ELSEVIER

International Journal of Pharmaceutics 226 (2001) 107–125

international journal of pharmaceutics

www.elsevier.com/locate/ijpharm

Application of linear mixed effects models to the evaluation of dissolution profiles

E. Adams [a,*], D. Coomans [b], J. Smeyers-Verbeke [a], D.L. Massart [a]

[a] *Vrije Universiteit Brussel, Pharmaceutical Institute, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *James Cook University of North Queensland, Department of Mathematics and Statistics, Townsville Q4811, Australia*

Received 5 April 2001; received in revised form 18 June 2001; accepted 19 June 2001

Abstract

The performance of linear mixed effects models for the comparison of dissolution profiles is examined. This type of model is frequently used by statisticians, but is rather unknown to people that work in dissolution laboratories. Hence, an extensive theoretical part was introduced to make the methodology more accessible. Firstly, repeated measures ANOVA is discussed, followed by the 'real' linear mixed effects models. The theory is applied to two types of dissolution data: one corresponding to an immediate and another to a slow release formulation. We tried to use as much as possible the standard settings of the statistical software (S-plus). Suggestions are given to solve problems encountered during model fitting. It was found that the statistical limits are much more discriminative than the similarity factor.

Keywords: Dissolution; Linear mixed models; Pharmaceutical statistics; Longitudinal data

1. Introduction

The evaluation of dissolution profiles has an extensive application in the drug development and manufacture of tablets and capsules. It is used during the optimization of the pharmaceutical dosage form, stability tests, the control of the production process and the establishment of the similarity between a new or generic formulation and an existing one. It is also useful to develop in vitro–in vivo correlations. In the guidance for scale-up and post approval changes (SUPAC) of the Food and Drug Administration (FDA), it is mentioned that for certain moderate changes, similarity between two batches can be established by using only in vitro dissolution testing (FDA Guidance for Industry, 1995, 1997).

In most cases, two batches (a reference and a test batch) have to be compared with each other. In dissolution testing, the percentage of the active component dissolved is measured at certain time points. To obtain meaningful conclusions, it is better to evaluate multiple time points or the entire dissolution profile rather than to check at only one time point whether a minimum percentage is dissolved. Several methods to compare dis-

* Corresponding author. Tel.: +32-247-74723; fax: +32-247-74735.
E-mail address: eadams@fabi.vub.ac.be (E. Adams).

0378-5173/01/$ - see front matter
PII: S0378-5173(01)00775-X

solution profiles were described. They are often divided into ANOVA-based methods, model-dependent and model-independent methods.

Although ANOVA-based methods rely on an underlying model, they are not considered as model-dependent since they do not really fit a curve through the observed data. They test the dissolution profiles for statistical differences in 'level' (size) and 'shape' (parallelism) (Mauger et al., 1986; Yuksel et al., 2000). The results (P-values) showed that these methods were too discriminative and investigated statistical rather than pharmaceutical equivalence (Polli et al., 1996, 1997; O'Hara et al., 1998).

Other model-independent techniques are the computation of the difference factor f_1 and the similarity factor f_2, originally described by Moore and Flanner (1996) and now recommended (especially f_2) by the FDA (FDA Guidance for Industry, 1995, 1997). Values between 0 and 15 for f_1 and between 50 and 100 for f_2 indicate equivalent dissolution profiles. The f_2 factor can be calculated as:

$$f_2 = 50 \times \log\left\{\left[1 + \frac{1}{n}\sum_{t=1}^{n} w_t(R_t - T_t)^2\right]^{-0.5} \times 100\right\} \quad (1)$$

with R_t and T_t the percentage dissolved at time t (for $t = 1, 2, \ldots, n$) for the reference and the test batch, respectively, and w_t the optional weight factor (mostly $w_t = 1$). Two sets are considered equivalent when there is no more than an empirically determined 10% average difference at any sample time point. This yields a minimal f_2 factor of 50.

Tsong et al. (1996) calculated the Mahalanobis distance between the respective means of the reference and the test batch using the pooled variance–covariance matrix. Another possibility to compare dissolution profiles is the calculation of the areas under the curve (AUC) or the dissolution efficiencies (Anderson et al., 1998). When the AUC is calculated using the trapezoidal rule, the method is model-independent. When first a curve is fitted, it is model-dependent. Recently, the use of principal component analysis for the evaluation of dissolution profiles was described (Adams et al., 2001). The first two principal components allow us to detect differences in level and shape. Confidence limits can be constructed using the bootstrap technique.

The so-called model-dependent methods all use some kind of curve fitting procedure. Different mathematical fit functions were investigated: the zero and first-order, Hixson–Crowell, Higuchi, quadratic, Weibull, Gompertz and logistic (Yuksel et al., 2000; Polli et al., 1997). All these models are employed to reduce the dimensions of the observed data to a minimal number of parameters. Although this simplifies the data handling and interpretation (profiles can statistically be compared through these model parameters), it also implies that the models (with mostly only 1 or 2 parameters) are not very flexible. It is even possible that a function fits the data of the reference set well while it gives inaccurate results for the test set, due to a different shape of the dissolution profiles. Although none of the functions described above are suitable to fit all types of dissolution curves, the Weibull seems to be one of the better choices. Another way of modeling is the use of mixed effects models. They show a greater flexibility and provide more statistical information compared to mathematical functions like the Weibull. A distinction can be made between linear (LME) and non-linear mixed effects (NLME) models. NLME models suppose that the underlying mechanism that produces the response (here: the dissolution process) is known. As an interesting consequence, the model parameters have a physical interpretation. An example of this mechanistic model applied to dissolution data was worked out by Crowder (1996) using specialized software. Although the results were good, the method is complicated and difficult to implement for people who are not familiar with this type of model.

In this paper, the application of LME models to dissolution data is examined using standard statistical software. Since most of the information that can be found about these models is very statistically orientated, a comprehensive theoretical part is added to introduce this technique to non-statisticians. The theoretical aspects will be illustrated by analyzing real pharmaceutical dissolution data.

2. Theory

2.1. Introduction

Throughout this text matrices are represented as capital bold letters and vectors as small bold letters. Scalars are indicated with small letters not in bold. Predicted or estimated values for a vector $\boldsymbol{x}$ or a matrix $\boldsymbol{X}$ are represented as $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{X}}$, respectively. The transpose $\boldsymbol{X}^{\mathrm{T}}$ of a matrix $\boldsymbol{X}$ is obtained by interchanging the rows and columns of $\boldsymbol{X}$. The inverse $\boldsymbol{X}^{-1}$ of a (non-singular) square matrix $\boldsymbol{X}$ is such that $\boldsymbol{X}\,\boldsymbol{X}^{-1} = \boldsymbol{X}^{-1}\,\boldsymbol{X} = \boldsymbol{I}$, where $\boldsymbol{I}$ is an identity matrix with all elements on the principal diagonal (from top-left to bottom-right) equal to 1 and all other elements being zero.

2.2. Mixed model analysis

Mixed model analysis allows us to investigate factors whose levels can be controlled by the researcher (fixed) as well as factors whose levels are beyond the researcher's control (random). For dissolution experiments, this can be illustrated as follows:

$$\begin{bmatrix} \mathrm{Batch}^{\mathrm{F}} \\ | \\ \mathrm{Tablet}^{\mathrm{R}} \end{bmatrix} \times \mathrm{Time}^{\mathrm{F}}$$

Every tablet studied belongs to a certain, fixed (F) batch (tablet nested in batch). From each batch studied, a number of tablets are chosen at random (R) and the dissolution curves are determined. The factor 'Time' is crossed with 'Tablet' (which is nested in 'Batch') since the percentage drug dissolved is determined for each selected tablet at previously determined, fixed (F) time points. The combination of fixed and random effects points to the mixed character.

In general, there are two possible approaches for the statistical comparison of two or more data sets. One can take one set as the reference, derive confidence limits and check if the test set falls within these limits. Another possibility is to examine whether the differences between the sets to be compared are significantly different from zero. Mixed effects models are more suitable for the latter approach.

2.3. ANOVA-based methods for longitudinal data

These methods are based on the classical, well-known ANOVA methods. Since for each tablet the percentage dissolved is measured at several time points, the analysis can be considered as a repeated measures ANOVA. This approach is known as the level and shape approach (Mauger et al., 1986). Although they are in general not considered as LME models, the underlying model for dissolution data can be written as follows:

$$y_{hij} = \mu + \beta_h + \gamma_{hj} + \tau_j + U_{hi} + Z_{hij} \qquad (2)$$

for $h = 1,\ldots, g$; $i = 1,\ldots, n_k$ and $j = 1,\ldots, p$ with g the number of batches, n_k the number of tablets in batch k, $N = \Sigma_{k=1}^{g} n_k$ the total number of tablets and p the number of measurement points. y_{hij} represents the measurement result at time j for tablet i belonging to batch h, μ is the overall mean of all measurements, β_h the effect of batch h on the result where $\Sigma_{h=1}^{g}\beta_h = 0$, γ_{hj} the interaction between batch h and time j where $\Sigma_{j=1}^{p}\gamma_{hj} = 0$ for each batch h and τ_j the effect of time j where $\Sigma_{j=1}^{p}\tau_j = 0$. U_{hi} represents the (random) effect of tablet i (belonging to batch h) and Z_{hij} the (random) measurement error. Two assumptions are made: $U_{hi} \sim N(0,v^2)$ and $Z_{hij} \sim N(0,\sigma^2)$. This must be interpreted that U_{hi} and Z_{hij} are both normally distributed with mean zero and variances v^2 and σ^2, respectively. In some statistical software programs (like SPSS), μ represents the mean of one of the batches. When μ is equalized to the mean of batch 2, this means in fact that β_2 is set to 0 and that β_1 expresses the difference between both batches. It is also required that the reference and the test batch are measured at the same time points. Since the time scale is not effectively used in the calculations, unequally spaced measurement points are considered as equally spaced.

A typical ANOVA table for the study of dissolution data is given in Table 1. The sources of variation can be subdivided in 'between-subjects' and 'within-subjects' factors (here: subjects = tablets) (SPSS, 1997). A 'between-subjects' factor is any factor that divides the subjects into discrete subgroups (here: batches). A 'within-subjects' factor is any factor that distinguishes measurements made on the same subject (here: each factor in

connection with the time points at which each of the tablets is measured). For each factor, the sum of squares (SS) can be calculated (Massart et al., 1997). The mean square (MS) is a variance estimate that is obtained by dividing the sum of squares (SS) by the corresponding degrees of freedom (d.f.). The following five terms can be studied (see also Table 1):

- *Batch*: allows us to study the difference in level or size between the mean profiles of the batches.
- *Tablet (Batch)*: is used to calculate the error term for the 'between-subjects' factors.
- *Time*: in fact this factor gives an answer to the question: do the tablets dissolve as a function of time? Probably they do, so this factor is not very interesting to study.
- *Time*Batch*: this interaction allows us to study the difference in shape (time profile) between the mean profiles of the batches.
- *Time*tablet (Batch)*: is used to calculate the error term for the 'within-subjects' factors.

So, the most interesting factors to study are *Batch* and *Time*Batch*. Correlated to model Eq. (2), the first *F*-test ($MS_B/MS_{Ta(B)}$) tests the hypothesis $\beta_h = 0$ (for $h = 1, \ldots, g$) or the difference in level or size between the batches. The other interesting *F*-test ($MS_{Tm*B}/MS_{Tm*Ta(B)}$) tests the hypothesis $\gamma_{hj} = \gamma_j$ (for $h = 1,\ldots,g$ for each time point $j = 1,\ldots, p$) or if the mean profiles are parallel (have the same shape). Although both a univariate and a multivariate approach for the study of the within-subjects effect (shape) are possible, the univariate is preferred (Yuksel et al., 2000; SPSS, 1997). This univariate repeated measures ANOVA requires certain assumptions about the homogeneity of the variance–covariance matrix: it assumes on the one hand that the variances are equal and on the other hand that the covariances between the within subjects (measurement times) are equal too. These typical assumptions can be tested by Mauchly's test of sphericity. When they are not met, a correction factor like the Greenhouse–Geisser ε can be used to adjust the degrees of freedom in the *F*-test.

2.4. Linear mixed effects models

A linear mixed effects model is in general represented in a somewhat other way than Eq. (2) and can be defined as:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \boldsymbol{\varepsilon}_i \quad (3)$$

with the assumptions: $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$, $\varepsilon_i \sim N(0, \Sigma_i)$ and $\boldsymbol{b}_1,\ldots, \boldsymbol{b}_N$, $\varepsilon_1,\ldots, \varepsilon_N$ independent.

Eq. (3) must be interpreted as follows: $\boldsymbol{y}_i$ is the $(n_i \times 1)$ response vector for subject i (here: tablet i), with n_i the number of time points at which subject i is measured for $1 \le i \le N$ and N the number of subjects. $\boldsymbol{X}_i$ $(n_i \times p)$ and $\boldsymbol{Z}_i$ $(n_i \times q)$ are (known) design matrices for the p fixed and q random effects, respectively. $\boldsymbol{X}_i$ $(n_i \times p)$ equals $\boldsymbol{Z}_i$

Table 1
Example of an ANOVA table for the comparison of dissolution profiles

Source of variation	d.f.	MS	*F*-test	Model (2)
'Between-subjects'				
Batch [B]	$g-1$	MS_B	$MS_B/MS_{Ta(B)}$	β_h
Tablet [Ta(B)] (nested in batch)	$N-g$	$MS_{Ta(B)}$		U_{hi}
'Within-subjects'				
Time [Tm]	$p-1$	MS_{Tm}	$MS_{Tm}/MS_{Tm*Ta(B)}$	τ_j
Time*batch [Tm*B]	$(p-1)(g-1)$	MS_{Tm*B}	$MS_{Tm*B}/MS_{Tm*Ta(B)}$	γ_{hj}
Time*tablet [Tm*Ta(B)] (nested in batch)	$(p-1)(N-g)$	$MS_{Tm*Ta(B)}$		Z_{hij}
Total	$Np-1$			

The corresponding term of model Eq. (2) is given in the last column. d.f. = degrees of freedom, MS = mean square, g = number of batches, N = total number of tablets, p = number of time points.

$(n_i \times q) \times \boldsymbol{M}_i\,(q \times p)$ where matrix $\boldsymbol{M}_i$ indicates to which batch tablet i belongs. $\boldsymbol{\beta}$ $\boldsymbol{(p \times 1)}$ is the vector containing the p fixed effects, $b_i(q \times 1)$ the vector containing the q random effects and $\boldsymbol{\varepsilon}_i$ $(n_i \times 1)$ the vector of the residual components or random errors. Furthermore, $\boldsymbol{b}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be normally distributed with mean vector 0 and covariance matrices $\boldsymbol{D}$ $(q \times q)$ and $\boldsymbol{\Sigma}_i$ $(n_i \times n_i)$, respectively. $\boldsymbol{\Sigma}_i$ depends only on i through its dimensions. The assumption that $\boldsymbol{b}_i$ and ε_i are independent means that the random effects and the random errors are supposed to be independent.

An example of Eq. (3) applied to dissolution data is given below where the percentage dissolved of tablet i is measured at four time points:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{pmatrix} = \begin{pmatrix} 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \end{pmatrix} \times \begin{pmatrix} 1 & B_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & B_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & B_i & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & B_i \end{pmatrix} \times \begin{pmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{10} \\ \beta_{11} \\ \beta_{20} \\ \beta_{21} \\ \beta_{30} \\ \beta_{31} \end{pmatrix} + \begin{pmatrix} 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \\ 1 & t & t^2 & t^3 \end{pmatrix} \times \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \\ b_{3i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{pmatrix} \tag{4}$$

$$\boldsymbol{y}_i = \underbrace{\boldsymbol{Z}_i \times \boldsymbol{M}_i}_{\boldsymbol{X}_i} \times \boldsymbol{\beta} + \boldsymbol{Z}_i \times \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$$

An alternative way to represent Eq. 4 is:

$$\begin{aligned} y_{ij} &= (\beta_{00} + \beta_{01}B_i) + (\beta_{10} + \beta_{11}B_i) \times t_{ij} \\ &\quad + (\beta_{20} + \beta_{21}B_i) \times t_{ij}^2 + (\beta_{30} + \beta_{31}B_i) \times t_{ij}^3 + b_{0i} \\ &\quad + b_{1i} \times t_{ij} + b_{2i} \times t_{ij}^2 + b_{3i} \times t_{ij}^3 + \varepsilon_{ij} \end{aligned}$$

or

$$\begin{aligned} y_{ij} &= (\beta_{00} + \beta_{01}B_i + b_{0i}) + (\beta_{10} + \beta_{11}B_i + b_{1i}) \times t_{ij} \\ &\quad + (\beta_{20} + \beta_{21}B_i + b_{2i}) \times t_{ij}^2 \\ &\quad + (\beta_{30} + \beta_{31}B_i + b_{3i}) \times t_{ij}^3 + \varepsilon_{ij} \end{aligned} \tag{5}$$

with: y_{ij} the percentage dissolved at time point j for tablet i; β_{00}, β_{10}, β_{20}, β_{30} the fixed effects parameters for the reference set; β_{01}, β_{11}, β_{21}, β_{31} the differences between the reference and the test batch for the fixed effects; b_{0i}, b_{1i}, b_{2i}, b_{3i} the random effects associated with tablet i; ε_{ij} the residual components or random errors; and B_i the indicator variables (0 if tablet i belongs to the reference and 1 if to the test batch).

In model Eq. (5), an intercept, linear, quadratic and cubic time trend can be distinguished. So in essence a polynomial function is fitted to the data. When the tablets are measured at more than four time points, it is possible in theory to add higher order terms. However, this makes the formula more complicated so that more parameters have to be estimated.

Although models (2) and (5) look different, some corresponding terms can be indicated: $\beta_{00} \sim \mu$, $\beta_{01}B_i \sim \beta_h$, $(b_{0i}, b_{1i}, b_{2i}, b_{3i}) \sim U_{hi}$, $(\beta_{10} \times t, \beta_{20} \times t^2, \beta_{30} \times t^3) \sim \tau_j$, $(\beta_{11}B_i \times t, \beta_{21}B_i \times t^2, \beta_{31}B_i \times t^3) \sim \gamma_{hj}$ and $\varepsilon_{ij} \sim Z_{hij}$. Compared to the ANOVA model described in Section 2.3, model Eq. (5) is more sophisticated and allows a greater refinement of the random effects $(b_{0i}, b_{1i}, b_{2i}, b_{3i})$ and the linear $((\beta_{10} + \beta_{11}\ B_i) \times t)$, quadratic $((\beta_{20} + \beta_{21}\ B_i) \times t^2)$ and cubic $((\beta_{30} + \beta_{31}\ B_i) \times t^3)$ time trend. It shows many advantages when working with unbalanced designs like missing values or data sets measured at different or unequally spaced time points.

One of the main differences between both approaches is the estimation of the model parameters. While the ANOVA-based approach uses the least squares method, the LME models discussed here make use of the more sophisticated maximum likelihood estimation (a basic introduction to this method is given by Kuttatharmakul et al., (2000). In this estimation procedure, a distinction is possible between 'maximum likelihood' (ML) and 'restricted maximum likelihood' (REML). For the estimation of the fixed effects (β parameters), the differences between ML and REML are mostly negligible. For the random effects and the residuals, REML is preferred because it corrects for the loss of degrees of freedom due to the estimation of the fixed effects (Verbeke and Molenberghs, 1997). In a first step, only $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$ are estimated by REML. However, since $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{\Sigma}}_i$ are not independent of $\hat{\boldsymbol{\beta}}$ [see Eq. (6)], the maximization is mostly done in an iterative way using the Newton–Raphson and the expectation–maximization (EM) algorithm. A hybrid approach is recommended, starting with some EM iterations (bring the parameters quickly into the region of the optimum, but converge slowly near the optimum) followed by Newton–Raphson iterations (quite unstable and computationally intensive far from the optimum, but converge quickly near the optimum) (Pinheiro and Bates, 2000). The maximum likelihood estimator of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{y}_i \quad (6)$$

with

$$\hat{\boldsymbol{V}}_i = \boldsymbol{Z}_i \hat{\boldsymbol{D}} \boldsymbol{Z}_i' + \hat{\Sigma}_i$$

$\boldsymbol{V}_i$ is also called the general covariance matrix because $\boldsymbol{y}_i$ has marginal normal distributions with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}_i = \mathrm{Var}(\boldsymbol{y}_i)$. Several structures can be used for the covariance matrices $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$. One usually specifies the unstructured type (UN) which does not assume the covariance matrices to be of any specific form. An overview of some frequently used covariance structures can be found in Verbeke and Molenberghs (1997). The components of $\boldsymbol{b}_i$ can be predicted using the best linear unbiased prediction (BLUP):

$$\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{D}} \boldsymbol{Z}_i' \hat{\boldsymbol{V}}^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}) \quad (7)$$

The goodness of fit of the model can be evaluated using the Aikaike Information Criterion (AIC) or its variant: the Bayesian Information Criterion (BIC), also known as the Schwarz's Bayesian criterion. In S-plus, these criteria are defined as follows (Pinheiro and Bates, 2000):

$$\mathrm{AIC} = -2\mathrm{logLik} + 2n_{\mathrm{par}} \quad (8)$$

$$\mathrm{BIC} = -2\mathrm{logLik} + n_{\mathrm{par}} \log(N) \quad (9)$$

with logLik the logarithm of the likelihood function that is maximized in the ML or REML procedure, n_{par} the number of parameters in the model and N the total number of observations used to fit the model.

Defined this way, the lower the AIC and BIC values, the better. It is interesting to mention that the SAS software gives a somewhat different definition so that there the higher values are deemed the better (Verbeke and Molenberghs, 1997). The advantage of using the AIC or BIC is that these criteria take into account the complexity of the model ($\sim n_{\mathrm{par}}$).

For the comparison of two models, the likelihood ratio test (LRT) is used:

$$\mathrm{LRT} = 2\log\frac{\mathrm{Lik}_2}{\mathrm{Lik}_1} = 2(\log\mathrm{Lik}_2 - \log\mathrm{Lik}_1) \quad \text{for } \mathrm{Lik}_2 > \mathrm{Lik}_1 \quad (10)$$

If n_i is the number of parameters to be estimated in model i, a P-value can be calculated for LRT from a χ^2 distribution with $n_2 - n_1$ degrees of freedom. It is important to keep in mind that to compare models with a different fixed effects structure, estimates based on ML are preferred.

2.5. Working procedure to fit LME models

Since the estimation of all the parameters in model Eq. (5) by the ML or REML process sometimes shows convergence problems, in a first step, the random effects parameters (b_{0i}, b_{1i}, b_{2i} and b_{3i}) are reduced to one (b_{0i}):

$$y_{ij} = (\beta_{00} + \beta_{01}B_i + b_{0i}) + (\beta_{10} + \beta_{11}B_i) \times t_{ij} + (\beta_{20} + \beta_{21}B_i) \times t_{ij}^2 + (\beta_{30} + \beta_{31}B_i) \times t_{ij}^3 + \varepsilon_{ij} \quad (11)$$

Together with the estimation of the parameter values, also their significance is given in the output.

In a second step, the fixed effects part of the model is simplified by omitting the fixed effects parameters with a low significance (for example P-value > 0.05). This is done in a hierarchical way, starting with the highest-order terms. Lower-order terms are kept when a higher-order term is significant and a fixed effect parameter β_{00}, β_{10}, β_{20} or β_{30} is not removed when the corresponding difference indicating parameter β_{01}, β_{11}, β_{21} or β_{31} is significant. Each time a parameter has been deleted, the model must be refitted.

Next the random effects parameters are added successively ($b_{1i} \rightarrow b_{2i} \rightarrow b_{3i}$). The goodness of fit of each model is evaluated and compared with the previous, less complex model using the AIC and LRT. For the reason mentioned in Section 2.4 (comparison of models), it is better to use ML during model refinement and REML for the definitive estimates. When adding the b-parameters one by one, it is possible that at a certain step, no convergence is reached with the standard settings. Although it is sometimes possible to obtain convergence by modifying these standard settings, it can be a time consuming procedure. It is better to adapt the time scale by centering or to use orthogonal polynomial contrasts (an example will be given in Section 4.2). When the fixed and random effects parameters are determined, some diagnostic plots can be constructed like the residuals versus fit, the autocorrelation of the residuals and the response versus fit. According to the assumption $\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, the residuals should be homogeneously spread around zero. This means that a horizontal cluster is obtained around the zero line of the residual versus fit plot. The autocorrelation function of the residuals (the correlation is referred to as *auto*correlation in the context of time-series data) takes values between -1 and 1 such that the correlation is 1 for identical time points ($lag = 0$, where the *lag* unit describes the shift or 'distance' between two successive time points). For $lag > 0$, the autocorrelation should be small. When the spread of the residuals is not homogeneous and autocorrelation for $lag > 0$ is high, a certain structure (autoregressive of order 1 or AR(1)) for the covariance matrix $\boldsymbol{D}$ can be applied (Pinheiro and Bates, 2000).

2.6. Software

For the ANOVA-based methods, SPSS 8.0 for Windows (SPSS, Chicago, IL) was employed in the mode General Linear Model-Repeated Measures. S-plus 2000 (MathSoft, Seattle, WA) was used to study the LME models. Unless otherwise mentioned, the standard settings of the program were used.

3. Data

The methods described above were applied to two data sets. These data are the same as used in a previous paper about the evaluation of dissolution profiles with principal component analysis (Adams et al., 2001).

Data A were obtained from the industry and consisted of a reference batch (Ar) and a test batch (At). Twelve tablets of each batch were measured at four time points (15, 30, 45 and 60 min). Fig. 1 shows the 12 dissolution profiles of Ar, together with four profiles of At. As can be seen, at the first time point measured (15 min), the percentage dissolved is already relatively high. The mean profiles of both batches are shown in Fig. 2. The f_2 factor, calculated for data A, is 83. So, according to the FDA guidelines, these batches can be considered as similar ($f_2 > 50$).

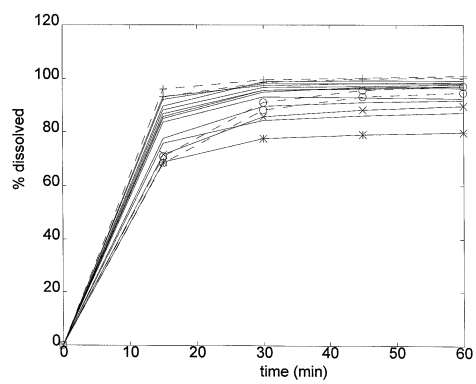

Fig. 1. The 12 dissolution profiles of Ar with dissolution profile 6 indicated with '– × –' and profile 8 with '–*–'. Dissolution profiles 2 and 7 of At are indicated by '--○--' and profiles 11 and 12 of At by '-- + --'.

Data B were obtained from the literature (Tsong and Hammerstrom, 1994) and also consist of a reference (Br) and a test set (Bt). In this case, measurements on 12 tablets of each batch were performed at seven time points: 1, 2, 3, 4, 6, 8 and 10 h. Notice that these time points are unequally spaced. The individual profiles of Br are shown in Fig. 3. It can be seen that the profiles are more gradually increasing compared to data A. Fig. 4 shows the mean profiles of both sets. When the f_2 factor was computed for data B, 64 was obtained. Since this value is higher than the limit of 50, also these two sets are not considered pharmaceutically different.

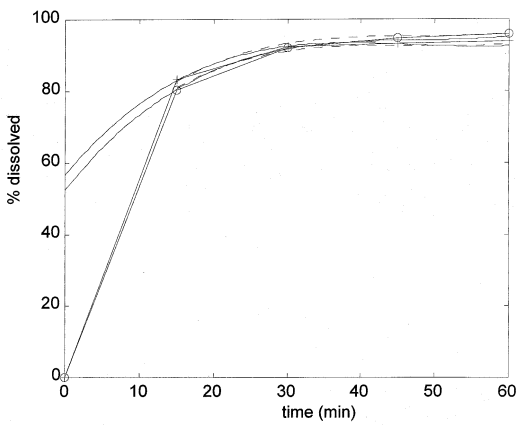

Fig. 2. Mean profiles of both batches of data A (reference '– + –', test '– ○ –'), together with the fitted average profiles using the parameters of Table 2 (—) and Table 3 (----) after time scale centering.

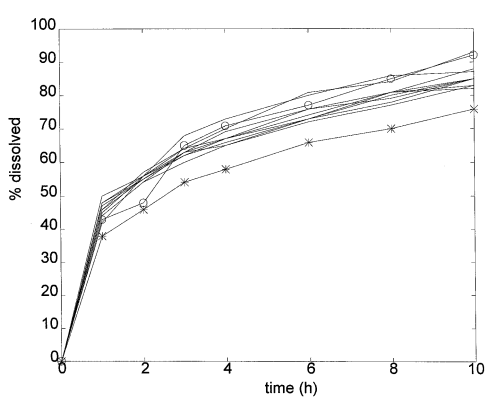

Fig. 3. The 12 dissolution profiles of Br with dissolution profile 2 indicated with '– ○ –' and profile 9 with '–*–'.

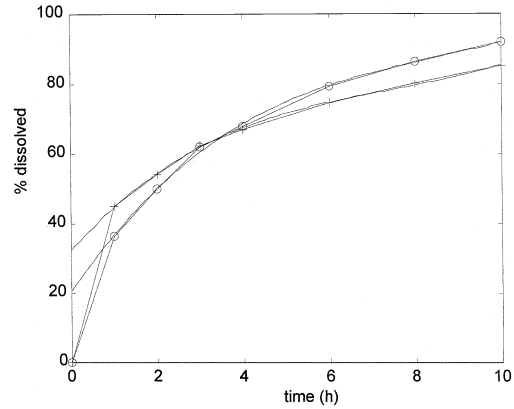

Fig. 4. Mean profiles of both batches of data B (reference '– + –', test '– ○ –'), together with the fitted average profiles using the parameters of Table 4.

For data A as well as data B, it is important to mention that no measurements were performed at time point 0. Consequently, the point (0,0) was not taken into account to fit the different models.

4. Results and discussion

4.1. Univariate repeated measures ANOVA

For data A ($g = 2$, $p = 4$, $N = 24$), the first F-test ($MS_B/MS_{Ta(B)}$) indicated that the general level of the profiles of both batches is similar (significance = 0.974). Concerning the other F-test ($MS_{Tm*B}/MS_{Tm*Ta(B)}$), Mauchly's test indicated that the assumptions about the variance–covariance matrix of the within-subjects were not met (significance = 0.000). Using a correction factor (Greenhouse–Geisser $\varepsilon = 0.351$), a significance of 0.032 was obtained (0.003 without correction). This means that this test indicates that the general shape of the profiles of both batches is different when a level of significance of 5% is considered. Of course, at a lower level (1% or even 0.1%) both sets are considered similar.

In a next step, the logarithm of data A was calculated. This transformation was done in an attempt to make the data more linear. However, the results were similar as for the original data: no

difference in level (significance = 0.987), Mauchly's test still indicated that a correction factor was necessary and the shape (significance = 0.047) was found different at a 5% significance level. So, using the logarithm of the original data does not show any advantage.

When data B ($g = 2$, $p = 7$, $N = 24$) were analyzed, the level of the profiles of both batches was also found similar (significance = 0.480). Mauchly's test (significance = 0.061) showed that the assumptions about the variance–covariance structure are met. Consequently a correction factor is not necessary to evaluate the general dissolution curve shape of both batches. The significance of the second F-test was 0.000 so that the shape of the two sets is considered different.

So, compared to the f_2 factor, where the two sets of data A as well as these of data B are considered pharmaceutically equivalent, the ANOVA-based methods are much more discriminative.

4.2. Linear mixed effects models

4.2.1. Data A

Following the steps of the working procedure (see Section 2.5), the parameters of model Eq. (11) were estimated by ML and the AIC_{ML} amounted to 531.1. β_{31} was found to be not significant (P-value = 0.9398). After removing this parameter, the model was refitted. The AIC_{ML} of this new model was somewhat lower (529.1). The LRT revealed that both models were equivalent ($P = 0.9369$), but the latter is preferred because it is less complex. β_{30} has in the latter model a P-value of 0.0009 and so it is not deleted. Since β_{21} is not significant (P-value = 0.3080), it is omitted and the new model is refitted ($AIC_{ML} = 528.2$). According to the LRT, the latter model is equivalent ($P = 0.2880$) to the previous one and because it contains less parameters, it is preferred. All parameters left in the model now are significant ($P < 0.0001$), except for β_{01} ($P = 0.1184$). However, this parameter is kept in the model because the corresponding parameter in the higher-order term (β_{11}) is significant at the 5% level. So, the fixed effects part of the model reveals that the mean curves of both sets of data A have a significant intercept (β_{00}), linear (β_{10}), quadratic (β_{20}) and cubic (β_{30}) time trend. A significant difference between both batches is found in the linear time trend (β_{11}).

Next, the random effects will be further refined. When b_{1i} was added to the model, the AIC_{ML} decreased to 491.8. According to the LRT, the addition of a random effects parameter in the linear time trend, significantly improves ($P < 0.0001$) the model. When the model was extended with a random effects parameter in the quadratic trend (b_{2i}), it led to no convergence using the standard settings of the LME models in S-plus. So, the equation of the provisional best model obtained for data A can be written as:

$$y_{ij} = (\beta_{00} + \beta_{01}B_i + b_{0i}) + (\beta_{10} + \beta_{11}B_i + b_{1i}) \times t_{ij} + \beta_{20} \times t_{ij}^2 + \beta_{30} \times t_{ij}^3 + \varepsilon_{ij} \quad (12)$$

The ML as well as the REML estimates for the parameters of model Eq. (12) are shown in Table 2. As can be seen, the estimates obtained by ML and REML are the same for the fixed effects parameters, but the standard error is slightly different. For the random effects too, small differences are observed as expected. The fixed effects parameters were used to the fit average profiles of both batches of data A (Fig. 2). One has to keep in mind that (0,0) was not measured and was not used in the calculations. So the calculation of the piece cut on the Y-axis, corresponding to the intercept of Eq. (12), is an extrapolation of the measured data. Since ANOVA and LME models are not well suitable for extrapolations outside the data space, one has to be careful with the statistical interpretation of the calculated intercept. The linear term of Eq. (12) describes the general slope, the quadratic term the curvature and the cubic term a second minimum/maximum like the dip at approximately 55 min. Fig. 5 shows the random effects plot. It can be seen that the profile of tablet 8 of the reference set has a clearly lower intercept than the other curves of that set. The curve of tablet 6 has also a low intercept, but shows a clearly different linear time trend ($\sim$ slope). This can be related to the profiles indicated in Fig. 1. Tablets 2 and 7 of the test batch (14 and 19 in Fig. 5) also show a clearly different linear time profile. The profiles of tablets 11 and 12 of the test set (23 and 24 in Fig. 5) must be rather flat with a high intercept and a small linear time trend. The profiles of these tablets are also plotted

Table 2
ML and REML estimates of the parameters of model Eq. (12) for data A

Parameter	ML estimate (standard error)	P-value (t-test)	REML estimate (standard error)	P-value (t-test)
β_{00}	56.69583 (3.502551)	<0.0001	56.69583 (3.504897)	<0.0001
β_{01}	−4.16667 (3.568575)	0.2555	−4.16667 (3.608899)	0.2607
β_{10}	2.47323 (0.252304)	<0.0001	2.47323 (0.249639)	<0.0001
β_{11}	0.11317 (0.045470)	0.0153	0.11317 (0.045984)	0.0164
β_{20}	−0.05406 (0.007375)	<0.0001	−0.05406 (0.007294)	<0.0001
β_{30}	0.00038 (0.000065)	<0.0001	0.00038 (0.000065)	<0.0001
S.D.(b_{0i})	8.287495		8.664086	
S.D.(b_{1i})	0.099406		0.104225	
S.D.(ε_{ij})	1.402439		1.432602	
AIC	491.8		527.5	

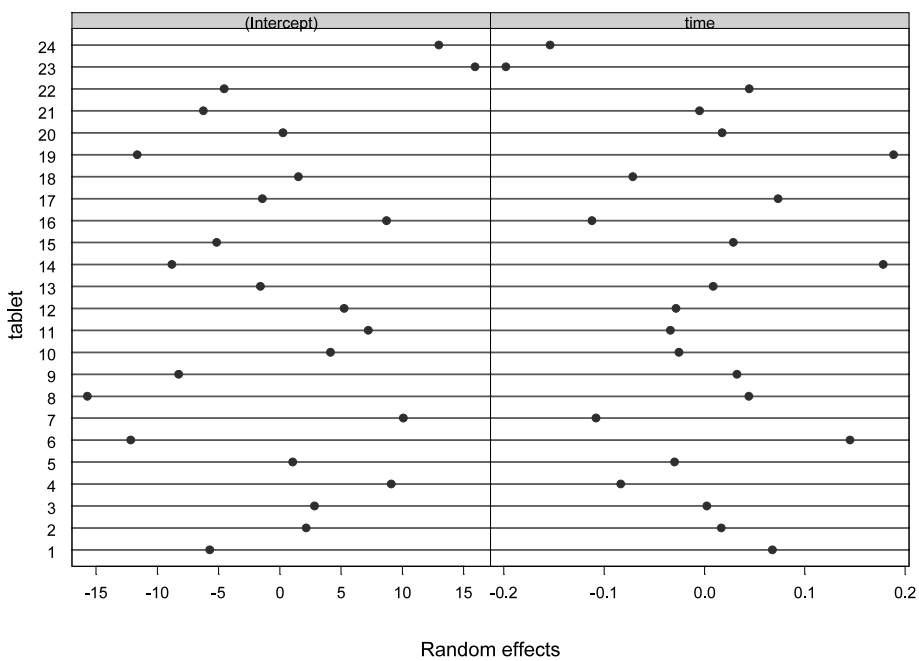

Fig. 5. Random effects plot of model Eq. (12) for the tablets of data A (1–12: reference, 13–24: test set). 'Intercept' is the piece cut on the Y-axis and 'time' represents the linear time trend.

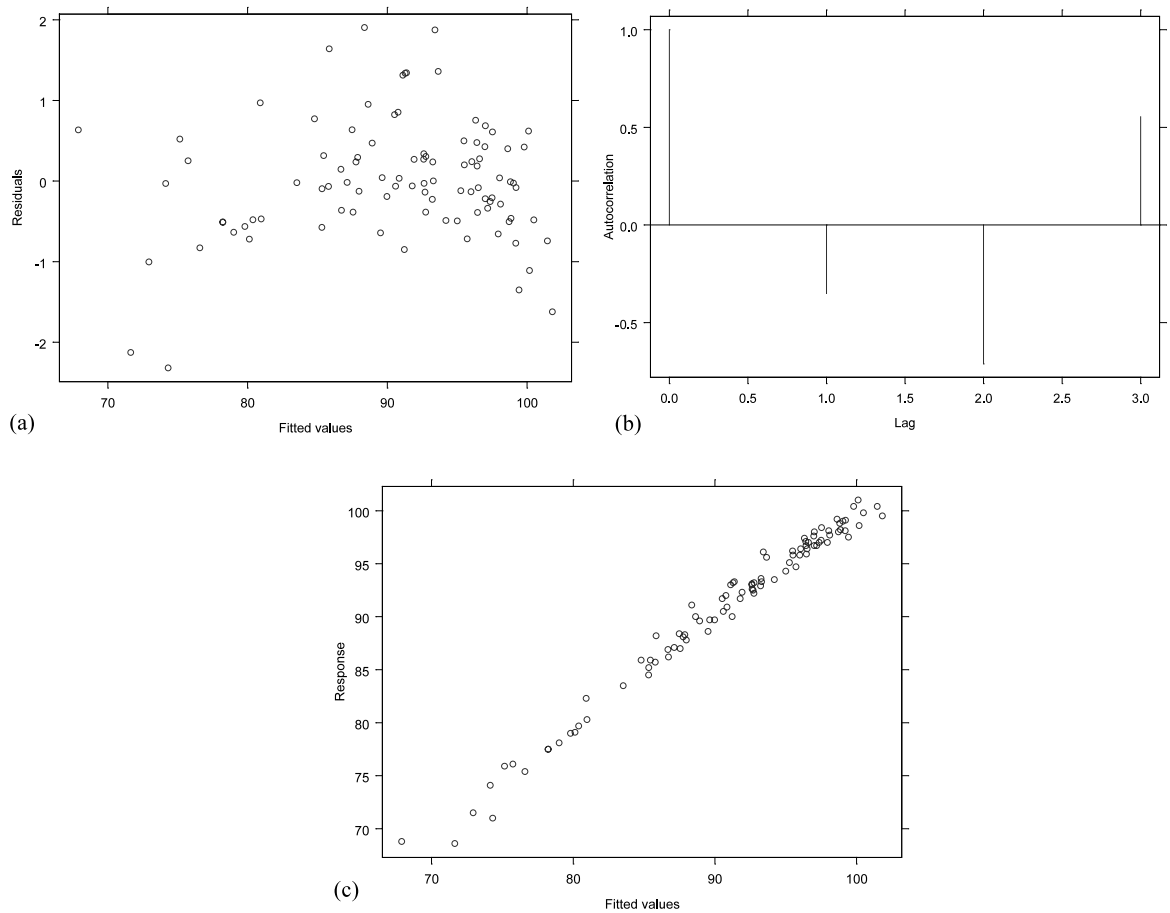

Fig. 6. Diagnostic plots of model Eq. (12) for data A. (a) Residuals versus fitted values; (b) autocorrelation of the residuals; (c) response versus fit.

in Fig. 1 and confirm these findings. Similar conclusions were drawn when the same data were analyzed using principal components (Adams et al., 2001). Fig. 6a shows a plot of the residuals versus fitted values. The residuals are not completely homogeneously spread, which means that there is still some correlation in these values. In other words, not all information is explained by the model. This is also noticed in the autocorrelation plot of the residuals (Fig. 6b). Remember also that Mauchly's test of sphericity (Section 4.1) indicated that the typical assumptions about the variance–covariance matrix of the within-subjects were not met for data A. Fig. 6c shows the response versus fitted values. Although some interesting conclusions can be drawn from model Eq. (12), not all possible information is used. In a next step, the model will be developed further. As described above, it was not possible to include b_{2i} in model Eq. (12) due to convergence problems using the standard settings of the program. Now some variants will be examined to see if better results can be obtained. In a first attempt, the number of iterations of the EM and Newton–Raphson algorithm was adapted. Default 25 EM iterations are performed and the maximum number of optimization iterations (maxIter) is 50, i.e. maximum 25 Newton–Raphson iterations. When EM was increased to 50 and maxIter to 100, convergence was reached when b_{2i} was added to

Eq. (12) ($AIC_{ML} = 418.4$). Even b_{3i} could be added in the ML mode ($AIC_{ML} = 387.2$), but not in the REML mode. When EM was adapted to 15 and maxIter = 100, convergence was also reached for REML ($AIC_{REML} = 430.4$). The addition of b_{2i} and b_{3i} means a greater refinement of the model (significant lower AIC_{ML} values ($P < 0.0001$)) and the plot shown in Fig. 5 can now be extended with a random effects plot for the quadratic and cubic time trend. The conclusions are about the same as discussed for Fig. 5. Only tablet 8 of the reference set could not be identified anymore as somewhat deviant from the other curves of the batch. The residuals are not yet homogeneously distributed and the autocorrelation at *lag* 1 is still too high. However, the main problem is that there are no rules how to adapt EM and maxIter and the trial-and-error approach can take a lot of time.

One of the causes of the convergence problem can be the unit of the time scale. When minutes are used as above, t^2 and t^3 for times like 60 min result in large figures. The corresponding coefficients (β_{20} and β_{30}) on the other hand become very small and can make the calculation unstable. When the time scale was expressed in hours, model Eq. (12) could be extended with b_{2i} without problems for the estimation of the parameters using the standard settings, but no convergence was obtained when b_{3i} was added. Another possibility to adapt the time scale is to center the time points (from each time point, the mean is subtracted so that -22.5, -7.5, 7.5 and 22.5 were used for data A). Using this time scale, it is possible to calculate the parameters of model Eq. (13) without convergence problems:

$$y_{ij} = (\beta_{00} + \beta_{01}B_i + b_{0i}) + (\beta_{10} + \beta_{11}B_i + b_{1i}) \times t_{ij} + (\beta_{20} + b_{2i}) \times t_{ij}^2 + (\beta_{30} + b_{3i}) \times t_{ij}^3 + \varepsilon_{ij} \quad (13)$$

The results for the REML estimation of the parameters are shown in Table 3. Due to the centering of the time axis, the *Y*-axis is now in the middle of the profile. This explains why the intercept is much higher compared to the value of Table 2. It also means that the calculation of the intercept is not based on extrapolation. The more random effects parameters are introduced in the model, the better the profile of each individual

Table 3
REML estimates of the parameters of model Eq. (13) for data A using centering of the time scale or orthogonal polynomial contrasts to estimate the random effects part

Parameter	Time scale centering		Orthogonal polynomial contrasts	
	REML estimate (standard error)	*P*-value (*t*-test)	REML estimate (standard error)	*P*-value (*t*-test)
β_{00}	92.53170 (1.484298)	<0.0001	53.96273 (3.658403)	<0.0001
β_{01}	2.35692 (2.032281)	0.2586	1.29955 (2.281932)	0.5748
β_{10}	0.08152 (0.018956)	0.0001	2.51572 (0.185603)	<0.0001
β_{11}	0.02818 (0.012947)	0.0330	0.02818 (0.012947)	0.0330
β_{20}	−0.01085 (0.000874)	<0.0001	−0.05406 (0.003937)	<0.0001
β_{30}	0.00038 (0.000028)	<0.0001	0.00038 (0.000028)	<0.0001
S.D.(b_{0i})	5.298758		5.796885	
S.D.(b_{1i})	0.085495		19.463823	
S.D.(b_{2i})	0.004267		9.396178	
S.D.(b_{3i})	0.000133		2.918168	
S.D.(ε_{ij})	0.165530		0.188583	
AIC_{REML}	430.4		430.4	

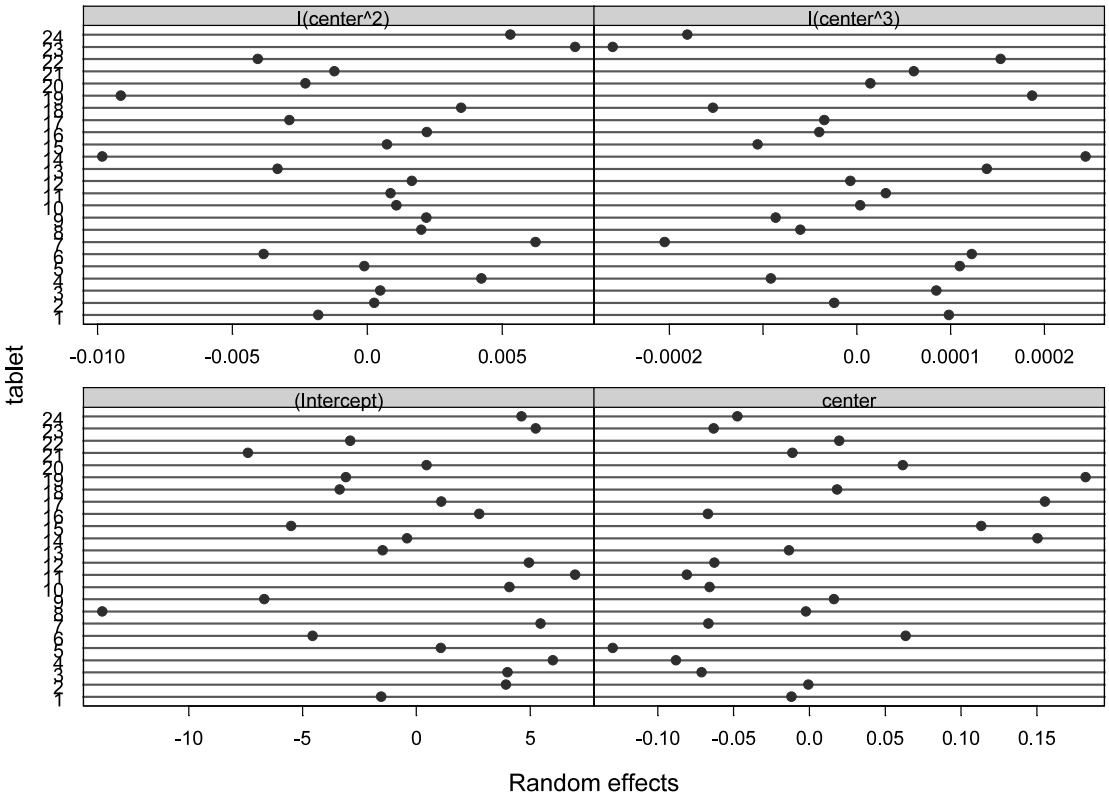

Fig. 7. Random effects plot of model Eq. (13) for the tablets of data A (1–12: reference, 13–24: test set), after centering the time scale. 'Intercept' is the piece cut on the *Y*-axis, 'center' represents the linear, 'center ∧ 2' the quadratic and 'center ∧ 3' the cubic time trend.

tablet is fitted [model Eq. (13) has a lower AIC than model Eq. (12)], but this does not necessarily mean that the average curve is fitted better as shown in Fig. 2. However, the average profile is not necessarily the best representative for a set of curves. This is certainly true when some curves of the set show a cubic trend and others do not. So, it is important to notice that refining the random effects part of the model will also influence the fixed effects part. Fig. 7 shows the random effects plot for model Eq. (13). Tablet 8 of the reference set can still be recognized as having a deviating intercept. Tablet 6 of the reference batch has the largest linear effect parameter of its set, but it is less pronounced than in Fig. 5. The profiles of tablets 2 and 7 of the test batch (14 and 19 in Fig. 7) can best be distinguished in the plot of the random quadratic trend. Test tablets 11 and 12 (23–24 in Fig. 7) also show somewhat deviating fitted parameters. When the range of the linear, quadratic and cubic parameters for the reference (1–12) and the test set (13–24) are compared, it can be concluded that the variability of the test set is larger. This is also in correspondence with the original data and with the conclusions previously found by principal component analysis in combination with the bootstrap technique (Adams et al., 2001). Fig. 8 shows some diagnostic plots: (a) although the residuals are distributed differently compared to Fig. 6, they are still not homogeneously distributed; (b) the autocorrelation at *lag* 1 is still high; (c) the response versus fit plot is clearly better compared to Fig. 6c.

Another method to deal with the time scaling problem is the use of orthogonal polynomial contrasts (Milliken and Johnson, 1984). This function

can easily be applied in S-plus using the poly-(time, 3) command in the formula. Without going into detail about the background of this method, the time points are in fact replaced by levels and made independent of the units used. The condition that the time points must be equally spaced is fulfilled for data A. Using orthogonal polynomial contrasts for the random effects, it is possible to calculate the parameters of model Eq. (13) without convergence problems. The results are also shown in Table 3. Remark that the AIC value is the same as obtained after centering of the time axis. In the random effects plot (Fig. 9), the intercept, quadratic and cubic part resembles Fig. 7, but the linear part is more similar to Fig. 5. The diagnostic plots were the same as in Fig. 8. This means that with none of the methods a homogeneous spread of the residuals nor small autocorrelations could be obtained. Giving the AR(1) structure to the covariance matrix $\boldsymbol{D}$ did not change much and the AIC was even higher.

It is interesting to mention that the same results as for the ANOVA-based methods can be obtained using orthogonal polynomial contrasts for the estimation of the fixed effects parameters of model Eq. (11). The intercept must now be interpreted similarly as described for the centering of the time scale. It indicates the average size of each profile. After estimating the parameters, several F-tests are performed: $\beta_{00}=0$, $\beta_{01}=0$, $\beta_{10}=\beta_{20}=\beta_{30}=0$ and $\beta_{11}=\beta_{21}=\beta_{31}=0$. The respective P-values are: <0.0001, 0.9739, <0.0001 and

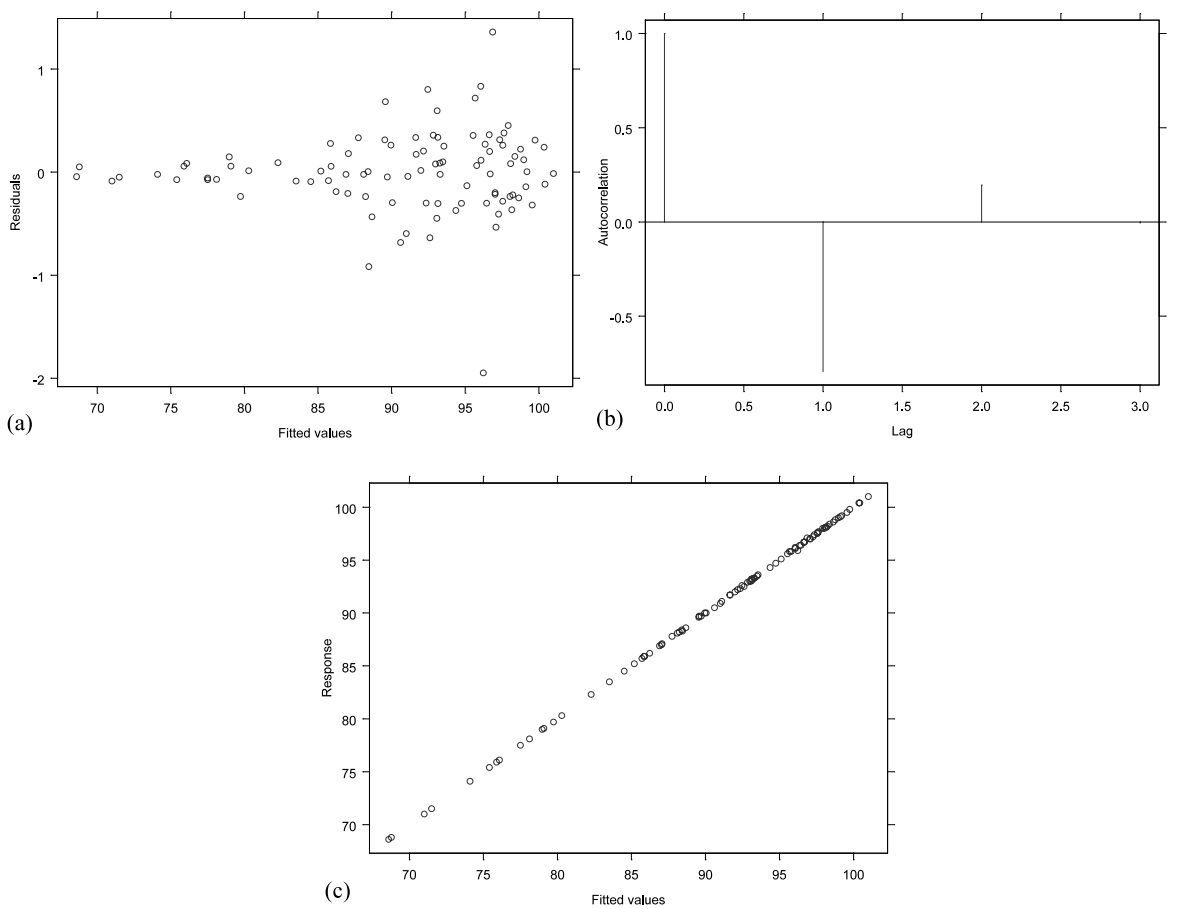

Fig. 8. Diagnostic plots of model Eq. (13) for data A. (a) residuals versus fitted values; (b) autocorrelation of the residuals; (c) response versus fit.

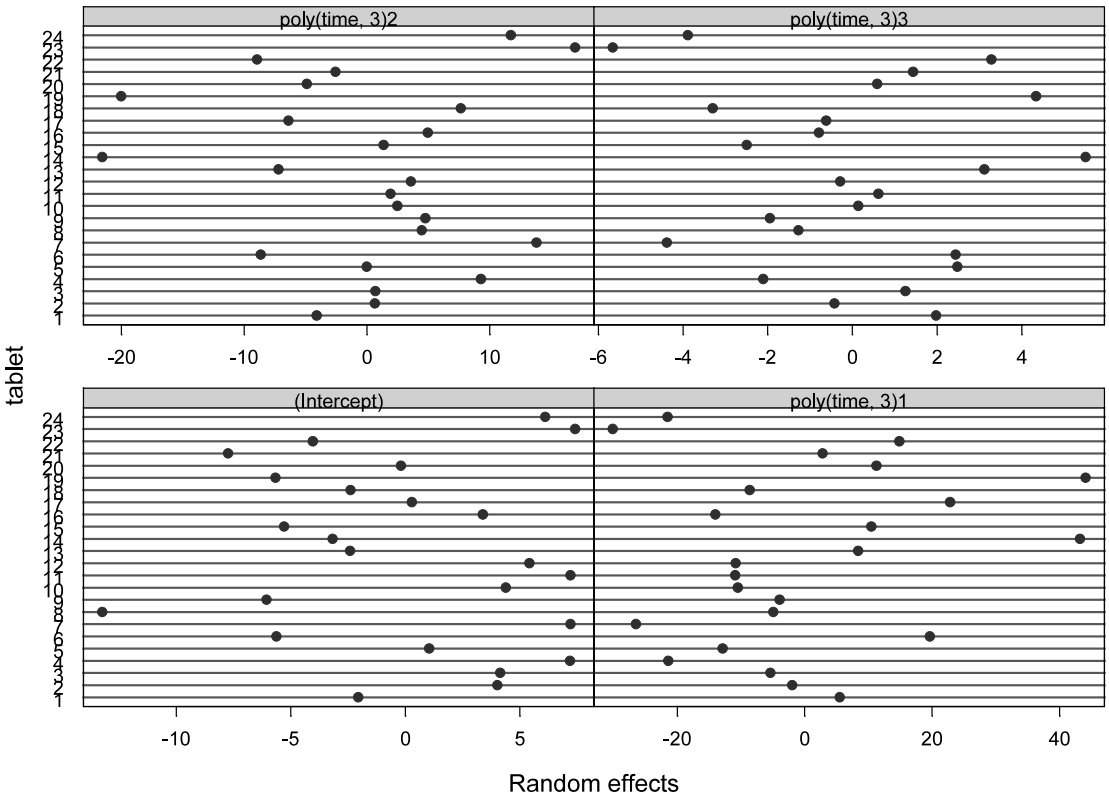

Fig. 9. Random effects plot of model Eq. (13) for the tablets of data A (1–12: reference, 13–24: test set), using orthogonal polynomial contrasts. 'Intercept' is the piece cut on the Y-axis, 'poly(time, 3)1' represents the linear, 'poly(time, 3)2' the quadratic and 'poly(time, 3)3' the cubic time trend.

0.0032. The fact that β_{00} is significantly different from 0 is not very interesting because it just means that there is a significant intercept. The second test, corresponding to $MS_B/MS_{Ta(B)}$ in Table 1, indicates that the average size of the mean profiles of each batch is similar. $\beta_{10} = \beta_{20} = \beta_{30} = 0$ tests whether the tablets dissolve as a function of time and is similar to $MS_{Tm}/MS_{Tm*Ta(B)}$. The last test corresponds to $MS_{Tm*B}/MS_{Tm*Ta(B)}$ and indicates that there is a global difference in time profile between the two batches at a 5% level of significance. So, the P-values found here are exactly the same as those obtained with the ANOVA-based methods (without the Greenhouse–Geisser correction factor) which means that the structure of model Eq. (11) is comparable to that of model Eq. (2). However, the LME models show many advantages: they allow to fit a real curve to the experimental data, the factor time can be examined more accurately since it is split up into its linear, quadratic and cubic trend (also called variance components) and the random effects part can also be defined more precisely by extending model Eq. (11) with b_{1i}, b_{2i} and b_{3i}. It also takes into account the real spacing between the data and can still be used when both batches are measured at different time points and/or contain missing values.

The most important question, however, is whether the two batches are similar or not. The difference between the batches is indicated by the parameters β_{01}, β_{11}, β_{21} and β_{31}. Parameters β_{21} and β_{31} were found not to be significant at the 5% level of significance. Model Eq. (13) corresponds to the model with the lowest AIC. Although time scale centering and orthogonal polynomial con-

trasts do not yield the same values for the parameters, the conclusions are the same. Like for the repeated measures ANOVA, these conclusions depend strongly on the level of significance. At a 5% level, β_{01} is not significant, but β_{11} is. So, both batches differ only in the linear part of the equation ($\sim$ slope). At a 1% level, both parameters are not significant and the two batches are considered similar. Remember that according to the actual FDA guidelines, the reference and test set are considered not to be different ($f_2 = 83$).

4.2.2. Data B

The working procedure described in Section 2.5 was also applied to data B. The AIC_{ML} value for model Eq. (11) amounted to 785.7. The P-value of β_{31} (0.1938) indicates that there is no significant difference between both batches in the cubic time trend. When β_{31} was removed and the parameters were recalculated by ML, the AIC_{ML} was 785.5. Both models were found equivalent by the LRT ($P = 0.1822$) so that the latter is preferred because it contains less parameters to be estimated. Since all parameters in this model are significant ($P < 0.0001$), the fixed effects structure could not be further simplified. The introduction of b_{1i} reduced the AIC_{ML} to 773.4. According to the LRT, this is a significant ($P = 0.0003$) improvement of the model. When the model was enlarged with b_{2i}, the AIC_{ML} increased to 775.7 so that this addition is not necessary to refine the model. So, the best model for data B is:

$$y_{ij} = (\beta_{00} + \beta_{01}B_i + b_{0i}) + (\beta_{10} + \beta_{11}B_i + b_{1i}) \times t_{ij} + (\beta_{20} + \beta_{21}B_i) \times t_{ij}^2 + \beta_{30} \times t_{ij}^3 + \varepsilon_{ij} \quad (14)$$

Table 4
REML estimates of the parameters of model Eq. (14) for data B

Parameter	REML estimate (standard error)	P-value (t-test)
β_{00}	32.32590 (1.071395)	<0.0001
β_{01}	−11.85023 (1.259125)	<0.0001
β_{10}	13.94953 (0.623487)	<0.0001
β_{11}	4.22639 (0.431214)	<0.0001
β_{20}	−1.56643 (0.123673)	<0.0001
β_{21}	−0.24056 (0.036823)	<0.0001
β_{30}	0.07041 (0.007379)	<0.0001
S.D.(b_{0i})	2.120088	
S.D.(b_{1i})	0.297404	
S.D.(ε_{ij})	1.798581	
AIC_{REML}	795.6	

The REML estimates for the parameters are given in Table 4. The fits on the average profiles of the reference and the test set are shown in Fig. 4. Also here one has to be careful with the interpretation of the intercepts. The random effects plot can be found in Fig. 10. Tablet 9 of the reference set has a clearly different intercept. In Fig. 3, it can be seen that the profile of tablet 9 is systematically lower than the other profiles. In this figure, the curve of tablet 2 is also indicated. The curve starts relatively low and ends relatively high. This explains the rather high linear time effect found in Fig. 10. When the variability of the tablets within each of the two sets is evaluated, the reference set has a slightly wider range. All these assessments are similar to those obtained using principal component analysis (Adams et al., 2001). Fig. 11 shows the diagnostic plots. The residuals (Fig. 11a) are homogeneously spread and the autocorrelation of the residuals (Fig. 11b) is low. This is in accordance with the results of Mauchly's test which indicated homogeneity of the variance–covariance matrix for data B. The response versus fitted values plot is shown in Fig. 11c. Next it was checked if the model would still improve when the AR(1) structure was defined for the covariance matrix $\boldsymbol{D}$. Since the AIC increased, this supplementary modeling can better be avoided.

Concerning the question whether both batches are similar or not, it can be seen in Table 4 that β_{01}, β_{11} as well as β_{21} all have P-values < 0.0001. This means that even at a level of significance of 0.01%, the two batches are considered statistically different. For comparison, the similarity factor indicated similarity ($f_2 = 64$).

5. Conclusion

The use of LME models was evaluated for the comparison of dissolution profiles. Besides the 'real' LME models, a repeated measures ANOVA was applied too and the link between both was demonstrated. Two different types of formula-

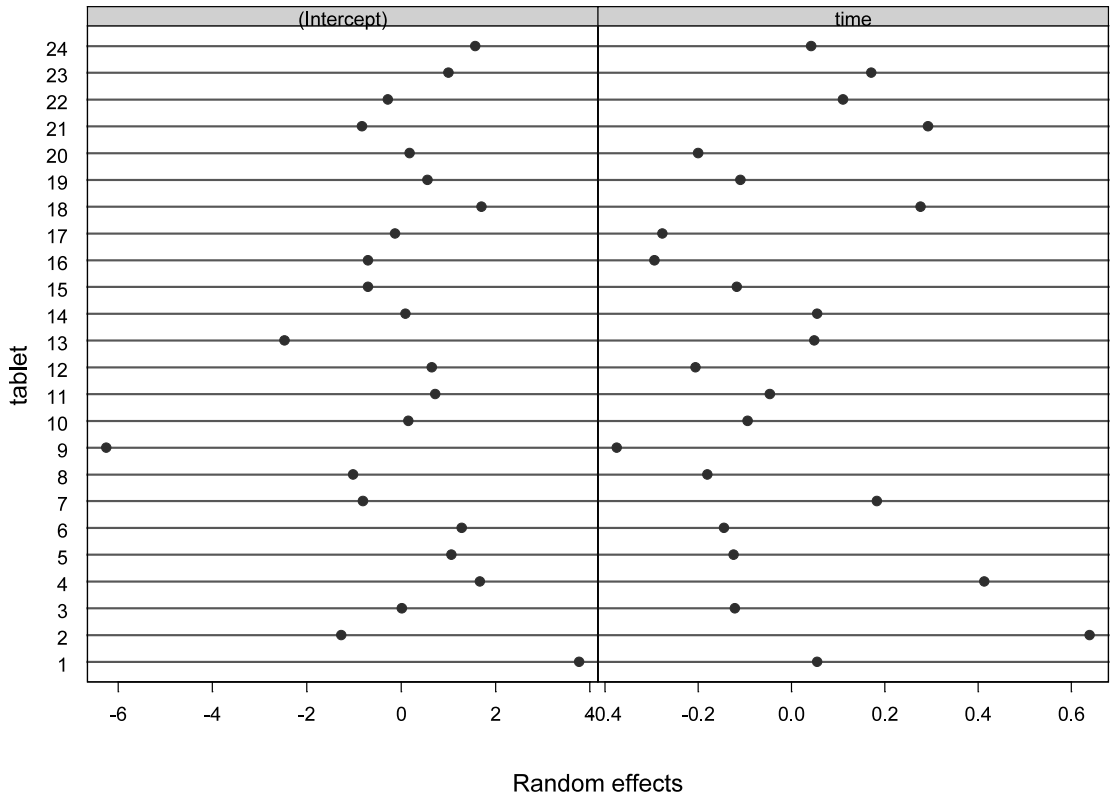

Fig. 10. Random effects plot of model Eq. (14) for the tablets of data B (1–12: reference, 13–24: test set). 'Intercept' is the piece cut on the *Y*-axis and 'time' represents the linear time trend.

tions were examined, each consisting of a reference and a test batch. Although some experience is required, the LME models allow us to analyze the data very accurately and can indicate small differences. Statistically seen, two sets are equivalent when all the difference indicating parameters do not significantly differ from zero (point hypothesis). Even at a level of significance of 0.01%, the statistical limits are much more discriminative than the f_2 factor (actually recommended by the FDA) and indicate differences where the f_2 factor does not. Since the statistical methods do not confirm what specialists would like to see, they are often considered as inconvenient. An improvement of the statistical approach can be the use of an interval hypothesis test instead of a point hypothesis (Massart et al., 1997). The acceptance limit can be defined in cooperation with the pharmaceutical industry.

Another problem using LME models is that the maximum likelihood method to estimate the parameters does not always reach convergence when the standard settings of the software are used. Centering the time axis or introducing orthogonal polynomial contrasts can solve the problem. When the typical assumption about the covariance matrix for the data is not fulfilled, the information left in the residuals can not be modeled in an easy way.

Acknowledgements

This research was financed by a post-doctoral fellowship of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), Brussels.

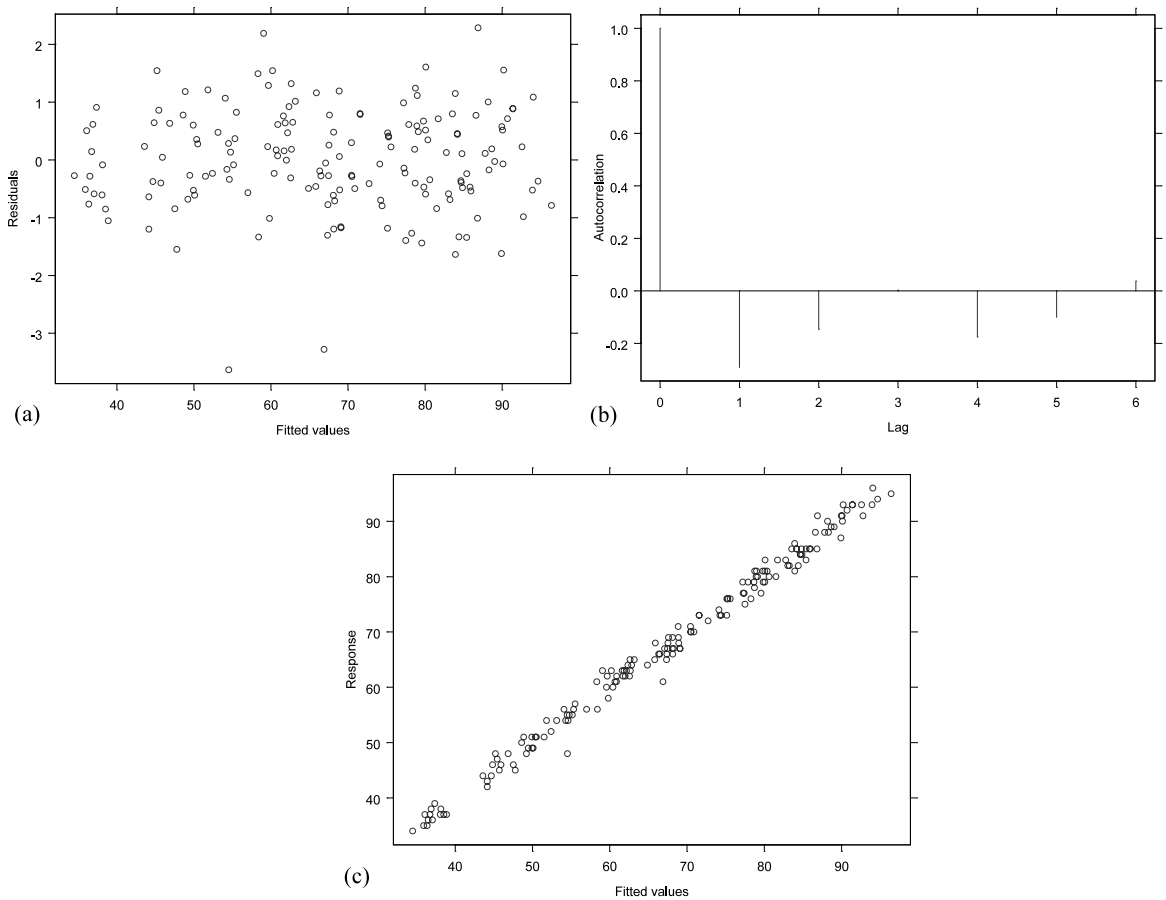

Fig. 11. Diagnostic plots of model Eq. (14) for data B. (a) residuals versus fitted values; (b) autocorrelation of the residuals; (c) response versus fit.

References

Adams, E., De Maesschalck, R., De Spiegeleer, B., Vander Heyden, Y., Smeyers-Verbeke, J., Massart, D.L., 2001. Evaluation of dissolution profiles using principal component analysis. Int. J. Pharm. 212, 41–53.

Anderson, N.H., Bauer, M., Boussac, N., Khan-Malek, R., Munden, P., Sardaro, M., 1998. An evaluation of fit factors and dissolution efficiency for the comparison of in vitro dissolution profiles. J. Pharm. Biomed. Anal. 17, 811–822.

Crowder, M.J., 1996. Keep timing the tablets: statistical analysis of pill dissolution rates. Appl. Stat. 45, 323–334.

FDA Guidance for Industry, 1995. Immediate Release Solid Oral Dosage Forms — Scale-up and Postapproval Changes: Chemistry, Manufacturing and Controls; In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation. Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD, November.

FDA Guidance for Industry, 1997. Dissolution Testing of Immediate Release Solid Oral Dosage Forms. Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD, August.

Kuttatharmakul, S., Smeyers-Verbeke, J., Massart, D.L., Coomans, D., Noack, S., 2000. The mean and standard deviation of data, some of which are below the detection limit: an introduction to maximum likelihood estimation. Trends Anal. Chem. 19, 215–222.

Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J., 1997. Handbook of Chemometrics and Qualimetrics, Part A. Elsevier, Amsterdam.

Mauger, J.W., Chilko, D., Howard, S., 1986. On the analysis of dissolution data. Drug Dev. Ind. Pharm. 12, 969–992.

Milliken, G.A., Johnson, D.E., 1984. Analysis of Messy Data. Designed Experiments, vol. 1. Wadsworth Inc, Belmont, CA.

Moore, J.W., Flanner, H.H., 1996. Mathematical comparison of dissolution profiles. Pharm. Technol. 20, 64–74.

O'Hara, T., Dunne, A., Butler, J., Devane, J., 1998. A review of methods used to compare dissolution profile data. Pharm. Sci. Technol. Today 1, 214–223.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-plus, Statistics and Computing. Springer-Verlag, New York.

Polli, J.E., Rekhi, G.S., Shah, V.P., 1996. Methods to compare dissolution profiles. Drug Inf. J. 30, 1113–1120.

Polli, J.E., Rekhi, G.S., Augsburger, J.J., Shah, V.P., 1997. Methods to compare dissolution profiles and a rationale for wide dissolution specifications for metoprolol tartrate tablets. J. Pharm. Sci. 86, 690–700.

SPSS, 1997. SPSS Advanced Statistics 7.5. SPSS Inc, Chicago, IL, USA.

Tsong, Y., Hammerstrom, T., 1994. Statistical issues in drug quality control based on dissolution testing. Proceedings of the Biopharmaceutical Section of the American Statistical Association, pp. 295–300.

Tsong, Y., Hammerstrom, T.H., Sathe, P.S., Shah, V.P., 1996. Statistical assessment of mean differences between dissolution data sets. Drug Inf. J. 30, 1105–1112.

Verbeke, G., Molenberghs, G., 1997. Linear Mixed Models in Practice, A SAS-Oriented Approach, Lecture Notes in Statistics 126. Springer-Verlag, New York.

Yuksel, N., Kanik, A.E., Baykara, T., 2000. Comparison of in vitro dissolution profiles by ANOVA-based, model-dependent and -independent methods. Int. J. Pharm. 209, 57–67.