model 1 with time-varying covariate weight (first row shows the results of the VAE-NODE approach, the second row those of the SAEM with linear interpolation and $W(t)$ as regressor). The third row displays the results of the expert-based growth model 2 (VAE and SAEM solution approximately coincide, $-2\mathcal{LL}$ and BICc corrected).

| | $V_{pop}/f_{V,pop}$ | | $k_{endo,pop}$ | | $\beta_V^W$ | | $\omega_V/\omega_{f_V}$ | | $\omega_{k_{endo}}$ | | $b$ | | $a_\Lambda/a$ | | $-2\mathcal{LL}$ | BICc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Estimate** | Value | R.S.E. | Value | R.S.E. | Value | R.S.E. | Value | R.S.E. | Value | R.S.E. | Value | R.S.E. | Value | R.S.E. | | |
| **VAE-NODE** | 4.21 | **0.35** | 3.71 | **13.8** | 0.89 | **1.13** | 0.21 | **3.59** | 1.06 | **9.41** | **0.228** | **2.25** | **0.045** | 5.61 | **3047** | **3087** |
| **SAEM-Lin.** | 4.28 | 3.35 | 4.19 | 15.8 | 0.81 | 4.01 | 0.21 | 12.7 | 1.11 | 10.3 | 0.267 | 3.70 | – | – | 3355 | 3388 |
| **Model 2** | 4.23 | 3.33 | 4.21 | 14.2 | 0.81 | 4.33 | 0.21 | 13.9 | 1.02 | 9.97 | 0.278 | 4.17 | 0.06 | **4.59** | 3523 | 3563 |

Fits of the $FT4$ concentration $C^{FT4}$ (left) and the predicted body weight $W$ (right) for individual IDs 1001, 1011 and 1016.




The VAE-NODE not only matched the expert-defined model but improved the likelihood substantially while decreasing the RSE values. It successfully recovered plausible growth-age trajectories that closely resemble the World Health Organization standards [2].