

PCA-based modelling in antidepressant trials: a pre-mechanistic approach



University of Pavia

E. Marostica¹, A. Russu¹, S. Zamuner², R. Gomeni^{3*}, G. De Nicolao¹

¹ Department of Industrial and Information Engineering, University of Pavia, Pavia, Italy

² Clinical Pharmacology / Modelling and Simulation, GlaxoSmithKline, Stockley Park, UK

³ Pharmacometrics, GlaxoSmithKline, Upper Merion, PA, USA

* Current address: Pharmacometrics, La Fouillade, France



GlaxoSmithKline

Introduction

- For many diseases, such as the psychiatric ones, the knowledge about disease progression and interaction between disease and drug action is often very limited or absent. The empirical models, adopted to describe the evolution of clinical endpoints¹, are characterized by arbitrarily chosen basis functions and are usually dataset-specific.
- This motivates the development of a flexible and general-purpose "pre-mechanistic" modelling approach appropriate for exploratory analysis and as a touchstone for subsequent mechanistic model building. Along this direction, the aim of this work is to introduce a method, based on Principal Component Analysis (PCA), that automatically provides regression functions reflecting the informative content of the data: the **PCA-based modelling tool**².

PCA-based modelling

- Population model in vector form $Y_i = \bar{X} + \tilde{X}_i + \varepsilon_i$, $Y_i \in \mathbb{R}^n$: observed data of the i -th subject
 $\bar{X}$: typical response, $\tilde{X}_i$: random effect, $\varepsilon_i \sim N(0, \sigma^2 I)$: measurement error

- Use Y_i , $i = 1, \dots, N$ to estimate $\Sigma_Y = \text{Var}[Y_i]$
- Singular Value Decomposition: $\Sigma_Y = U_Y D_Y U_Y^T$, U_Y : matrix of principal vectors of Σ_Y .
Remark: $U_Y = U_X$ (matrix of singular vectors of $\Sigma_X = \text{Var}[\tilde{X}_i]$)
- PCA-based model: $Y_i = \bar{X} + U\eta + \varepsilon_i = \bar{X} + \sum_{k=1}^q \eta_k \phi_k + \varepsilon_i$
 $U = [\phi_1 \dots \phi_q \dots \phi_n]$, $\phi_k \in \mathbb{R}^n$, $\eta \sim N(0, \Omega)$, $\Omega = \text{diag}\{\omega_1^2, \dots, \omega_q^2\} = D_Y - \sigma^2 I$
- Model order selection: use crossvalidation or complexity criteria (e.g. Mallows' Cp³ that accounts for goodness-of-fit and model complexity) to select $q \leq n$ components.

Validation strategy

- Population analyses of simulated and experimental data were performed.
- Three parametric models were employed to simulate a total of 150 datasets (100 subjects each)

Inverse Bateman (IB)

Weibull + linear (WL)

Weibull (W)

$$y(t) = A + s(e^{-k_{ad}t} - e^{-k_{el}t})$$

$$y(t) = Ae^{-(t/t_d)^b} + h_{rec}t$$

$$y(t) = Ae^{-(t/t_d)^b}$$

- The number of components was decided by selecting the lower order model between the ones minimizing the Mallows' Cp³ or the crossvalidatory Root Mean Square Error (RMSE) with respect to a validation set made of observations not used for the model identification.
- Parameter estimation was carried out with R 2.13.1⁴, according to the empirical Bayes paradigm.

Results: experimental data

- The experimental dataset was obtained from a Phase II depression trial (Figure 1). Only subjects in the placebo arm were considered.

Experimental Dataset

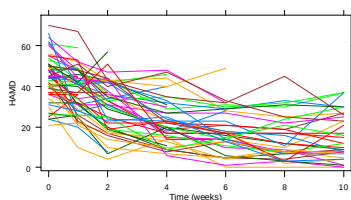


Figure 1: Spaghetti plot of experimental HAM-D data

- In the experimental scenario, the PCA-based model with 3 principal functions was chosen according to the order selection procedure. The 3 principal functions obtained from the experimental dataset are shown in Figure 2.

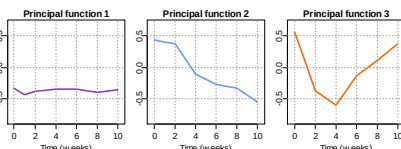


Figure 2: Principal functions of PCA with 3 components

- The proposed approach achieved very satisfactory individual fittings (Figure 3) and crossvalidatory performances (Figure 4) on the experimental data.

Individual fittings

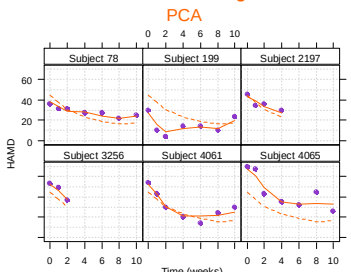


Figure 3: Individual fittings (solid line) and typical curve (dashed line) of a subset of subjects, obtained with the PCA-based model with 3 components

Prediction Analysis

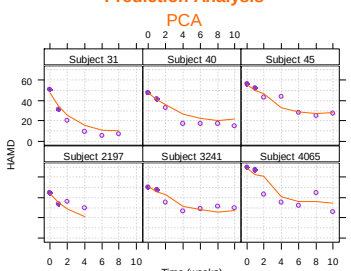


Figure 4: Crossvalidatory performances of the PCA-based model on experimental data. The filled violet points represent the training set, whereas the empty points are the validation set

Results: simulated data

- Figure 5 shows the crossvalidatory performances of the PCA-based model and the parametric ones on the data simulated with each one of the parametric model. Each parametric model was tested on the dataset generated with the model itself. Good crossvalidatory performances were obtained by means of the new approach.

Prediction Analysis

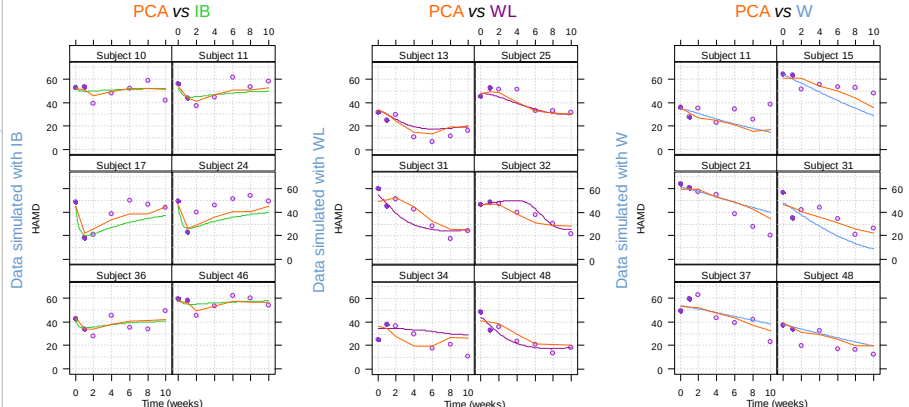


Figure 5: Crossvalidatory performances of the PCA-based model and the parametric ones on the data simulated with their own generating model. The filled violet points represent the training set, whereas the empty points are the validation set. PCA (orange) vs IB (green) (left panel), PCA (orange) vs WL (purple) (middle panel) and PCA (orange) vs W (blue) (right panel)

- The 3 parametric models and the new approach based on PCA have been evaluated in terms of "denoising", i.e. the ability to reconstruct the true individual profiles, and predictive performances. Figures 6-7 summarize the overall performance on the complete set of simulated data for each of the four considered models.
- The PCA-based approach provided satisfactory fitting performances, in terms of the RMSE with respect to the true profile, in all the 150 simulated datasets (Figure 6).
- Moreover, the proposed approach yielded good predictive performances, in terms of the crossvalidatory RMSE with respect to the true profile, in all the 150 simulated datasets (Figure 7).

Denosing: RMSE w.r.t. true profile

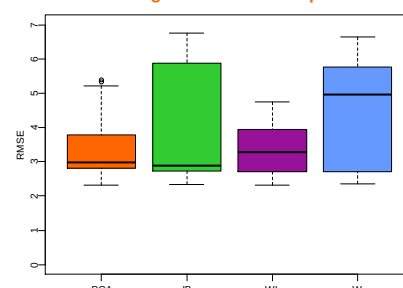


Figure 6: Boxplot of RMSE obtained with the PCA-based model and the 3 parametric ones.

Prediction of missing data: RMSE w.r.t. true profile

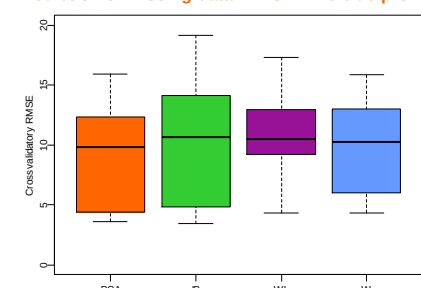


Figure 7: Boxplot of the crossvalidatory RMSE obtained with the PCA-based model and the 3 parametric ones.

Conclusions

- The proposed PCA-based approach can be valuable when the knowledge of the clinical and physiopathological mechanisms underlying the changes of clinical score over time is limited or absent.
- It automatically provides basis functions suitable to develop parsimonious population models and yields reliable reconstructions of individual profiles.
- This approach is useful for exploratory analysis and as a touchstone in order to benchmark the performances of mechanistic models.

References

- Gomeni R, Lavergne A, Merlo-Pich E (2009), Modelling placebo response in depression trials using a longitudinal model with informative dropout, European Journal of Pharmaceutical Science 36, pp. 4-10.
- Abdi H, Williams Lynne J (2010), Principal component analysis, WIREs Computational Statistics 2, pp. 433-459
- Mallows C L (1973), Some comments on Cp, Technometrics 15, pp. 661-675
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2011). <http://www.R-project.org>.