

Machine Learning algorithms for concentration prediction in anticancer drug therapy

Olga Teplytska^{1*}, Diego Valderrama^{2*}, Luca Marie Koltermann¹, Elena Trunz³, Alina Pollehn³, Eduard Schmulenson¹,

Achim Fritsch¹, Jan Hasenauer^{4,5}, Holger Fröhlich^{2,6#}, Ulrich Jaehde^{1#}

¹ Department of Clinical Pharmacy, Institute of Pharmacy, University of Bonn, Bonn, Germany

² Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany

³ Institute of Computer Science II, Visual Computing, University of Bonn, Bonn, Germany

⁴ Hausdorff Center for Mathematics, University of Bonn, Germany

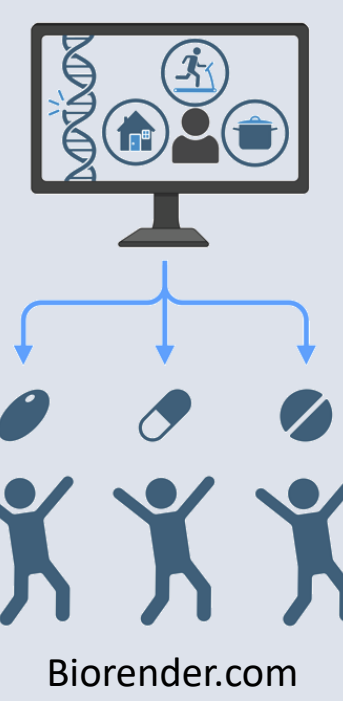
⁵ Medical Sciences-Institute (LIMES), University of Bonn, Germany

⁶ Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany

*shared first authorship, #shared senior authorship

Introduction

- In anticancer drug therapy, an individual dose must be carefully selected to avoid under- or overdosing of patients.
- Therapeutic Drug Monitoring (TDM), i.e. individual dosing guided by measured drug plasma concentrations can be applied, but empirical models are needed.
- Novel Machine Learning (ML) methods can be advantageous for predicting concentrations because they do not require assumptions on pharmacokinetic processes and influential covariates.



Aim

Our goal was to develop Scientific Machine Learning (SciML) models to predict plasma concentrations of anticancer drugs and compare them with population pharmacokinetic (Pop-PK) and classical Machine Learning (ML) models. We used clinical data of fluorouracil (5FU) as an example of an iv administered drug (540 samples) and sunitinib as an example of a po administered drug (302 samples).

Methods

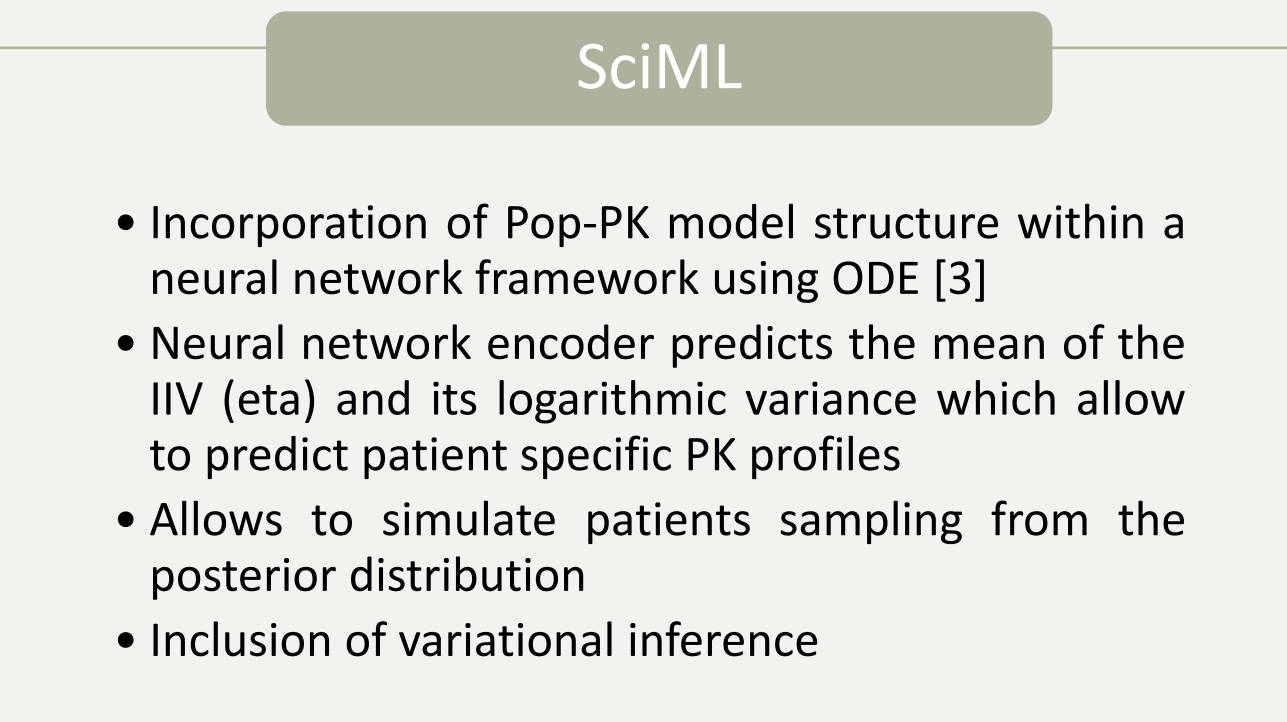
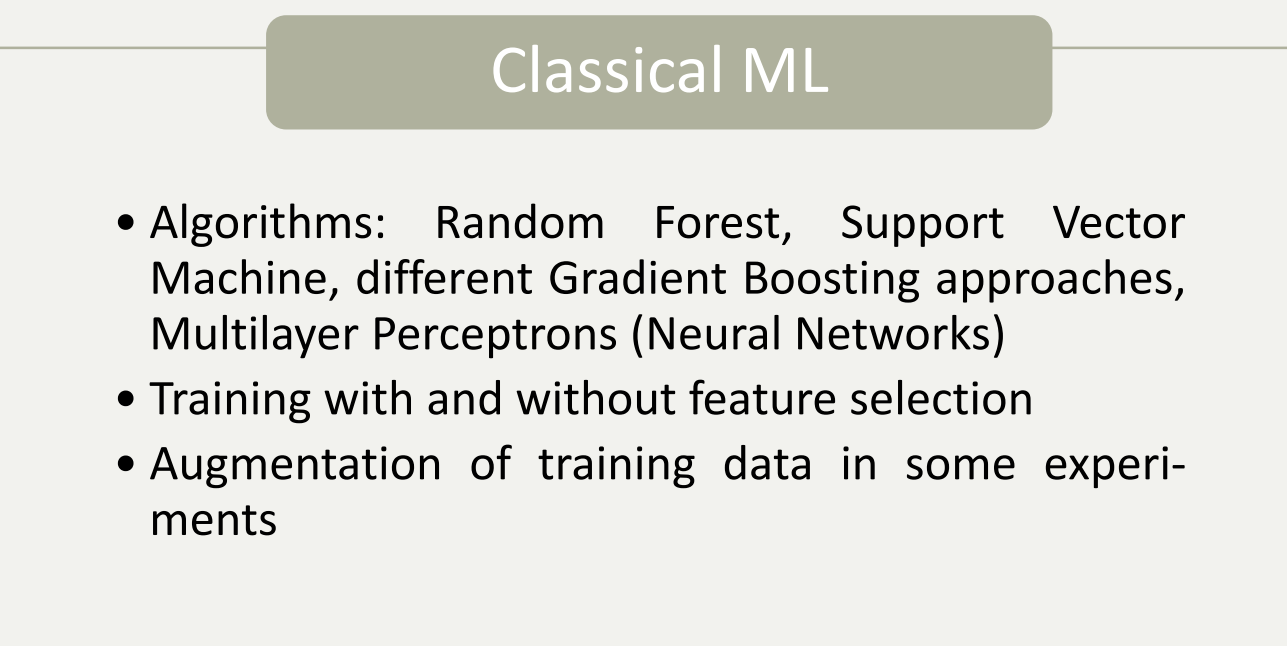
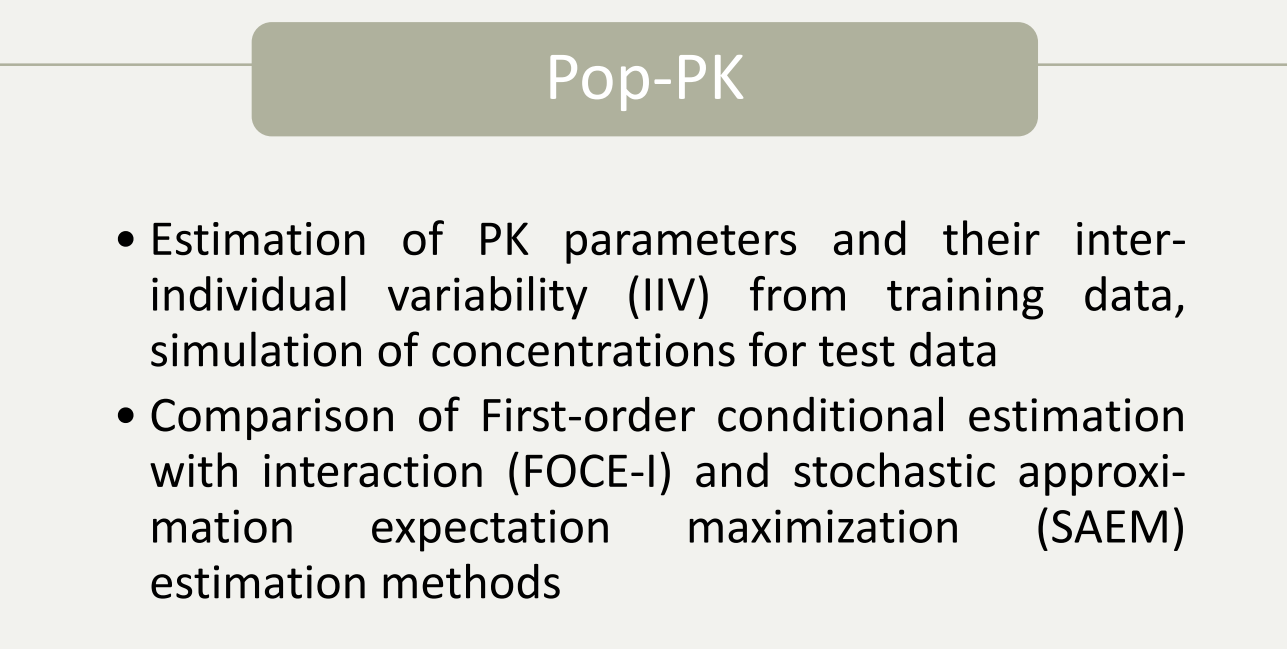
Baseline patient characteristics (median and range)
Fluorouracil (5FU)

Demographics (<i>n</i> = 156)	
Sex M/F	96/60
Age (years)	64.5 (35-83)
Body surface area (m ²)	1.915 (1.35-2.85)
Therapy related details	
5FU dose (mg) ^a	4000 (2700-5720)
5FU AUC (mg·h/L) ^b	18.75 (8.1-92.3)
Tumor entity	
Colorectal cancer	78
Gastroesophageal cancer	48
Pancreatic cancer	16 [1]

Sunitinib ^c		
	mRCC patients ^d (<i>n</i> = 26)	mCRC patients ^d (<i>n</i> = 21)
Sex M/F	25/1	12/9
Age (years)	64 (43–75)	61 (33–85)
Weight (kg)	83 (65–106)	73 (57–106)
Height (cm)	180 (155–186)	172 (149–184)
BMI (kg/m ²)	25.7 (22.5–34.5)	26.0 (13.3–39.3)

a Administered as continuous infusion over 24 h
b Calculated by multiplying the infusion time with the measured steady-state concentration
c Administered in a 4-weeks on, 2-weeks off regimen
d mCRC, Metastasized colorectal cancer; mRCC, Metastasized renal cell carcinoma

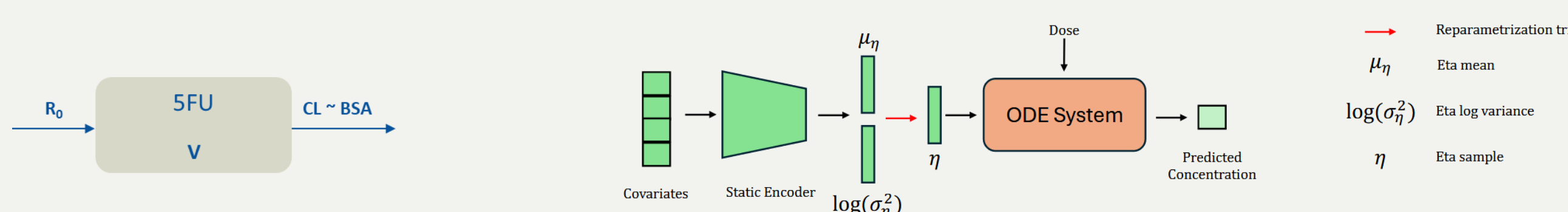
Algorithms for drug concentration prediction



10-fold train-test split (80/20) for each method

Pharmacokinetic and SciML model structures

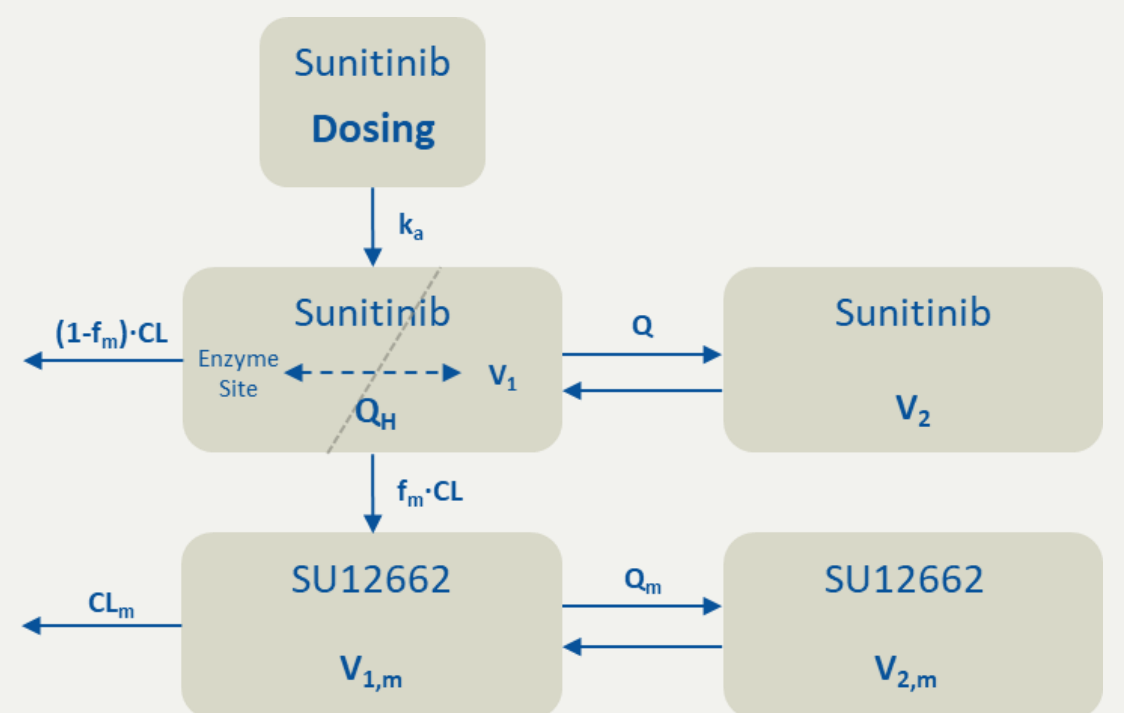
Fluorouracil (5FU)



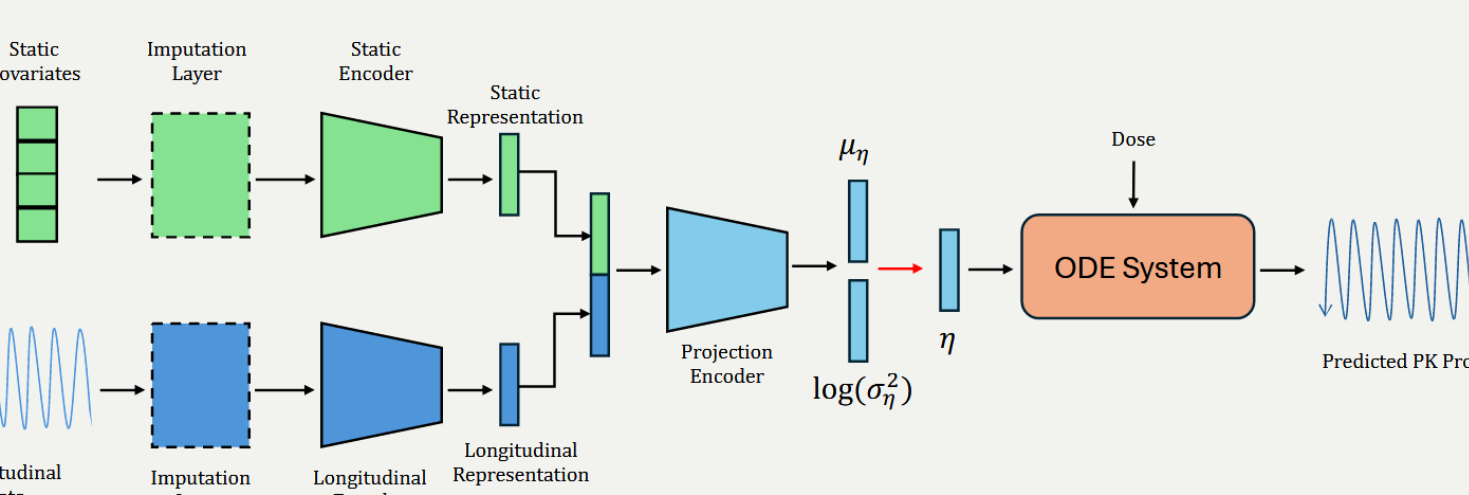
- One-compartment model
- Body surface area (BSA) (centered on median) is a linear covariate on CL [2]
- Fixed central volume of distribution V and IIV (V)

- Experiments done with fixed (MMPK-SciML*) and learnable population volume of distribution (MMPK-SciML)
- Loss: Evidence lower bound (ELBO) using a gaussian distribution as a prior

Sunitinib



- Two-compartment model
- Presystemic formation of the active metabolite SU12662 modelled via a hypothetical enzyme compartment (part of central compartment)
- Intercompartmental clearance, fraction metabolized and peripheral volume of distribution fixed



- Multimodal encoder containing three blocks: Static covariates, longitudinal covariates and projection encoder to combine both
- Loss: Different from Pop-PK modelling which uses a joined loss function, we only regularize our model using the ELBO with a gaussian prior for the parent concentration

Abbreviations:

BMI: Body mass index; CL_{mv}, clearance of the metabolite SU12262; f_m, fraction metabolized to SU12662; k_a, absorption rate constant; Q, inter-compartmental clearance of sunitinib; Q_{lv}, liver blood flow; Q_{mv}, intercompartmental clearance of the metabolite SU12662; R₀, infusion rate; V₁, volume of the central compartment of sunitinib; V_{1,m}, volume of the central compartment of the metabolite SU12662; V₂, volume of the peripheral compartment of sunitinib; V_{2,m}, volume of the peripheral compartment of the metabolite SU12662

References

- Schmulenson E, Zimmermann N, et al. Influence of the skeletal muscle index on pharmacokinetics and toxicity of fluorouracil. *Cancer Med* 2023; 12(3): 2580–9.
- Diekstra MH, et al. Population Modeling Integrating Pharmacokinetics, Pharmacodynamics, Pharmacogenetics, and Clinical Outcome in Patients With Sunitinib-Treated Cancer. *CPT Pharmacometrics Syst Pharmacol* 2017; 6(9): 604–13.
- Valderrama D, et al. Integrating machine learning with pharmacokinetic models: Benefits of scientific machine learning in adding neural networks components to existing PK models. *CPT Pharmacometrics Syst Pharmacol*. 2024; 13: 41-53.

Conclusion and Outlook

- A structural model is key to effectively capture the time course of plasma concentrations.
- SciML significantly outperformed Pop-PK models and classical ML algorithms in the case of 5FU, but Pop-PK methods performed better for sunitinib as the task was too complex for the sparse training data.
- In the next part of our collaborative project, we aim to incorporate our model architecture into frameworks for dose individualization, e.g. via Reinforcement Learning or Bayesian Optimization methods.

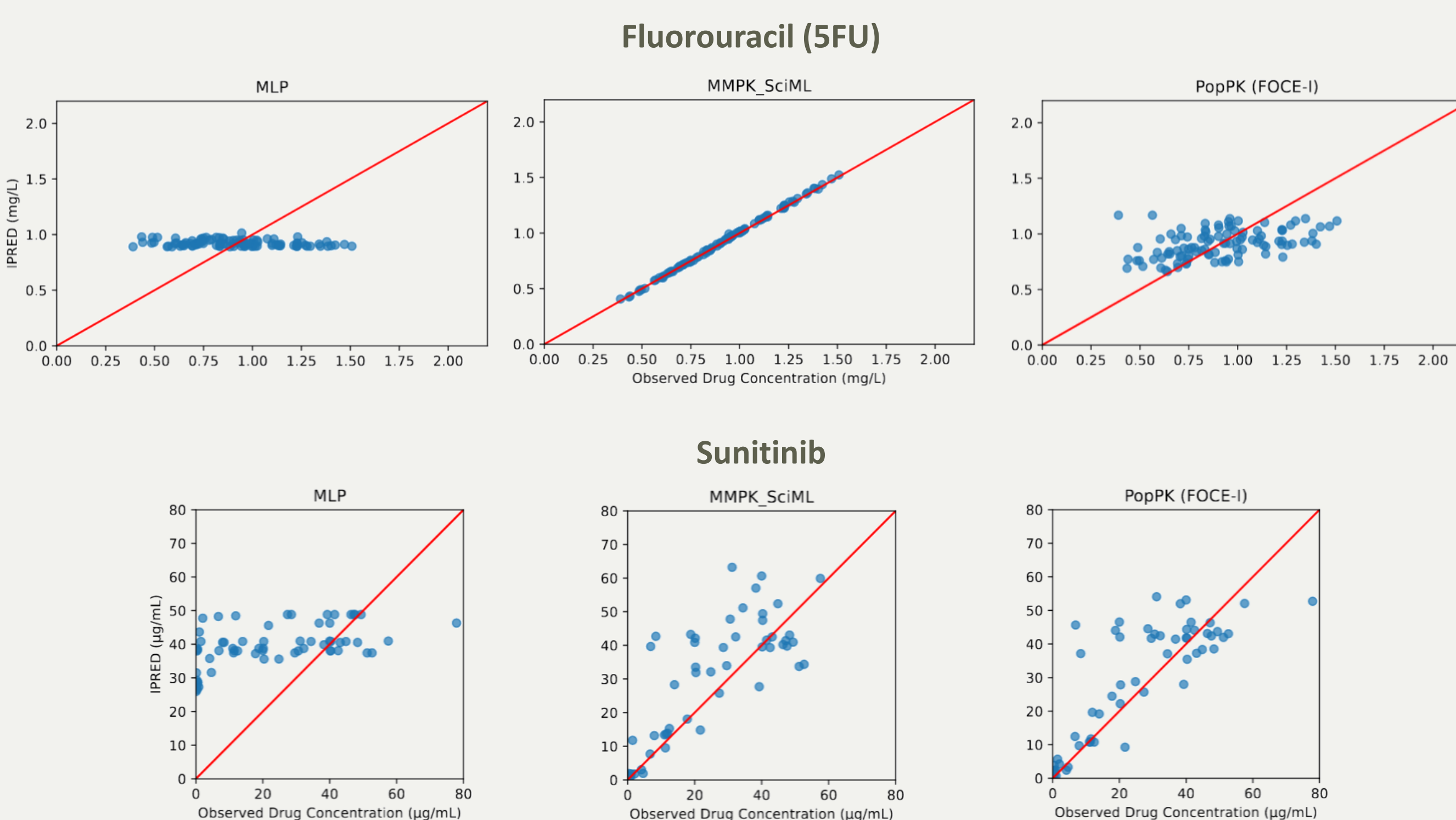
Results

Cross validation average metrics for Fluorouracil (5FU) and Sunitinib

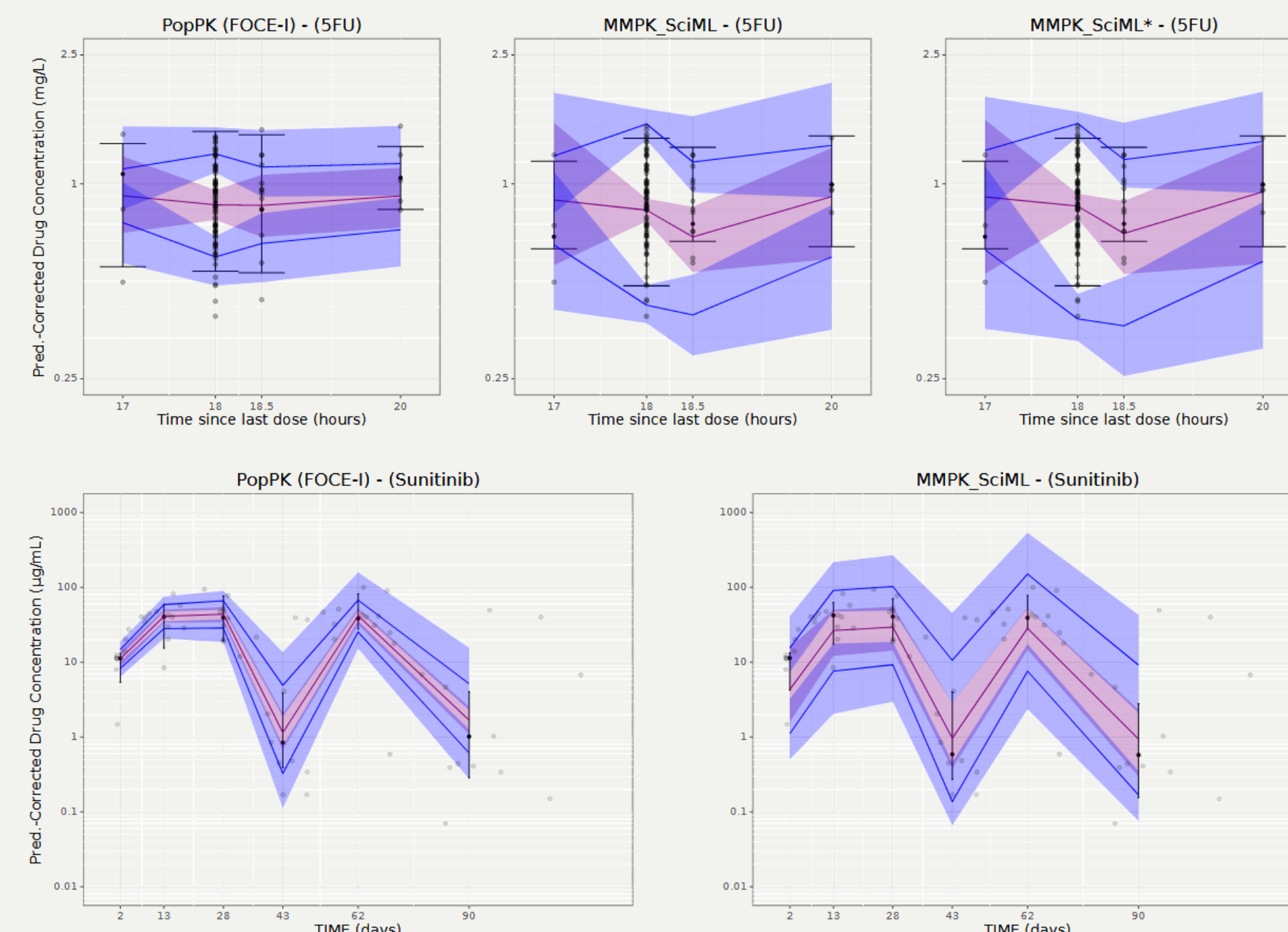
Model	Fluorouracil (5FU)		Sunitinib	
	MAE	RMSE	MAE	RMSE
Support Vector Machine	0.41 ± 0.09 (0.38 ± 0.05)	0.52 ± 0.08 (0.50 ± 0.07)	20.05 ± 2.65 (21.89 ± 3.54)	23.88 ± 3.40 (27.70 ± 5.32)
Random Forest	0.23 ± 0.02 (0.23 ± 0.03)	0.32 ± 0.12 (0.32 ± 0.12)	18.27 ± 2.35 (21.01 ± 3.48)	21.97 ± 3.15 (26.30 ± 5.29)
LightGBM	0.23 ± 0.02 (0.23 ± 0.02)	0.32 ± 0.12 (0.32 ± 0.12)	16.81 ± 1.67 (21.11 ± 3.36)	20.31 ± 1.98 (26.37 ± 5.19)
Multi-Layer Perceptron Two Hidden Layers	0.23 ± 0.03 (0.23 ± 0.03)	0.32 ± 0.11 (0.32 ± 0.12)	21.16 ± 3.59 (19.49 ± 3.72)	25.06 ± 4.22 (25.04 ± 5.13)
PopPK (FOCE-I)	0.22 ± 0.04	0.30 ± 0.12	9.68 ± 2.69	13.98 ± 3.80
PopPK (SAEM)	0.22 ± 0.04	0.31 ± 0.12	9.73 ± 2.75	14.12 ± 3.89
MMPK-SciML	0.03 ± 0.03	0.11 ± 0.16	12.55 ± 3.43	18.87 ± 5.12
MMPK-SciML *	0.04 ± 0.04	0.08 ± 0.12	-	-

Notes: For all algorithms we reported the Mean absolute error (MAE) and the Root Mean Squared Error (RMSE) as performance metrics. For the classical machine learning methods we depict the results of the versions with feature selection. In brackets we show the results with data augmentation. MMPK-SciML* refers to the model with a fixed volume for 5FU. Metrics of SciML are reported when using the sampling from the patient specific distribution for test subjects. Both metrics need to be as low as possible indicating small prediction errors.

Goodness of Fit Plots (selection)



pcVPC (selection)



Legend: Blue: 90%-confidence interval of the 10. and 90. percentile;
Pink: 90%-confidence interval of the median; Black dots: prediction-corrected concentrations

- In case of 5FU, the performance metrics as well as plots show a clear superiority of the SciML approach.
- For sunitinib, SciML performed significantly better than the classical ML approaches, but a bit worse than the PopPK model.
- There were no significant differences in parameter estimates and accuracy between FOCE-I and SAEM.
- Classical ML algorithms without a compartmental model structure were not able to learn the data.

Acknowledgements

We thank Ana Victoria Ponce Bobadilla for her input and explanations on how to create meaningful diagnostic plots for ML-enabled population PK models and Sven Stodtmann for the idea of encoding the concentration-time trajectory into parameters.

This work was partially funded by Federal Ministry of Education and Research within the projects „BNTrAinee“ (funding code 16DHBK1022).