Extending npde to evaluate Model Averaging: an application to viral dynamic models

S. El Messaoudi[@1] A. Lemenuel-Diot[3] J. Guedj[1] E. Comets[12]

[@]selma.el-messaoudi@inserm.fr, [1]Université Paris Cité, IAME, Inserm, F-75018, Paris, France,
[2]Université de Rennes, Inserm, EHESP, Irset-UMRS 1085, F-35000 Rennes, France,
[3]Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel

Context

Viral dynamics models

- Aim to provide a better understanding of viral infections by characterizing pathogenisis and by providing an estimation of treatment effects.
- Are poorly identifiable, and parameters are often fixed to arbitrary values.[1]
- Usually focus on one aspect of the disease, due to limited amount of data.[2]

Model averaging (MA)[3]

- Offer an alternative to model selection that takes into account uncertainty by combining the results of several candidate models.
- Provide better coverage rates for parameters with acceptable identifiability (relative bias $<50\%$) than model selection.

Objectives

We aim to extend and assess npde for evaluation of a model averaging framework.

Methods

Statistical model

Model for the observations defined as:

$$Y_{ijm} = f_m(t_j, \theta^b_{im}) + g_m(t_j, \theta^b_{im}) \times e_{ijm}$$

- f_m the structural model of model m
- g_m the error model of model m
- θ^b_{im} the vector of individual parameters under model m
- t_j the time of viral load measurement
- e_{ijm} the residual error.

Computation of MA

- M candidate models
- Weight w_m for model m proportional to AIC

$$w_m = \frac{e^{\frac{-\Delta AIC_m}{2}}}{\sum_{l=1}^{M} e^{\frac{-\Delta AIC_l}{2}}}$$

with $\Delta AIC_m = AIC_m - AIC_{min}$ $(l = 1, ..., m)$
and $AIC_{min} = min_{m=1,...,M}(AIC_m)$

Computation of normalised prediction distribution errors (npde)[4]

- Prediction discrepancies defined as the value of F_{ij} (cumulative distribution function (cdf) at observation y_{ij}).
- pd_{ij} approximated with Monte Carlo simulation using the design of an independant validation dataset v:

$$\widehat{pd}_{ij} = \frac{1}{K}\sum_{k=1}^{K} \delta_{ijk} \qquad \Rightarrow \qquad \widehat{pd}_{ij_{MA}} = \sum_{m=1}^{M} w_m \widehat{pd}_{ij_m}$$

with $\delta_{ijk} = 1$ if $y_{ij}^{sim(k)} < y_{ij}$ and 0 otherwise

- $\widehat{pd}_{ij_{MA}} \sim \mathcal{U}(0,1)$ when $K \to \infty$
- Decorrelation using the inverse of the cdf
- Normalisation

$$\widehat{npde}_{ij_{MA}} = \Phi^{-1}(\widehat{pde}_{ij_{MA}}) \sim \mathcal{N}(0,1)$$

- Statistical test: global test based on a combination of the mean, variance and distribution tests with a Bonferroni correction.

Evaluation of the performance of npde to evaluate models obtained by MA

Models used

- Four models used to characterize acute viral infections[5–8], taken as examples to evaluate MA in Gonçalves *et al*[9].
- Log-transformed viral loads (VL) $Y_{ij} = log_{10}VL$.
- Additive error model on Y_{ij} corresponding to a proportional model on VL.
- True parameters Ψ_0^m set for each model.[9]

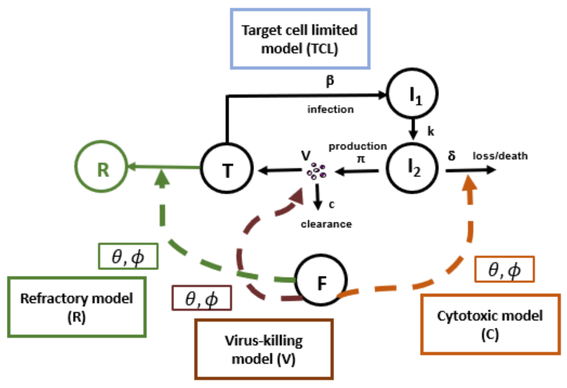

Figure 1: Schematic representation of the 4 models in this study.True parameters Ψ_m^0 defined in Gonçalves *et al*.[9].

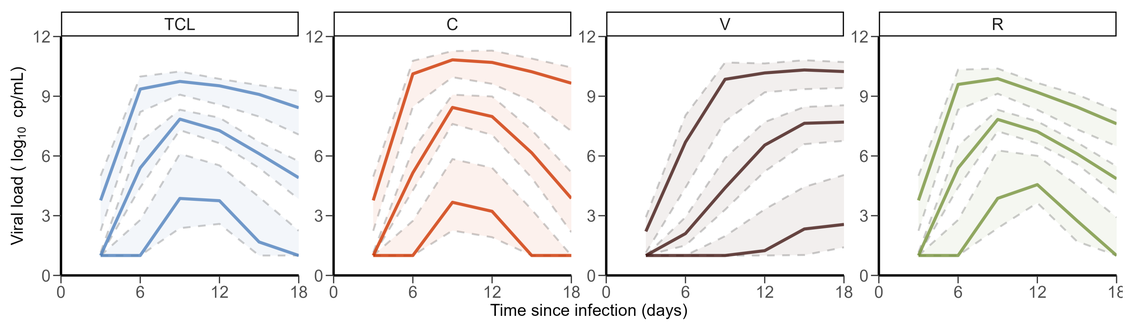

Figure 2: Envelope of 5th, 50th and 95th percentiles of simulated $log_{10}VL$ in 100 simulated datasets with 30 subjects under each model (see next section)

Simulation study

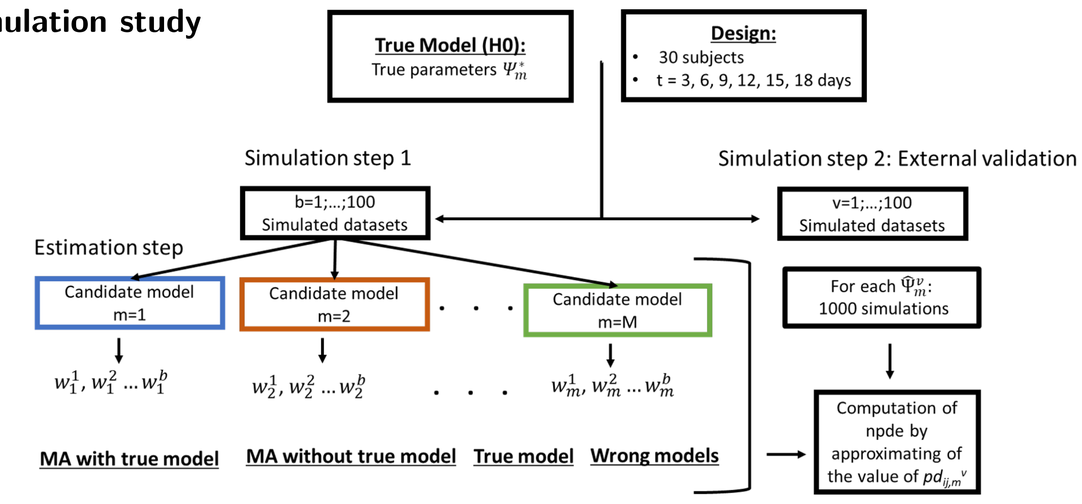

Figure 3: Workflow of the simulation-based study. b: building; v:validation.

Criteria used for evaluation

1. Rejection or acceptation of npde according to the global test with a p-value < 0.05.
2. Applied to the true model, MA with the true model, MA without the true model and wrong models.

Results

Illustration for one simulation under model C

- npde computed for one simulation under the different models.
- Similar patterns with true, MA or with wrong models (Figure 4).
- Predictive distributions overlayed ⇒ discrimination of tested versus true model difficult (Figure 5).

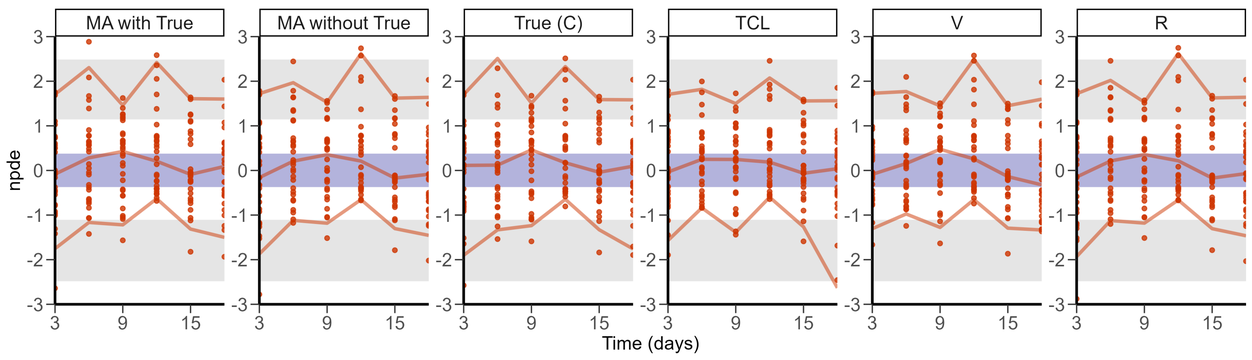

Figure 4: npde calculated with or without MA, with the true model(C) and the other models for a specific repetition.

Evaluation of the performance of npde

	Target-cell limited			Cytotoxic			Virus-killing			Refractory		
Rejection of npde (%)	MA	True model	Others	MA	True model	Others	MA	True model	Others	MA	True model	Others
Global	27	20	23-32	21	21	19-27	22	28	24-43	24	23	24-34
Fisher	31	23	23-28	27	28	23-26	27	31	27-42	23	21	19-27
SW	6	2	2-9	9	5	5-10	8	4	6-12	6	7	4-11
Wilcoxon	14	13	8-20	11	13	9-12	17	16	10-21	16	15	16-26

Table 1: Evaluation of MA and single models with external datasets.

- In most cases, true model selected with Model Selection (Table 1).
- Type I error α inflated to around 20% in the different simulations (rejection of the true model) ⇒ uncertainty from estimation step not taken into account ?
- Similar rejection rate whether using MA or wrong models ⇒ poor discrimination power.

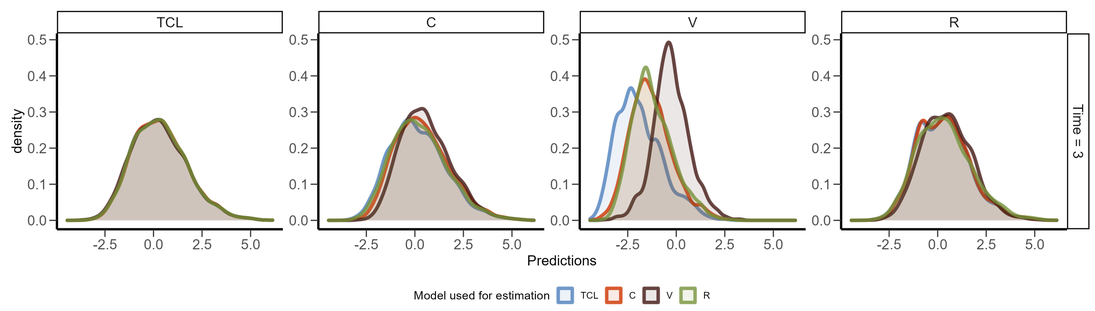

Figure 5: Predictive distribution of the observations of ID=1 at Time=3 with each model. Each panel corresponds to the true model.

Conclusion

- npde successfully extended to MA to provide diagnostic plots.
- However, failure to reject the wrong model and low discriminatory power.
  - Evaluation of (M, $\hat{\Psi}$) instead of (M, Ψ_0) in contrast to previous evaluations.
  - Estimation error not accounted for in the computation of pd.[10]
  - Models poorly identifiable.[9]
- Perspectives: account for estimation error to correct type I error inflation.[10]

[1] Guedj *et al. Bul Math Biol* 2007;
[2] Moore *et al. Bull Math Biol* 2018;
[3] Buatois *et al. AAPS* 2018;
[4] Comets *et al. AAPS* 2021;
[5] Madelain *et al. Nat Commun* 2018;
[6] Baccam *et al. J Virol* 2006;
[7] Pawelek *et al. PLoS Comp Biol* 2012;
[8] Li and Handel *J Theor Biol.* 2014;
[9] Gonçalves *et al.AAPS* 2020;
[10] Yano *et al.JPKPD* 2001;