Outlier detection in experimental data using a modified Hampel identifier

by

Silvia Selinski and Claudia Becker
Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany

Abstract

The present method allows to detect outlying observations in data which may be described by a deterministic function plus a stochastic component. This type of functional relationship often occurs in experimental data, in toxicological research, for instance. The Hampel identifier, an outlier identification method designed for location-scale models, is modified to account for the special structure of the data. Simulated standardisation values for the procedure are given for sample sizes from 16 to 21.

The procedure is applied to a toxicological study with one of the basic petrochemical compounds ethylene (ethene). This study was designed to determine the individual and population parameters, i. e. the parameters which describe the general behaviour of the investigated process in the whole population, as well as the intra- and interindividual variability of the processes of inhalation, exhalation, and metabolic elimination of the chemical ethylene in male Sprague-Dawley rats.

The results are discussed for various methods determining the functional relationship and for two possible approaches of applying the outlier identification method, one based on the simulated (exact) standardisation values for all sample sizes, the other based on taking a tabled value corresponding to the sample size 'nearest' to the real sample.

KEY WORDS: Outliers, Hampel identifier, toxicokinetics, nonlinear hierarchical models, population parameters, EM algorithm, ethylene

1. Introduction

In many experimental situations, in toxicological research, for instance, we have the situation of observations which differ much from the main part of the data. They seem to be surprisingly higher or lower than one would expect from the rest of the observations and from the 'knowledge' about the underlying processes which generate the data. Such observations are usually called 'outliers' although there exists no formal definition. Of course, it depends on the assumed model if an extreme observation is considered as surprisingly large or small, i. e. to arise from some other distribution than the remaining data.

For cases where the main mass of the data consists of independent observations, identically distributed according to some location-scale model, we have several methods to detect such deviating observations (see Barnett and Lewis, 1994; Gather and Becker, 1997; Hawkins, 1980, for some overviews). Especially in the case of univariate data, there exists a wide variety of such methods for several models. One popular procedure in this context is the so-called Hampel identifier, which is based on two robust measures of location and scale, the median and the median of the absolute deviations from the median (MAD for short), respectively (see Davies and Gather, 1993; Hampel, 1985). Observations too far from the median of the data with respect to their MAD are declared to be outliers. The Hampel identifier is introduced in detail in chapter 2.

Frequently, the data are not assumed to come from such a relatively simple location-scale model, but they are supposed to arise from some deterministic process plus a stochastic component which may contain several sources of variation. The choice of outlier identification methods is less extensive for such more complicated models like regression models or time series (also see Rousseeuw and Leroy, 1987). Assuming that the observations are linked to some parameter vector by a nonlinear function means in terms of outlier detection to compare the observations with the presumed functional relationship. The question is then which observations may be regarded as 'too far away from the main part of the data'. The idea of this approach is to compare the observations with a - perhaps preliminary – model and decide on this basis which observations may be considered as not consistent with the rest of the data. For this purpose a *reference line* is estimated from the model and the deviations of the observations from that line are used for the outlier identifying procedure.

The present method allows to detect outlying observations with respect to the model which is supposed to describe the main part of the data adequately. The Hampel identifier is modified to account for the special structure of the data and applied to a toxicological study with one of the basic petrochemical compounds ethylene (ethene).

This study was designed to determine the individual and population parameters, i. e. the parameters which describe the general behaviour of the investigated process in the whole population, as well as the intra- and interindividual variability of the processes of inhalation, exhalation, and metabolic elimination of the chemical ethylene in male Sprague-Dawley rats. The animals are exposed in a closed inhalation chamber and the decay of ethylene in the atmosphere of the exposition system is observed about 20 times per animal. These experiments were run 5 times for each of the 20 animals with the same concentration in group A (10 rats) and with 5 different doses in group B (10 rats) (for details, see Selinski 2000, 2001). The observed concentrations of ethylene may be described by a nonlinear function f which depends on the time since the application of ethylene into the system and on the kinetic constants which determine the exchange and metabolism of ethylene. Furthermore a stochastic component is assumed which contains the variation of the observations across the concentration-time curve as well as the intra- and the interindividual variation of the parameters which determine the concentration-time curve.

A nonlinear hierarchical model was fitted to the data and the parameters were estimated by the use of an EM algorithm. These estimates are used to construct a reference line for the observations and the modified Hampel identifier was applied to the absolute deviations from the expected values. The results from the usage of different estimates, individual mean, for instance, for the construction of the reference line are compared.

Standardisation values for the Hampel identifier are given for sample sizes from 16 to 21. The identification method is applied once using the exact values from simulations and again using an approximation by a tabled standardisation value for an average sample size. The results are compared for all estimates of the reference line. Approximations of standardisations or critical values are generally used if no tabled values are available for the real sample size.

2. The Hampel identifier

Analysing random data it often occurs that some of the observations differ much from the main part of the data. Such observations are usually called 'outliers' although there does not exist any general formal definition. Nevertheless, it seems to be generally accepted that outliers are observations, which are 'surprisingly far away from the main part of the data' and appear to be 'inconsistent with the rest of the data' (Gather, 1990; Barnett and Lewis, 1994). In the univariate data considered in this article, this means that the observations in question seem to be surprisingly lower or higher than one would expect from the rest of the data. Of course, it depends on the assumed model if an extreme observation is considered as surprisingly large or small.

Potential sources of outliers are (Barnett and Lewis, 1994):

- inherent variability: the natural variation of the observations over the population, unexpected events during the data generating process,
- measurement error: inadequacies in the measurement instrument, rounding of obtained values, mistakes in recording,
- executing error: variability due to the imperfect collection of the data, e.g. choosing a biased sample.

Outliers may influence the analysis of the data and may even falsify the results. They can also be interesting in themselves, since they can hint at unexpected events or unknown relationships. In both cases it is desirable to identify such outlying observations to either exclude them from analysis (or downweight them), or to investigate them further.

2.1. Outlier identification based on outlier regions

A general approach in modelling the occurrence of outliers is to specify a so-called outlier-region and suppose an unknown number k of non-regular observations to lie in this region (Davies and Gather, 1989, 1993).

DEFINITION 2.1: Let X be a univariate random variable with density f. For any $\alpha \in (0,1)$, the α *outlier region* of f is defined by

$$\text{out}(\alpha, f) := \{x \in IR \mid f(x) < \delta(\alpha)\}, \text{ where}$$

$$\delta(\alpha) := \sup_{\delta > 0} \{P(f(X) < \delta) \leq \alpha\}.$$

A number x is called an α *outlier* with respect to f if $x \in \text{out}(\alpha, f)$.

DEFINITION 2.2: Let $\boldsymbol{x}_N = (x_1, ..., x_N)$ be a sample of size N. Suppose that the sample contains $N - k$ regular observations iid with density f whereas the k nonregular observations lie in the outlier region $\text{out}(\alpha_N, f)$. Then $\boldsymbol{x}_N$ is called a sample of size N with a number k of α_N outliers.

The value of α_N can be specified e.g. by choosing $\alpha_N = 1 - (1 - \alpha)^{1/N}$. Hence, for a sample of size N of the target distribution, an observation lies in the outlier region only with probability α (Davies and Gather, 1993).

Neither the number k nor the parameters which specify f, like the expectation μ and the variance σ^2, for instance, are supposed to be known but it is reasonable to assume that $0 \leq k \leq N/2$. The aim is to identify those observations of $\boldsymbol{x}_N$ which lie in the outlier region $\text{out}(\alpha_N, f)$ or, equivalently, to estimate the α_N outlier region using the sample $\boldsymbol{x}_N$ which contains an unknown number k of outliers.

DEFINITION 2.3: Let $\boldsymbol{x_N} = (x_1, ..., x_N)$ be a sample of size N. Let $L(\boldsymbol{x_N}, \alpha_N)$ denote a lower bound and $R(\boldsymbol{x_N}, \alpha_N)$ denote an upper bound for some fixed value $\alpha \in (0,1)$. An outlier identifier is defined by specifying a region

$$OR(\boldsymbol{x}_N, \alpha_N) := (-\infty, L(\boldsymbol{x}_N, \alpha_N)] \cup [R(\boldsymbol{x}_N, \alpha_N), +\infty)$$

with all numbers $x \in OR(\boldsymbol{x_N}, \alpha_N)$ being classified as α_N outliers by the identifier.

The performance of an identifier depends much on how L and R are chosen. Often, we determine these bounds by taking a location statistic m and a scale statistic s and setting $L = m - const\ s$, $R = m + const\ s$. In this case, the robustness of the used location and scale statistics against outliers is essential for the performance of the identifier. In general, it can be stated that using estimators of location and scale with high breakdown point in such identifiers yields procedures with good performance properties like high masking and swamping breakdown points, for example. For details see Becker and Gather, 1999; Davies and Gather, 1993; Gather and Becker, 1997.

The so-called Hampel identifier depends on the robust location and scale statistics median and median absolute deviation (see Hampel, 1985; Davies and Gather, 1993; Gather and Becker, 1997).

DEFINITION 2.4: (Hampel identifier)

Let $\boldsymbol{x_N} = (x_1, ..., x_N)$ be a sample of size N and let $x_{(1)}, ..., x_{(N)}$ be the respective order statistics. The Hampel identifier is defined by identifying all $x \in (-\infty, L(\boldsymbol{x}_N, \alpha_N)] \cup [R(\boldsymbol{x}_N, \alpha_N), +\infty)$ as α_N outliers, where

$L(\boldsymbol{x_N}, \alpha_N) := \text{med}(\boldsymbol{x_N}) - \text{MAD}(\boldsymbol{x_N})\ g(N, \alpha_N)$ and

$R(\boldsymbol{x_N}, \alpha_N) := \text{med}(\boldsymbol{x_N}) + \text{MAD}(\boldsymbol{x_N})\ g(N, \alpha_N)$, with

$$\mathrm{med}(\boldsymbol{x}_N) = \begin{cases} x_{\left(\frac{N+1}{2}\right)} & N\ odd \\ (x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)})/2 & N\ even \end{cases}$$

denoting the median and

$$\mathrm{MAD}(\boldsymbol{x}_N) = \mathrm{med}\left(\left(|x_1 - med(\boldsymbol{x}_N)|, \ldots, |x_N - med(\boldsymbol{x}_N)|\right)\right)$$

denoting the median absolute deviation.

The values of $g(N, \alpha_N)$ may be obtained by requiring

$$P(OR(\boldsymbol{x}_N, \alpha_N) \subset out(\alpha_N, f)) = 1 - \alpha \tag{2.1}$$

or by requiring

$$P(no\ outliers\ identified\ in\ \boldsymbol{x}_N) = 1 - \alpha\,. \tag{2.2}$$

For the case of normal distributions Davies and Gather (1993) provide values of $g(N, \alpha_N)$ for $\alpha = 0.05$ and $N = 20$, 50, and 100 as well as formulas obtained by simulations for $\alpha = 0.05$ and $\alpha = 0.01$ and $N > 10$. Otherwise values of $g(N, \alpha_N)$ have to be simulated.

The Hampel identifier performs well with respect to several criteria like the average proportion of correctly identified outliers, the asymptotic bias, resistance against masking and swamping (Davies and Gather, 1993; Gather and Becker, 1997).

2.2. Outlier identification using a reference line

Trying to identify outliers in data sets as presented here, the time-series structure of the observations has to be taken into account. The observations can be supposed to vary across a theoretical concentration-time curve, which depends on the assumptions on the processes, and circumstances, which generate the data. Moreover, it often occurs that part of the data are systematically over- or underestimated. Hence, the Hampel identifier has to be modified.

DEFINITION 2.5: (Hampel identifier in case of models with deterministic component)
Let $\boldsymbol{y}_N = (y_1, ..., y_N)$ be a set of observations, which may be described by some model M. The expectations of the observations are given by $E(y_n) = h(\theta, t_n)$, $n = 1, ..., N$, according to model M, where h is some linear or nonlinear function, θ is a parameter vector and t_n is the time point on which y_n is observed. Let $\tilde{y}_N = (\tilde{y}_1, \ldots, \tilde{y}_N)$ be the *reference line*, i.e. the estimated set of observations generated according to model M, and let $x_n = |y_n - \tilde{y}_n|$ denote the absolute residuals of the observations with respect to the reference line. The modified Hampel identifier is given by the rule of identifying all

$x = |y - \tilde{y}| \in (-\infty, L(\boldsymbol{x}_N, \alpha_N)] \cup [R(\boldsymbol{x}_N, \alpha_N), +\infty)$ as α_N outliers, where

$$L(\boldsymbol{x}_N, \alpha_N) := \text{med}(\boldsymbol{x}_N) - \text{MAD}(\boldsymbol{x}_N)\, g(N, \alpha_N) \text{ and}$$

$$R(\boldsymbol{x}_N, \alpha_N) := \text{med}(\boldsymbol{x}_N) + \text{MAD}(\boldsymbol{x}_N)\, g(N, \alpha_N).$$

The performance of the Hampel identifier depends much on the fit of the model to the main part of the data. In cases where the fit of a model or the estimation procedure are supposed to be improved by the elimination of outliers, a first estimation may be used as a reference line for detecting outlying observations. After elimination of the α_N outliers the estimation may be repeated and the fit of the model is compared with the result of the first estimation step. In case of many outliers or a repeated non-convergence of a set of observations the procedure could be repeated again using the new estimates to decide if an observation of the complete data set is considered as an outlier or not. This is the same proceeding as in consecutive outlier testing with inward procedures (cf. Barnett and Lewis, 1994, p. 127ff; Hawkins, 1980, p. 63ff). The performance of such procedures depends on the performance of the estimators used to fit the model (see Gather and Becker, 1997, for a discussion). Here, we fit the reference line essentially by means of maximum likelihood (ML) estimates because of their availability in most cases. From location-scale models it is well known that the use of ML estimates within these methods involves the danger of masking, meaning that outliers are not detected because they “mask” each other. For this reason, the use of robust estimators is recommended. The development of an according procedure for the models considered here and the comparison with the currently proposed method is the next step and will be done in future research.

3. Example: Ethylene study

The outlier identifying method presented in the previous chapter was motivated by an inhalation study with one of the basic petrochemical compounds, ethylene (ethene). The study was performed at the *Institute of Occupational Physiology at the University of Dortmund* (*IfADo*) as part of the risk assessment of chemical carcinogens. It aims to determine the population mean kinetic parameters of uptake, elimination, and metabolism of ethylene and to quantify the variability due to interindividual and interoccasion differences.

The following section gives an outline to the data and the modelling approach. For details see Schirm and Selinski (2000), Selinski (2001), and Selinski *et al*. (2000).

3.1 Project

The substance of interest of the present inhalation study is the volatile chemical ethylene (ethene) ($H_2C{=}CH_2$). Ethylene is an important industrial bulk chemical, which is also present in the environment. In mammalian organisms ethylene is partly transformed, by hepatic metabolising enzymes (cytochrome P-450) to ethylene oxide (Filser and Bolt, 1983). Ethylene oxide, also a physiological body constituent (Bolt, 1996, 1998; Bolt *et al*., 1997), is biologically reactive and thereby genotoxic (Kirkovski *et al*., 1998; Filser and Bolt, 1984; Bolt and Filser, 1987; Bolt *et al*., 1984). As previous inhalation experiments with ethylene have indicated the metabolism may be well approximated by first order kinetics at concentrations below 800 ppm (*parts per million)*. This approximation is used in the present study where the maximum concentrations were about 500 ppm ethylene. At higher concentrations the metabolism of ethylene becomes more and more saturated (Bolt and Filser, 1987).

The present inhalation study with ethylene consists of two groups of experiments (group A and B) investigating the inter- and intraindividual behaviour of the processes of uptake, exhalation, and metabolism under equal and under different experimental conditions, respectively.

The experiments were carried out using the 'closed chamber technique' as reviewed by Filser (1992), which allows investigations of kinetics of volatile chemicals *in vivo*. This technique is based on a closed inhalation chamber where during the exposure period the declining atmospheric concentrations of the substance (ethylene) are analytically determined.

In the inhalation chamber, the experimental animals are exposed to the gas or vapour of interest (ethylene) (see figure 1). The exhaled CO_2 is absorbed by soda lime, and its volume is replaced by pure oxygen. At the beginning of each experiment, the test material (ethylene) is injected into the chamber. During the experiment the atmospheric concentration within the chamber is measured by gas chromatography (Bolt *et al.*, 1984). Due to the way of application, the actual concentration in the inhalation chamber at the beginning of each experiment, i. e. at zero time, is not exactly known and must be treated as an additional parameter.

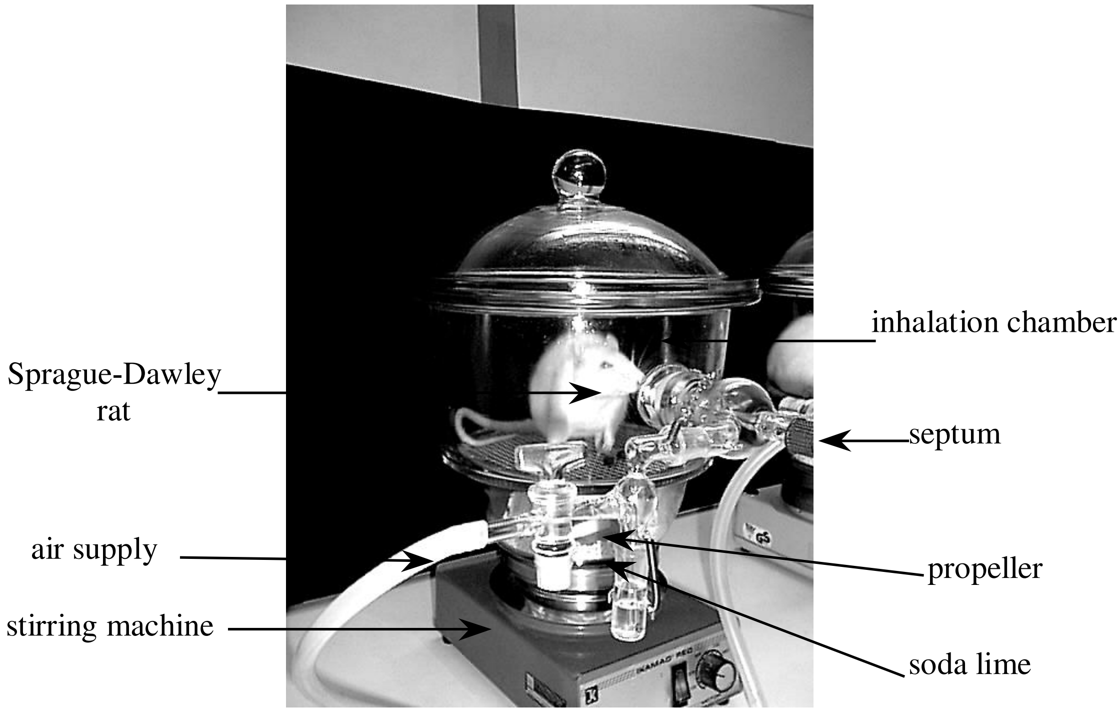

Figure 1: *Experimental design for investigating the kinetics of volatile compounds in vivo according to Filser (1992).*

In the first group of experiments (group A) ten animals were exposed five times each to an initial concentration of about 100 ppm ethylene for a time period of about eight hours. Thus we finally obtained five short time series per animal observed under identical conditions.

The experimental design of the second group (group B) was similar to the first, except for the application of different initial concentrations in the inhalation chamber. Observing another ten rats, we obtained five concentration-time curves per animal at five different initial concentrations of 20 ppm, 50 ppm, 100 ppm, 200 ppm, and 500 ppm ethylene (see Quinke *et al.*, 2000, for further details).

The applied ethylene doses were below the concentration of saturation of ethylene metabolism of about 800 ppm. Hence the data can be analysed approximating the real kinetic processes by first order kinetics using a two-compartment model.

3.2 Two-compartment model

The two-compartment model used by Filser (1992) for the characterisation of exposure to volatile xenobiotics describes uptake, endogenous production, excretion, and the metabolic elimination of the substance. The model is depicted as follows: a xenobiotic gas, in this case ethylene, enters the body and is exhaled. This process is described by introducing two compartments, the first, C_1, representing the environment outside the body, here the inhalation chamber of the exposition system, and the second compartment, C_2, the body itself. The volatile xenobiotic migrates from one compartment to the other through a theoretical interface. During this process, some portion of the xenobiotic within the organism, at any stage, is eliminated by metabolic processes, and another portion is again exhaled (see figure 2).

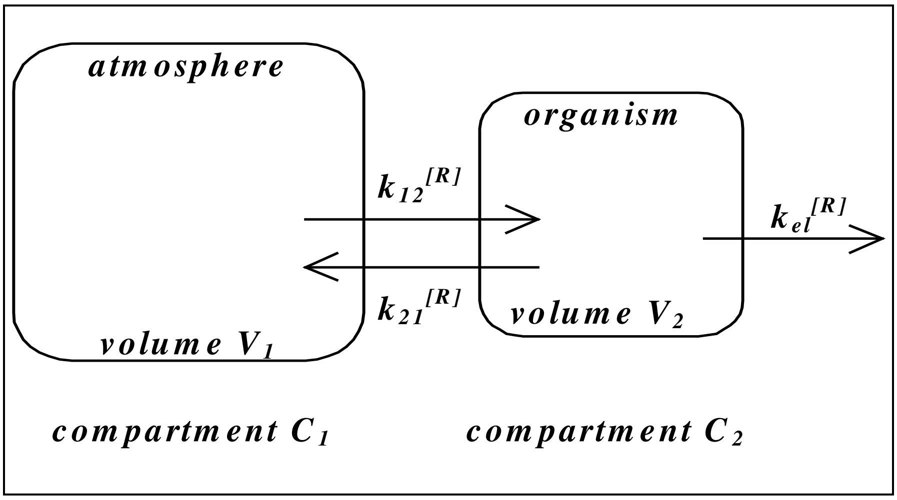

Figure 2. *Two-compartment block model in the case of metabolic turnover*

In case of the present ethylene study the inhalation chamber and its atmosphere form the first compartment where the decay of ethylene is observed. The second compartment is represented by the animal assuming that ethylene is distributed within the whole organism. Ethylene is inhaled, exhaled and metabolised to its reactive metabolite ethylene oxide.

Let $y_l(t)$, l = 1, 2, denote the concentration of a xenobiotic in compartment l at time t and let V_l describe the volume of the compartment. A preliminary assumption is that the compound, in this case ethylene, is metabolised within the body, and that there is no metabolism back to the parent ethylene, the latter being very likely on toxicological grounds.

In the case of overall first order kinetics, each partial process can be characterised by one rate or velocity constant k, that is $k_{12}^{[R]}$ for the uptake, $k_{21}^{[R]}$ for the exhalation, and $k_{el}^{[R]}$ for the metabolic elimination (see figure 2). Thus the concentration of ethylene in the two compartments is given by (Becka *et al.*, 1993; Urfer and Becka, 1996):

$$y_1(t) = y(0) \cdot \left\{ \frac{\left(k_{12}^{[R]} + \lambda_1\right)\exp\{\lambda_2 t\} - \left(k_{21}^{[R]} + \lambda_2\right)\exp\{\lambda_1 t\}}{(\lambda_1 - \lambda_2)} \right\}, \tag{3.1}$$

and

$$y_2(t) = y(0) \cdot \left\{ \frac{\left(k_{12}^{[R]} + \lambda_1\right)\left(k_{12}^{[R]} + \lambda_2\right)}{(\lambda_1 - \lambda_2)V_2 / V_1 k_{21}^{[R]}} \cdot \left[\exp\{\lambda_2 t\} - \exp\{\lambda_1 t\}\right] \right\} \tag{3.2}$$

where $\lambda_{1,2} = \frac{1}{2}\left\{ -\left(k_{12}^{[R]} + k_{21}^{[R]} + k_{el}^{[R]}\right) \pm \sqrt{\left(k_{12}^{[R]} + k_{21}^{[R]} + k_{el}^{[R]}\right)^2 - 4k_{12}^{[R]}k_{el}^{[R]}} \right\}$, $y(0)$ is the initial concentration in compartment 1, V_1 and V_2 are the volumes of distribution of compartment 1 (inhalation chamber) and compartment 2 (organism), respectively.

In the practical application we have to take into account, that the individual organisms have different volumes which are also varying between repeated experimental occasions. In general, the kinetic parameters of the individuals are estimated first and then standardised to eliminate the effect of the volume (i.e., slightly different body weights of the rats). As we use the estimated parameters of the individuals for further calculations, we estimate the standardised kinetic parameters directly (Selinski *et al.*, 2000).

According to Filser (1992) the individual rates of uptake $k_{12}^{[R]}$, exhalation $k_{21}^{[R]}$ and metabolic elimination $k_{el}^{[R]}$ are related to the respective rates k_{12}, k_{21} and k_{el} for a standard rat of 1000 ml by

$$\begin{aligned} k_{12}^{[R]} &= k_{12} \cdot v_2^{2/3}, & & \\ k_{21}^{[R]} &= k_{21} \cdot v_2^{1/3}, & &\text{and} \\ k_{el}^{[R]} &= k_{el} \cdot v_2, & &\text{where} \end{aligned} \tag{3.3}$$

$v_2 = \left(\dfrac{1000}{V_2}\right)$ depends on the actual volume of the organism V_2 and the standard volume 1000 ml.

Substituting the real kinetic parameters in (3.1) and (3.2) yields

$$f(\beta,t) = y_1(t) = y(0) \cdot \left\{ \frac{\left(k_{12}v_2^{2/3} + \lambda_1\right)\exp\{\lambda_2 t\} - \left(k_{21}v_2^{1/3} + \lambda_2\right)\exp\{\lambda_1 t\}}{(\lambda_1 - \lambda_2)} \right\}, \tag{3.4}$$

and

$$y_2(t) = y(0) \cdot \left\{ \frac{\left(k_{12}v_2^{2/3} + \lambda_1\right)\left(k_{12}v_2^{2/3} + \lambda_2\right)}{(\lambda_1 - \lambda_2)\alpha_2 k_{21} v_2^{1/3}} \cdot \left[\exp\{\lambda_2 t\} - \exp\{\lambda_1 t\}\right] \right\}, \tag{3.5}$$

where

$$\lambda_{1ik,2ik} = \frac{1}{2}\left\{ -\left(k_{12ik}v_{2ik}^{2/3} + k_{21ik}v_{2ik}^{1/3} + k_{elik}v_{2ik}\right) \pm \sqrt{\left(k_{12ik}v_{2ik}^{2/3} + k_{21ik}v_{2ik}^{1/3} + k_{elik}v_{2ik}\right)^2 - 4k_{12ik}k_{elik}v_{2ik}^{5/3}} \right\}$$

and $\beta = (k_{12}, k_{21}, k_{el}, y(0))^T$ is the vector of the standardised kinetic parameters and the initial concentration in the first compartment $y(0)$.

3.3 Population models

Population models find a broad application in toxicokinetics or – more general – in pharmacology, where individual experimental outcomes should be pooled together to obtain a set of parameters describing 'in general' the individual behaviour. The individuals are assumed to be a random sample of a population so that their individual sets of parameters and characteristics form a representative data base for the estimation of the parameters and characteristics of the whole population: *population (mean) parameters*. The relationship between observations and parameters is usually nonlinear.

The present approach is referred to as hierarchical Bayes models introduced by Lindley and Smith (1972) for linear and Racine-Poon (1985), Racine-Poon and Smith (1990) for nonlinear hierarchical models. The idea is the following: The observations of each individual are characterised by an individual parameter vector β_i. These parameter vectors β_i vary across a population mean β in the manner of a random sample. The population mean may be known, unknown or there may be some information available. So, the prior information about the parameter vectors is decomposed into several conditional levels of distributions. Estimates are obtained as posterior means of the individual and population parameter vectors.

Four-stage nonlinear hierarchical models are an extension of the classical hierarchical models as proposed by Racine-Poon and Smith (1990) to the situation of repeated measurements where the intraindividual variability has to be considered as in the present inhalation study. Repetition of experiments under equal or different conditions arises in areas such as biomedical and agricultural growth studies, assay development and calibration, pharmacodynamic and pharmacokinetic studies.

This section introduces four-stage models for both experimental situations of the ethylene study: repeated measurements under equal and under different experimental conditions. The estimation is performed using an EM algorithm, which is introduced in section 3.4. Finally the outlier identifying procedure presented in chapter 2 is applied.

Four-stage hierarchical models

We distinguish two cases: first the classical repeated measurement design, corresponding to group A of the ethylene study, secondly two or more sets of observations per individual evaluated under different experimental conditions, group B in case of the ethylene study.

The relationship between the observations and the parameters is the following:

$$y_{ijk} = f(\beta_{ik}, t_j) + \varepsilon_{ijk}, \tag{3.6}$$

where $i = 1, \ldots, I$ denotes the individual,

$j = 1, \ldots, J_{ik}$ is the index of the time point t at which y was observed,

$k = 1, \ldots, K$ denotes the occasion, and

$\beta_{ik} = (\varphi_{ik}^T,\ y_{ik}(0))^T$ is a vector of dimension $p = 4$ with $y_{ik}(0)$ being the initial concentration of the ith individual at exposure k and $\varphi_{ik} = (k_{12ik}, k_{21ik}, k_{elik})^T$ being the vector of the standardised kinetic parameters (see previous section 3.2).

In case of the inhalation study with ethylene we have $I = 10$ individuals in each group and $K = 5$ exposure occasions for all animals but one in group A which died at the end of the fourth exposure due to reasons not related to the experiment. The number of observations J differs from individual to individual and from occasion to occasion. Indices i and k are omitted here just for simplification. The time points t_j are usually the same for all animals and occasions with few exceptions, which were completely recorded.

Our main interest are not the individual responses to the experimental conditions but is focused on a population mean process, which underlies the different individual processes. The individual kinetic parameter vectors φ_{ik} may be regarded as to vary at random across an individual mean parameter vector φ_i, which describes the general behaviour of the respective processes for that individual. Furthermore the individual mean processes are supposed to vary across a population mean process with parameter vector φ in the manner of a random sample. Additionally, we suppose that the variances of the observed concentration-time curves differ from individual to individual and from occasion to occasion.

Estimation of inter- and intra-individual variability in repeated measurement data (group A)

Regarding the experiments of group A of the inhalation study, where each of the ten Sprague-Dawley rats was exposed five times to a concentration of about 100 ppm ethylene, we propose a four-stage nonlinear hierarchical model.

We assume that the observations y_{ijk} of the concentration of ethylene in the atmosphere of the exposition system are independent and have the following distribution:

given β_{ik}, τ_{ik}^2: $\quad y_{ijk} \sim N(f(\beta_{ik}, t_j),\ \tau_{ik}^2)$, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$,

with $\beta_{ik} = (\varphi_{ik}^T,\ y_{ik}(0))^T$, and $\varphi_{ik} = (k_{12ik}, k_{21ik}, k_{elik})^T$,

given β_i, Ω_i: $\quad \beta_{ik} \sim N(\beta_i, \Omega_i)$, $\quad i = 1, \ldots, I$ and $k = 1, \ldots, K$,

with $\beta_i = (\varphi_i^T,\ y_i(0))^T$, and $\varphi_i = (k_{12i}, k_{21i}, k_{eli})^T$,

given β, Σ: $\quad \beta_i \sim N(\beta, \Sigma)$, $\quad i = 1, \ldots, I$,

with $\beta = (\varphi^T,\ y(0))^T$, and $\varphi = (k_{12}, k_{21}, k_{el})^T$,

$$p(\beta) \propto 1 \qquad \forall\ \beta \in IR^4.$$

To obtain Bayes estimates for the population mean and individual parameter vectors β, β_i, and β_{ik} the nonlinear hierarchical model is transformed into a linear one, such as provided by Lindley and Smith (1972). For that purpose the observations y_{ijk} are replaced by an 'almost' sufficient statistic ζ_{ik} with

$$\zeta_{ik} \sim N\,(\,\beta_{ik},\ \tau_{ik}^2\, C_{ik})\,, \qquad i = 1, \ldots, I,\ k = 1, \ldots, K,$$

where $\tau_{ik}^2\, C_{ik}$ is the inverse information matrix:

$$\left(\tau_{ik}^2 C_{ik}\right)^{-1} = E\left[-\frac{\partial^2}{\partial\beta_{ik}\partial\beta_{ik}^T}\ln L\left(y_{111},\ldots,y_{IJK}\left|\beta_{11},\ldots,\beta_{IK},\tau_{11}^2,\ldots,\tau_{IK}^2\right.\right)\right]. \tag{3.7}$$

For example, ζ_{ik} can be chosen as the mean of the posterior density of β_{ik} . In the case of uninformative priors for the variances τ_{ik}^2, the posterior distribution of β_{ik} can be well approximated by its likelihood, so that the maximum likelihood estimate of β_{ik} can be used as a good approximation for ζ_{ik} (Racine-Poon, 1985). For the calculation of the information matrix, see Selinski and Urfer (1998) or Selinski (2001).

To specify f we suppose that our concentration-time curves can be well approximated by first order kinetic processes. Hence, the concentration-time curve in the exposition system is given by

$$f\left(\beta_{ik},t_j\right) = y_{ik}(0)\cdot\left\{\frac{\left(k_{12ik}v_{2ik}^{2/3}+\lambda_{1ik}\right)\exp\left\{\lambda_{2ik}t_j\right\}-\left(k_{21ik}v_{2ik}^{1/3}+\lambda_{2ik}\right)\exp\left\{\lambda_{1ik}t_j\right\}}{\left(\lambda_{1ik}-\lambda_{2ik}\right)}\right\}, \tag{3.8}$$

$i = 1, \ldots, I$, $k = 1, \ldots, K$, where $v_{2ik} = \left(\dfrac{V_{2ik}}{1000}\right)$ depends on the volume of the ith rat at the kth occasion V_{2ik} and

$$\lambda_{1ik,2ik} = \frac{1}{2}\left\{-\left(k_{12ik}v_{2ik}^{2/3}+k_{21ik}v_{2ik}^{1/3}+k_{elik}v_{2ik}\right)\pm\sqrt{\left(k_{12ik}v_{2ik}^{2/3}+k_{21ik}v_{2ik}^{1/3}+k_{elik}v_{2ik}\right)^2-4k_{12ik}k_{elik}v_{2ik}^{5/3}}\right\}$$

with $\lambda_{2ik} < \lambda_{1ik} < 0$.

Thus, we obtain the following linear hierarchical model A.

Linear hierarchical model A

given β_{ik}, τ_{ik}^2: $\quad \zeta_{ik} \sim N(\beta_{ik}, \tau_{ik}^2 C_{ik})$, $\quad i = 1, \ldots, I,\ k = 1, \ldots, K$

given β_i, Ω_i: $\quad \beta_{ik} \sim N(\beta_i, \Omega_i)$, $\quad i = 1, \ldots, I,\ k = 1, \ldots, K$

given β, Σ: $\quad \beta_i \sim N(\beta, \Sigma)$, $\quad i = 1, \ldots, I$

$\quad p(\beta) \propto 1, \quad \forall \beta \in IR^4.$

where $\tau_{ik}^{-2} C_{ik}^{-1}$ is the information matrix as given in (3.7).

Bayes estimates may now be derived from the well known formulas of Lindley and Smith (1972). These estimates are based on the observations and the covariance matrices. The latter are usually unknown and in the present study we are especially interested in the covariance structure of the investigated processes. Thus, the estimation is performed by the use of an EM algorithm which is introduced in the next section.

Estimation in case of different doses (group B)

Analysing the experiments of group B where each of the ten rats was exposed to concentrations of 20, 50, 100, 200, and 500 ppm ethylene it has to be taken into account that the dose varies from occasion to occasion. Thus, the individual and occasion-dependent initial concentration $y_{ik}(0)$ varies across an occasion-dependent mean $y_k(0)$, about 20 ppm for $k = 1$, for instance.

For simplification of the notation the rats 11 to 20 in the ethylene data set (group B) are numbered from $i = 1$ to 10.

Nonlinear hierarchical model

The observations y_{ijk} of the concentration of ethylene in the atmosphere of the exposition system are supposed to be independent and have the following distribution:

given φ_{ik}, $y_{ik}(0), \tau_{ik}^2$: $\quad y_{ijk} \sim N(f(\varphi_{ik}, y_{ik}(0), t_j), \tau_{ik}^2)$, $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K,$

with $\varphi_{ik} = (k_{12ik}, k_{21ik}, k_{elik})^T$

given φ_i, Ω_i: $\quad \varphi_{ik} \sim N(\varphi_i, \Omega_i)$, $\quad i = 1, \ldots, I, k = 1, \ldots, K,$

with $\varphi_i = (k_{12i}, k_{21i}, k_{eli})^T$,

given φ, Σ: $\quad \varphi_i \sim N(\varphi, \Sigma)$, $\quad i = 1, \ldots, I,$

with $\varphi = (k_{12}, k_{21}, k_{el})^T$

$$\mathrm{p}(\varphi) \propto 1, \qquad \forall\ \varphi \in I\!R^3.$$

Linear hierarchical model B

The nonlinear hierarchical model is transformed into a linear one by substituting the observations y_{ijk} by the maximum likelihood estimates ζ_{ik}. Thus, we receive the following linear model B:

given $\boldsymbol{\theta}$, V: $\qquad\qquad \tilde{\zeta} \sim N(\boldsymbol{\theta}, V)$,

where $\tilde{\zeta} = (\tilde{\zeta}_{11}^T, \ldots, \tilde{\zeta}_{IK}^T)^T$ with $\tilde{\zeta}_{ik} = \left(\hat{k}_{12ik}, \hat{k}_{21ik}, \hat{k}_{elik}\right)^T$ being the three first components of the maximum likelihood estimate ζ_{ik} of β_{ik}, $\boldsymbol{\theta} = (\theta_1^T, \ldots, \theta_I^T)^T$, $\theta_i^T = (\varphi_{i1}^T, \ldots, \varphi_{iK}^T)^T$, $V = diag\{(\tau_{1,1}^2 \tilde{C}_{1,1}), \ldots, (\tau_{IK}^2 \tilde{C}_{IK})\}$, and $\tau_{ik}^2 \tilde{C}_{ik}$ denotes the left upper 3×3 matrix of the inverse of the Information matrix $\left(\tau_{ik}^2 C_{ik}\right)^{-1}$.

given ψ, Ω: $\qquad\qquad \boldsymbol{\theta} \sim N(Z_2\psi, \Omega)$,

where , $\boldsymbol{\theta} = (\theta_1^T, \ldots, \theta_I^T)^T$, $\theta_i^T = (\varphi_{i1}^T, \ldots, \varphi_{iK}^T)^T$, $\psi = (\varphi_1^T, \ldots, \varphi_I^T)^T$

$$\Omega = \begin{pmatrix} \Omega_1 & & & & \\ & \ddots & & 0 & \\ & & \Omega_1 & & \\ & & & \Omega_2 & \\ & 0 & & & \ddots & \\ & & & & & \Omega_I \end{pmatrix}, \text{ and } Z_2 = \begin{pmatrix} I & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ I & 0 & & & \\ 0 & I & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \\ 0 & I & 0 & & \vdots \\ \vdots & 0 & \ddots & & \\ & & & \ddots & 0 \\ \vdots & \vdots & & 0 & I \\ & & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & I \end{pmatrix} \text{ is a suitable design matrix.}$$

given φ, Λ: $\qquad\qquad \psi \sim N(Z_3\varphi, \Lambda)$,

where $\varphi = (k_{12}, k_{21}, k_{el})^T$, $\Lambda = diag\{\Sigma, \ldots, \Sigma\}$, and $Z_3 = (I_3, \ldots, I_3)^T$ is a suitable design matrix,

$$p(\varphi) \propto 1, \qquad \forall \varphi \in IR^3.$$

Where estimates of $y_{ik}(0)$ are required, e.g. for the calculation of the residuals, the maximum likelihood estimate $\hat{y}_{ik}(0)$ may be used.

3.4 EM algorithm

In general the EM algorithm as proposed by Dempster *et al.* (1977) aims to estimate the 'missing data' of an *incomplete* data set by an iterative procedure. The term 'missing data' means not only missing values of a random sample but may also refer to unknown parameters, for instance. Estimates of the 'missing data' and of the (hyper)parameters are obtained by computing iteratively the expectation of the posterior density given the observation and the current estimates of the (hyper)parameters (E-step) and the maximum of the posterior density conditional on the observations and the current estimates of the 'missing data' (M-step). The EM algorithm finds broad application, for example to grouped, censored, and truncated data, finite mixture models, iteratively reweighted least squares, factor analysis, estimation of variance components and estimation of hyperparameters. The latter is done here for the evaluation of covariance matrices, individual and population parameters in hierarchical models. The algorithm may be used for both: estimation within a maximum likelihood and within a Bayesian framework.

The EM algorithm in case of four-stage hierarchical models

Fitting hierarchical models by the use of the results of Lindley and Smith (1972) the Bayes estimates require the knowledge of certain hyperparameters such as τ_{ik}^2, Ω_i, and Σ in case of the presented four-stage models. However, we have usually only vague knowledge about these hyperparameters. Furthermore, it is just the aim of the present inhalation study to gain information about them, especially with regard to the interoccasion and interindividual variability. Hence, we estimate both the parameter vectors and the covariance matrices using an EM algorithm as proposed by Dempster *et al.* (1977).

As prior density for the inverse covariance matrices the Wishart distribution is chosen, so that

$\Omega_i^{-1} \sim W_p(\rho_1, R_1)$, $i = 1, \ldots, I$, and

$\Sigma^{-1} \sim W_p(\rho_2, R_2)$,

where $W_p(\rho, R)$ denotes the Wishart distribution with degrees of freedom ρ and matrix R with p denoting the size of the quadratic matrix R. Vague knowledge about the inverse covariance matrices $\Omega_1^{-1}, \ldots, \Omega_I^{-1}$, and Σ^{-1} can be expressed by choosing ρ_1 and ρ_2 as small as possible, i. e., $\rho_1 = \rho_2 = p = 4$ in case of model A. The choice of R_1 and R_2, respectively, seems to have little influence on the estimates (Racine-Poon, 1985).

The unknown variances τ_{ik}^2, $i = 1, \ldots, I$, $k = 1, \ldots, K$, are replaced by their maximum likelihood estimates

$$\hat{\tau}_{ik}^2 = \frac{1}{J} \cdot \sum_{j=1}^{J} \left(y_{ijk} - f\left(\zeta_{ik}, t_j\right)\right)^2 \qquad , i = 1, \ldots, I, k = 1, \ldots, K. \tag{3.9}$$

which may be used to approximate the Bayes estimates due to the equivalence of the posterior mode and the maximum likelihood estimates of τ_{ik}^2 in the special case of our four-stage models.

Model A

With the assumptions and the notation of model A the rth iteration of the EM algorithm is given as follows:

<u>E-step</u>

$$\beta^{(r)} = \left[\sum_{i=1}^{I}\sum_{k=1}^{K}\left(\hat{\tau}_{ik}^2 C_{ik} + \Omega_i^{(r-1)} + \Sigma^{(r-1)}\right)^{-1}\right]^{-1} \cdot \sum_{i=1}^{I}\sum_{k=1}^{K}\left(\hat{\tau}_{ik}^2 C_{ik} + \Omega_i^{(r-1)} + \Sigma^{(r-1)}\right)^{-1} \zeta_{ik} \tag{3.10}$$

$$\beta_i^{(r)} = \left[\left[\sum_{k=1}^{K}\left(\hat{\tau}_{ik}^2 C_{ik} + \Omega_i^{(r-1)}\right)^{-1}\right] + \Sigma^{(r-1)^{-1}}\right]^{-1} \cdot \left[\left(\sum_{k=1}^{K}\left(\hat{\tau}_{ik}^2 C_{ik} + \Omega_i^{(r-1)}\right)^{-1} \cdot \zeta_{ik}\right) + \Sigma^{(r-1)^{-1}} \cdot \beta^{(r)}\right] \tag{3.11}$$

$$\beta_{ik}^{(r)} = \left[\left(\hat{\tau}_{ik}^2 C_{ik}\right)^{-1} + \left(\Omega_i^{(r-1)} + \Sigma^{(r-1)}\right)^{-1}\right]^{-1} \cdot \left[\left(\hat{\tau}_{ik}^2 C_{ik}\right)^{-1} \cdot \zeta_{ik} + \left(\Omega_i^{(r-1)} + \Sigma^{(r-1)}\right)^{-1} \cdot \beta^{(r)}\right]. \tag{3.12}$$

<u>M-step</u>

$$\Omega_i^{(r)} = \frac{R_1^{-1} + \sum_{k=1}^{K}\left(\beta_{ik}^{(r)} - \beta_i^{(r)}\right) \cdot \left(\beta_{ik}^{(r)} - \beta_i^{(r)}\right)^T}{K + \rho_1 - p - 1}, \qquad i = 1, \ldots, I, \text{ and} \tag{3.13}$$

$$\Sigma^{(r)} = \frac{R_2^{-1} + \sum_{i=1}^{I}\left(\beta_i^{(r)} - \beta^{(r)}\right)\left(\beta_i^{(r)} - \beta^{(r)}\right)^T}{I + \rho_2 - p - 1} \tag{3.14}$$

Model B

With the assumptions and the notation of model B the rth iteration of the EM algorithm is given as follows:

<u>E-step</u>

$$\varphi^{(r)} = \left[Z_3^T Z_2^T \left\{\hat{V} + \Omega^{(r-1)} + Z_2 \Lambda^{(r-1)} Z_2^T\right\}^{-1} Z_2 Z_3\right]^{-1} Z_3^T Z_2^T \left\{\hat{V} + \Omega^{(r-1)} + Z_2 \Lambda^{(r-1)} Z_2^T\right\}^{-1} \tilde{\zeta} \tag{3.15}$$

$$\psi^{(r)} = \left[Z_2^T \left(\hat{V} + \Omega^{(r-1)}\right)^{-1} Z_2 + \Lambda^{(r-1)^{-1}}\right]^{-1} \left[Z_2^T \left(\hat{V} + \Omega^{(r-1)}\right)^{-1} \tilde{\zeta} + \Lambda^{(r-1)^{-1}} Z_3 \varphi^{(r)}\right], \tag{3.16}$$

and

$$\theta^{(r)} = \left[\hat{V}^{-1} + \left\{\Omega^{(r-1)} + Z_2 \Lambda^{(r-1)} Z_2^T\right\}^{-1}\right]^{-1} \left[\hat{V}^{-1} \tilde{\zeta} + \left\{\Omega^{(r-1)} + Z_2 \Lambda^{(r-1)} Z_2^T\right\}^{-1} Z_2 Z_3 \varphi^{(r)}\right]. \tag{3.17}$$

<u>M-step</u>

$$\Omega_i^{(r)} = \frac{R_1^{-1} + \sum_{k=1}^{K}\left(\varphi_{ik}^{(r)} - \varphi_i^{(r)}\right) \cdot \left(\varphi_{ik}^{(r)} - \varphi_i^{(r)}\right)^T}{K + \rho_1 - p - 1}, \qquad i = 1, \ldots, I, \text{ and} \tag{3.18}$$

$$\Sigma^{(r)} = \frac{R_2^{-1} + \sum_{i=1}^{I} \left(\varphi_i^{(r)} - \varphi^{(r)}\right)\left(\varphi_i^{(r)} - \varphi^{(r)}\right)^T}{I + \rho_2 - p - 1}. \tag{3.19}$$

Both steps are repeated until $\Omega_1^{(r)}, \ldots, \Omega_I^{(r)}$, and $\Sigma^{(r)}$ converge. Racine-Poon (1985) suggests as criterion for convergence, that the maximum change in the elements of the covariance matrices between successive iterations should be less than 0.001.

Reasonable starting values $\Omega_1^{(0)}, \ldots, \Omega_I^{(0)}$, and $\Sigma^{(0)}$ are given by

$$\Omega_i^{(0)} = \frac{R_1^{-1} + \sum_{k=1}^{K} \left(\zeta_{ik} - \overline{\zeta}_{i.}\right)\left(\zeta_{ik} - \overline{\zeta}_{i.}\right)^T}{K + \rho_2 - p - 2}, \; i = 1, \ldots, I, \text{ and} \tag{3.20}$$

$$\Sigma^{(0)} = \frac{R_2^{-1} + \sum_{i=1}^{I} \left(\overline{\zeta}_{i.} - \overline{\zeta}_{..}\right)\left(\overline{\zeta}_{i.} - \overline{\zeta}_{..}\right)^T}{I + \rho_2 - p - 3} \tag{3.21}$$

in case of model A, where $\overline{\zeta}_{i.} = \frac{1}{K}\sum_{k=1}^{K} \zeta_{ik}$ and $\overline{\zeta}_{..} = \frac{1}{I}\sum_{i=1}^{I} \overline{\zeta}_{i.} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} \zeta_{ik}$.

In case of model B, ζ_{ik} is substituted by the vector of the maximum likelihood estimates of the kinetic parameters $\tilde{\zeta}_{ik}$.

3.5 Application of the modified Hampel Identifier

Toxicokinetic data often contain observations which are not consistent with the general behaviour of the main part of the data and the understanding of the processes involved in the generation of the data. In case of small data sets it is usually possible to identify those observations which differ much from the rest of the data clearly without any 'objective' tool for outlier identification. For larger data sets and cases which are not quite clear such an 'objective' method is necessary to identify outlying observations. For the models considered in the ethylene study, an outlier identification procedure which is suitable for location-scale models, the Hampel identifier, was accordingly modified (see chapter 2).

In case of the present data of the ethylene study the set of observations $\boldsymbol{y}_N$ consists of the measurements of the concentrations in the atmosphere of the inhalation chamber of a single experiment, i.e. $\boldsymbol{y}_{ik} = (y_{i1k}, \ldots, y_{iJk})$ for some $i = 1, \ldots, I$, $k = 1, \ldots, K$. So, every set of observations for each rat and each dosing occasion is analysed separately. The residuals are calculated and possible outliers are identified using the described procedure. As the sample sizes were about $N = 20$, the first approach consists in using the standardisation $g(20, \alpha_{20})$ from Davies and Gather (1993). The standardisation (2.1) and $\alpha = 0.05$ were chosen, hence $g(20, \alpha_{20}) = 5.82$. The second and more expensive approach is to first simulate the standardisation values $g(N, \alpha_N)$ for the exact sample sizes (here ranging from $N = 16$ to $N = 21$) and then to work with these "precise" values. The results of both approaches are compared. For sets of observations where no maximum likelihood estimates ζ_{ik} and individual and occasion dependent estimates β_{ik} are available – due to non-convergence of the maximum likelihood estimation procedure – it is possible to estimate the respective concentration-time curve using the individual mean β_i or φ_i or the population mean β or φ. The initial concentration may be estimated by using

$$f(\varphi, y(0), t) = y(0)g(\varphi, t),$$

with f specified by eq. (3.4) and

$$g(\varphi,t) = \left\{ \frac{\left(k_{12}v_2^{2/3} + \lambda_1\right)\exp\{\lambda_2 t\} - \left(k_{21}v_2^{1/3} + \lambda_2\right)\exp\{\lambda_1 t\}}{(\lambda_1 - \lambda_2)} \right\}.$$

Thus, estimating g by the use of the respective individual or population mean and substituting f by the observations yields

$$\tilde{y}_{ijk}(0) = y_{ijk} \Big/ g\left(\varphi_i^*, t_j\right) \text{ and} \tag{3.22}$$

$$\tilde{y}_{ijk}(0) = y_{ijk} \Big/ g\left(\varphi^*, t_j\right), \text{ respectively,} \tag{3.23}$$

$i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$, where φ_i^* and φ^* are the Bayes estimates of the individual mean kinetic parameter vectors φ_i and of the population mean kinetic parameter vector φ, respectively.

The estimate of $y_{ik}(0)$ may then be obtained as

$$\tilde{y}_{ik}(0) = \frac{1}{J} \sum_{j=1}^{J} \tilde{y}_{ijk}(0). \tag{3.24}$$

Hence, it is possible to include sets of observations where the maximum likelihood estimation was not successful – perhaps due to existing outliers – into the outlier identification procedure. Nevertheless, in cases where this method leads to poorly fitting estimates of the data the performance of the Hampel identifier would be rather bad.

Generally, the present modelling approach consists of three parts:

- maximum likelihood estimation of parameters providing the 'data' for the Bayes estimation
- EM algorithm providing estimates of parameters and hyperparameters
- outlier identification using the previous estimates

which can be repeated to obtain a satisfying fit of the model or perhaps to modify the model after some iterations. It seems to be reasonable to alter the model after the third unsatisfying EM estimation.

Although hierarchical modelling and EM estimation provide a certain protection against outliers (Robert, 1994), the maximum likelihood estimation is certainly affected so that an elimination of outliers using a reference line from a previous estimation procedure could be a useful tool to improve the estimation.

4. Results

In the following, the results of applying the modified Hampel identifier to the ethylene data are discussed. The identification method is applied in two different ways:

first, for all samples under consideration, the standardisation value $g(N, \alpha_N)$ corresponding to the respective sample size N is determined by simulation, and the outlier identification is performed on basis of these exact values. Second, a tabled standardisation value for an 'average' sample size of all considered samples is used as an approximation. This is a usual proceeding if no tabled values are available for the real sample size. The results are then compared to decide whether the difference in the results caused by the approximation is relevant.

Table 1 gives the simulated standardisation values for the Hampel identifier for sample sizes from 16 to 21.

Table 1. *Simulated standardisation values $g(N, \alpha_N)$ for the Hampel identifier according to standardisation (2.1), $\alpha = .05$.*

N	$g(N, \alpha_N)$	N	$g(N, \alpha_N)$
16	6.09	19	5.99
17	6.27	20	5.82
18	6.08	21	5.87

4.1 Identification of outliers in group A

The single data sets for the individuals at the different dosing occasions were analysed with respect to possible outliers using the modified Hampel identifying procedure described in section 2. The residuals were computed using the individual and occasion dependent Bayes estimates β^*_{ik} from model A. Additionally, reference lines were constructed by the use of the maximum likelihood estimates ζ_{ik}, by the estimated individual means β_i^*, and by the estimated population mean β^*. The results are compared with respect to the influence of the estimate on the identification of outliers and the effect of substitution of the chosen individual and occasion dependent estimate by the individual or the population mean.

The modified Hampel identifier was standardised using the simulated values of $g(N, \alpha_N)$, $\alpha = .05$, from table 1. The procedure was repeated, this time approximating the standardisation by $g(20, \alpha_{20}) = 5.82$, $\alpha = .05$, from Davies and Gather (1993) as the number of observations J_{ik} was about 20 for $i = 1, ..., 10$ and $k = 1, ..., 5$. Thus, the effect of the approximation of the standardisation on the identification of outliers was observed.

Table 2 gives the upper and lower bound for the outlier region as specified in definition 2.5 for the data of every single inhalation experiment based on β_{ik}^*.

***Table 2**. Lower and upper bound of the α_N- outlier region of group A determined by the use of the Hampel identifier.*

rat	dose	J_{ik}	$L(x_N, \alpha_N)$	$R(x_N, \alpha_N)$
1	1	19	-0.9112477	1.8469077
	2	20	-5.1319621	8.2092821
	3	21	-6.197436	13.030336
	4	20	-1.3363561	6.4811261
	5	20	-1.0922147	2.3309347
2	1	19	-1.2611003	2.3684803
	2	20	-2.3060923	4.3201523
	3	21	-10.3808912	17.1700712
	4	21	-4.1520851	9.0170251
	5	21	-5.5830083	11.4880083
3	1	19	-2.4239292	3.9384092
	2	20	-11.7724956	22.3103556
	3	21	-1.5865529	3.4108129
	4	21	-0.6704028	1.6123228
	5	19	-0.6088355	4.0807355
4	1	19	-1.629495	2.569495
	2	20	-1.4500249	3.8426249
	3	21	-1.3563067	3.6262667
	4	21	-1.2776793	2.3310793
	5	21	-2.3888117	3.7149317
5	1	19	-0.3040584	0.9916984

	2	20	-1.2766367	2.1069367
	3	21	-1.2405029	2.4419829
	4	20	0.2139601	7.7298499
	5	18	-0.8827382	14.0669482
6	1	21	-1.9182938	3.9369138
	2	20	-4.0483568	7.7282968
	3	21	-2.2980068	3.6147268
	4	21	-1.1012089	2.2736889
	5	21	-1.9999311	3.1366711
7	1	20	-2.4866053	4.8892553
	2	21	-1.2431173	2.8634173
	3	19	-1.3819064	3.1149064
	4	21	-0.9986904	3.4146104
	5	20	-2.6586297	4.2623397
8	1	20	-2.3686071	4.2476271
	2	19	-2.0463594	3.8125794
	3	19	-1.4889307	3.3261907
	4	21	-1.582507	2.809427
	5	20	-4.8864884	8.4601684
9	1	21	-4.3225879	6.4802079
	2	19	-4.2012541	13.8237341
	3	19	-1.7963907	3.5817907
	4	20	-1.881953	3.724453
	5	21	-1.1036918	2.0794918
10	1	21	-2.1951778	3.7093378
	2	21	-1.6900844	4.1930644
	3	21	-3.1143014	5.6345814
	4	16	-1.4873949	2.5028949

Table 3 shows which observations are identified as outliers.

Table 3. *Outliers in group A, time in hours since application of ethylene.*

rat	occasion	time	rat	occasion	time
3	4	7:55	6	4	0:25
4	2	2:55	6	5	0:25
5	1	0:25	7	4	7:15
5	1	0:50	9	3	5:50
5	2	7:30	9	4	8:10
5	2	8:20	10	1	0:25

Note, that only part of the irregular observations would have been detected as 'outliers' at first sight at the data.

As the estimates of the individual and population parameters were satisfying according to the coefficient of determining

$$R^2 = 1 - \sum_{j=1}^{J}\left(y_j - \tilde{y}_j\right)^2 \Big/ \sum_{j}^{J}\left(y_j - \bar{y}_j\right)^2 ,$$

where y_j denote the observations and the $\tilde{y}_j$ are the estimated observations no further elimination and subsequent maximum likelihood and EM estimation was carried out.

Performing the identification procedure also based on the maximum likelihood estimates, the estimated individual and population means yields the following results (table 4) which are illustrated by figures 3 – 6 (see also figures 7 and 8).

Table 4. *Outliers identified using β^*_{ik}, ζ_{ik}, β^*_i, and β^*, respectively, i = 1, . . ., 10, k = 1, . . ., 5, for the calculation of the reference line. Observations identified as outliers are marked by a cross.*

			Identifier based on			
rat	occasion	time (in h)	β^*_{ik}	ζ_{ik}	β^*_i	β^*
1	5	8:20				×
2	4	4:35		×		
2	5	6:15				×
2	5	6:40				×
3	2	5:50			×	
3	2	6:15			×	×
3	2	6:40			×	×
3	2	7:05			×	×
3	2	7:30			×	×
3	2	7:55			×	×
3	2	8:20			×	×
3	3	8:45		×	×	×
3	4	6:40		×		
3	4	7:55	×	×	×	×
3	5	7:30		×	×	×
3	5	8:45		×	×	
4	1	0:50		×		
4	2	2:55	×	×	×	×
4	2	5:50	×	×		
5	1	0:25	×			
5	1	0:50	×			
5	2	7:30	×	×		
5	2	8:20	×	×		
6	1	0:25		×	×	

6	1	4:10		×	×	
6	4	0:25	×	×		
6	5	0:25	×	×		
7	2	0:25			×	
7	4	7:15	×			
7	4	7:55				×
7	4	8:20				×
7	4	8:45				×
8	5	1:15				×
8	5	2:55				×
9	3	5:50	×	×	×	×
9	4	8:10	×			×
10	1	0:25	×	×	×	×
10	4	6:40			×	×

Note, that there are few observation which are identified as outliers at the basis of all estimators for the parameters of the deterministic model. Furthermore, the fit of the respective reference line to the data is often poor in case of the estimated individual mean and population mean where the latter yields often better results. Thus, a careful look at the data, especially when using mean parameters, seems to be indispensable to decide if there is really an 'outlier' or a simple lack of fit of the reference line. Some typical graphs are given below.

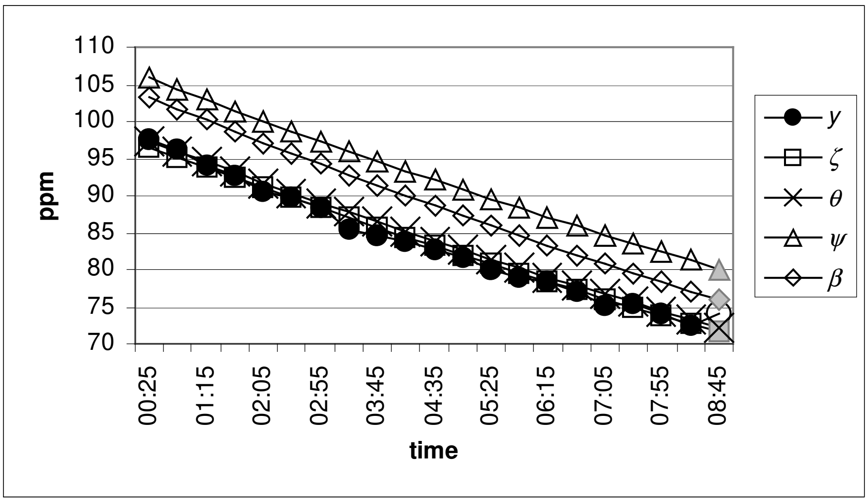

Figure 3. *Observations y and reference lines based on the maximum likelihood estimates* ζ, *the individual and occasion-dependent Bayes estimates* θ, *the estimated individual means* ψ, *and the estimated population mean* β, *respectively, for rat 3,* 3^{rd} *occasion, group A. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

Note, that in figure 3 the last observation was identified as outlier using all estimates but the Bayes estimate of $\theta = (\beta_{I1}, \ldots, \beta_{IK})^T$ although the difference between the estimated concentration-time curves of the maximum likelihood estimate and the Bayes estimate is rather small. The fit of both estimate is quite good $(R^2_{\zeta_{33}} = 0.99, R^2_{\beta_{33}} = 0.98)$. Although the fit of the individual and the population mean is really bad the last observation of this data set is identified plausibly as outlier.

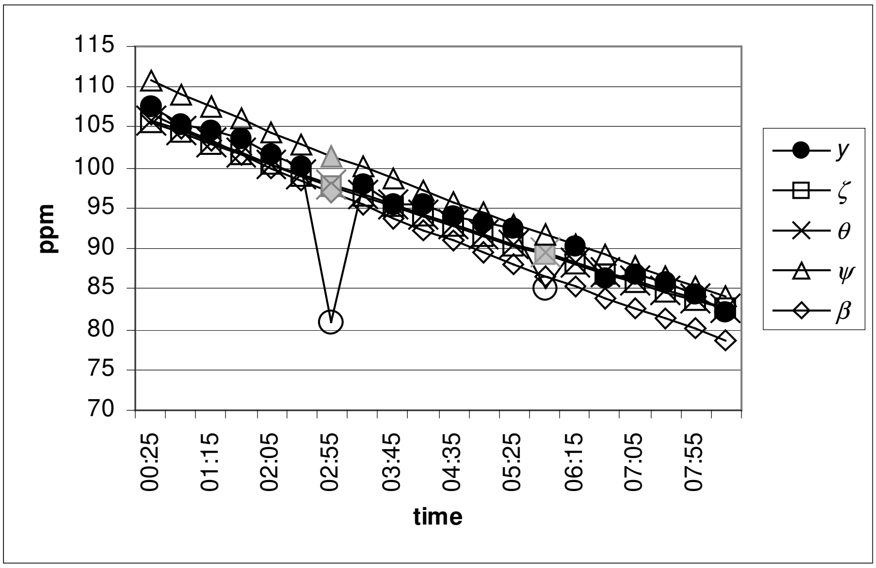

Figure 4. *Observations y and reference lines based on the maximum likelihood estimates* ζ*, the individual and occasion-dependent Bayes estimates* θ*, the estimated individual means* ψ*, and the estimated population mean* β*, respectively, for rat 4, 2*nd *occasion, group A. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

Though the fit of all reference lines is rather bad (R^2 from 0.58 to 0.75), it can be seen in figure 4 that the first outlier at 2:55 *h* since application of ethylene was identified clearly by use of each estimate. Furthermore, the observation at 5:50 *h* was identified as outlier using the maximum likelihood estimate and the individual and occasion-dependent Bayes estimate of β_{42}.

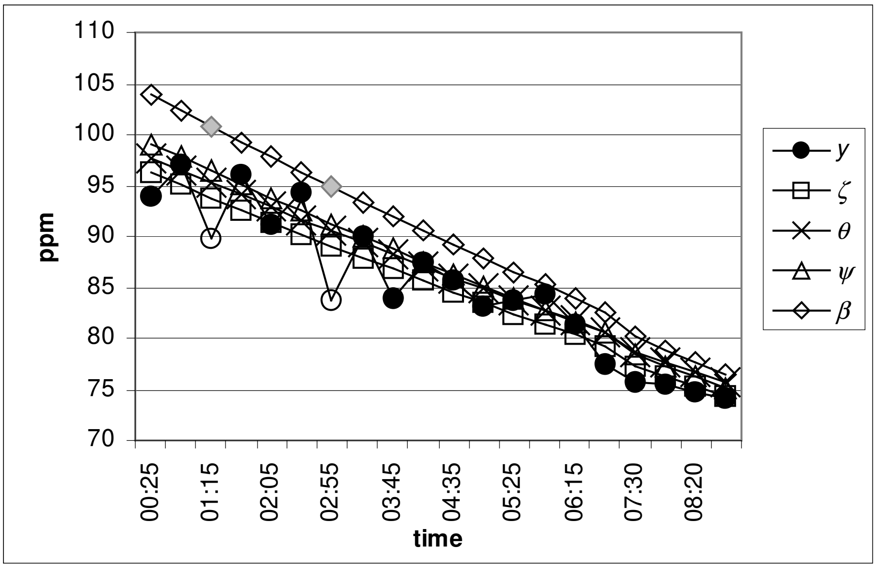

Figure 5. *Observations y and reference lines based on the maximum likelihood estimates* ζ*, the individual and occasion-dependent Bayes estimates* θ*, the estimated individual means* ψ*, and the estimated population mean* β*, respectively, for rat 8, 5*th *occasion, group A. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

Figure 5 shows the dependency of the outlier identification on the model and the fit of the data. Outliers are identified here only by the use of the estimated population mean for the calculation of the reference line ($R^2_\beta = 0.36$, else R^2 from 0.81 to 0.88).

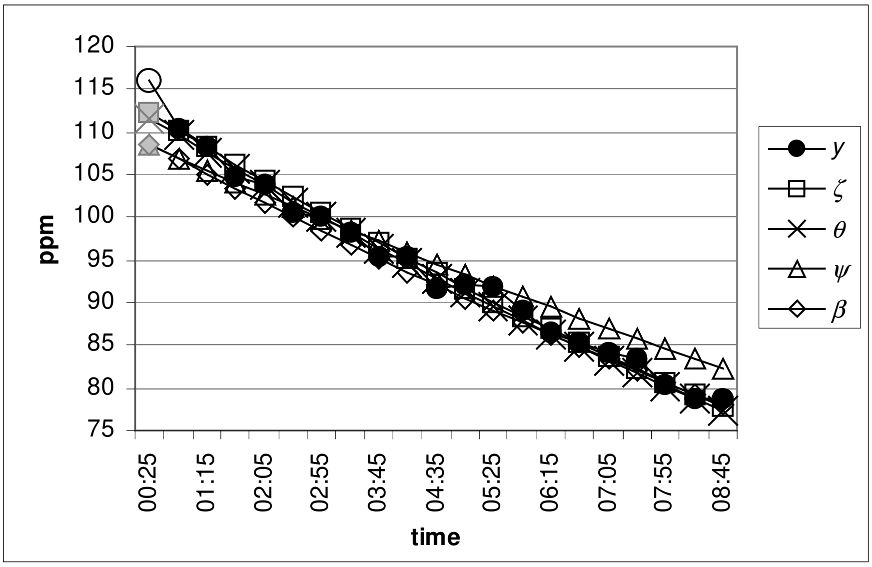

Figure 6. *Observations y and reference lines based on the maximum likelihood estimates* ζ*, the individual and occasion-dependent Bayes estimates* θ*, the estimated individual means* ψ*, and the estimated population mean* β*, respectively, for rat 10,* 1^{st} *occasion, group A. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

In case of good fit, as shown in figure 6 for rat 10 at first occasion where R^2 is between 0.92 and 0.99, even small deviations from the supposed concentration-time curve may be identified as outliers.

Using the approximation of the standardisation $g(20, \alpha_{20}) = 5.82$, $\alpha = .05$, from Davies and Gather (1993) yields slightly different lower and upper bounds of the respective outlier regions but for all estimators of the reference line the observations identified as outliers are exactly the same.

4.2 Identification of outliers in group B

Corresponding to the proceeding for group A the modified Hampel identifier as described in section 2 is used to detect possible outliers in the data of group B. Due to the lack of fit of the Bayes estimates φ_{ik}^{*}, the residuals were calculated using the maximum likelihood estimates ζ_{ik}. In case of the data of rat 11, 2nd and 5th dose, and rat 16, 5th dose, where no maximum likelihood estimates were available, the population mean φ was used instead. The population mean was preferred to the individual mean as the first provided a better fit to the data. The initial concentration was estimated using eq. (3.22) and (3.23) of section 3.5. Thus, it was possible to obtain a reference line for the construction of an α_N outlier region.

Additionally, the residuals were estimated by the use of the Bayes estimates of individual means β_i^{*}, and the population mean β^{*}, respectively. Furthermore, all results were computed for the standardisation $g(N, \alpha_N)$ from table 1 and for the approximation $g(20, \alpha_{20}) = 5.82$ from Davies and Gather (1993), where $\alpha = .05$ in both cases.

Table 5 gives the lower and upper bound for the α_N outlier region, as specified in definition 2.5, for the observations of every single inhalation experiment. The estimation of the residuals was based on the maximum likelihood estimates ζ_{ik} except for rat 1, 2nd and 5th occasion, and rat 6, 5th occasion, where no maximum likelihood estimates were available.

***Table 5**. Lower and upper bound of the α_N- outlier region determined by the use of the Hampel identifier. Estimates based on the population mean are printed bold.*

rat	dose	J_{ik}	$L(x_N. \alpha_N)$	$R(x_N. \alpha_N)$
11	1	21	-0,2841110	0,7407910
	2	21	**-5,1696606**	**10,0360206**
	3	21	-0,5761980	2,0416380
	4	21	-2,4045774	4,4638374
	5	20	**-7,0606730**	**12,0597730**
12	1	21	-0,2110814	0,3566014
	2	21	-1,3074840	2,1111840
	3	21	-1,1123918	2,6879518
	4	20	-1,4619400	3,8284400
	5	20	-7,6526084	15,0642484
13	1	21	-0,6050284	1,2878684
	2	20	-0,7778164	1,4805764
	3	21	-1,2242896	2,1080096

	4	20	-2,5148430	4,0751430
	5	21	-4,8218580	9,6338580
14	1	21	-0,1945366	0,5402966
	2	21	-0,8605570	1,7159570
	3	20	-0,5528605	1,3709405
	4	20	-3,0541951	4,8529151
	5	21	-7,5056920	12,7548920
15	1	21	-0,0167828	0,5365828
	2	21	-1,1792246	1,8999046
	3	21	-0,7239230	2,0306830
	4	21	-1,2661693	3,1001693
	5	21	-9,9015640	14,9172440
16	1	21	-0,8152432	1,3237232
	2	21	-0,8988106	1,8171506
	3	21	-1,8805286	3,4183486
	4	21	-2,9999606	6,9502606
	5	21	**-7,2831032**	**11,6407432**
17	1	21	-0,5519696	1,1390896
	2	21	-0,7586380	1,5099980
	3	21	-1,7376710	3,0702310
	4	21	-2,5484972	5,9540572
	5	21	-10,8637180	18,4097180
18	1	20	-0,2766110	0,7250110
	2	19	-0,6991320	1,5427320
	3	20	-0,7032793	1,5620393
	4	21	-4,0032186	7,5172386
	5	21	-2,9866706	7,1847106
19	1	20	-0,8227933	1,2848033
	2	21	-0,8453054	1,8548254
	3	21	-1,5204196	4,0583996
	4	21	-4,1272774	8,6833574
	5	21	-6,9795800	13,2565600
20	1	21	-0,5181892	0,8734892
	2	21	-0,6511838	1,2995638
	3	21	-1,7231070	3,3327270
	4	21	-3,7732944	7,2022944
	5	21	-2,3303078	5,4311278

The following observations were identified as α_N outliers (table 6).

Table 6. *Outliers in group B, estimation of the concentration-time curve performed by the use of the Bayes estimate of* ζ_{ik} *from model B, time in hours since application of ethylene.*

rat	occasion	time	rat	occasion	time
1	3	0:25	5	1	3:20
1	3	7:05	7	3	3:20
2	1	0:25	8	1	4:10
3	2	8:20	8	3	2:30
3	5	0:25	8	3	7:30
4	1	3:45	8	5	3:20

4	2	0:25	10	1	5:00
4	3	3:20	10	5	0:25
4	3	3:45	10	5	5:25
4	4	6:15	10	5	8:20

Note, that no outliers are detected in the data sets of rat 11, 2nd or 5th dose, or rat 16, 5th dose, where the Marquardt algorithm in PROC NLIN did not converge. The detection of outliers is not related to the performance of φ^*_{ik}.

Using the Bayes estimates of the individual means and of the population mean, respectively, yields quite diverse observations which were classified as outliers (see table 7).

Table 7*. Outliers identified using φ^*_i, φ^*, and ζ_{ik}, respectively, i = 1, . . ., 10, k = 1, . . ., 5, for the calculation of the reference line. Observations identified as outliers are marked by a cross.*

			Identifier based on		
rat	dose	time (in h)	ζ_{ik}	φ^*_i	φ^*
1	3	0:25	×		
1	3	7:05	×		
2	1	0:25	×		
2	2	0:25		×	
2	2	0:50		×	
2	2	1:15		×	
2	3	0:25		×	
2	3	0:50		×	
2	3	1:15		×	
2	4	0:25		×	
2	4	0:50		×	
2	5	0:25			×
3	2	8:20	×		
3	5	0:25	×		×
4	1	0:25		×	
4	1	0:50		×	
4	1	3:45	×		
4	2	0:25	×	×	
4	2	0:50		×	
4	3	0:25		×	
4	3	0:50		×	
4	3	3:20	×		
4	3	3:45	×		

4	4	0:50		×	
4	4	1:15		×	
4	4	1:40		×	
4	4	6:15	×		
4	5	0:25		×	
5	1	3:20	×		
7	3	3:20	×		
8	1	4:10	×		
8	3	2:30	×		
8	3	7:05	×		
8	5	3:20	×		
9	1	0:25		×	×
9	1	0:50		×	
9	1	1:40			×
9	1	2:30			×
9	2	0:25		×	
9	2	0:50		×	
9	2	1:15		×	
9	2	1:40		×	
9	3	0:25		×	
9	3	0:50		×	
9	3	1:15		×	
9	3	1:40		×	
9	3	2:05		×	
9	4	0:25		×	
9	4	0:50		×	
9	4	1:15		×	
9	4	1:40		×	
9	5	0:25		×	
9	5	0:50		×	
9	5	1:15		×	
10	1	4:35	×		
10	5	0:25	×		
10	5	5:25	×		
10	5	8:20	×		

Remarkably, the parallel identification of outliers occurs only seldom for the data sets of group B. This is probably due to the lack of fit of the Bayes estimates. Some typical situations are shown by figures 7 and 8.

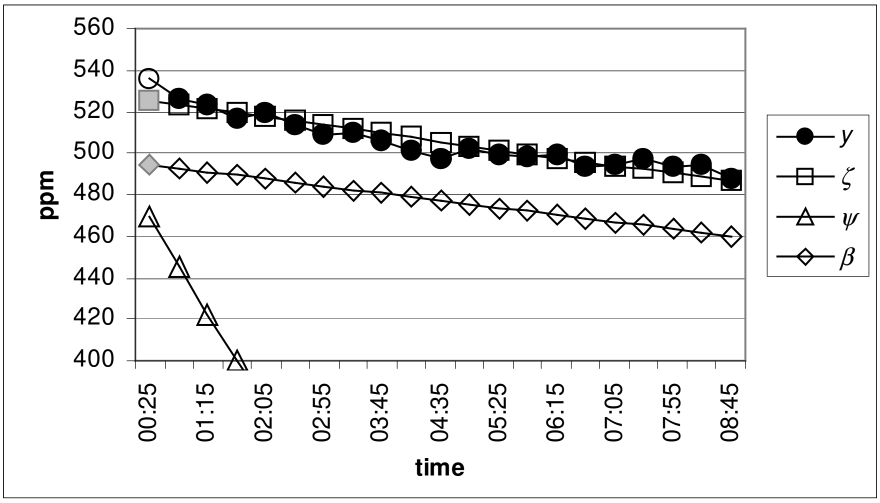

Figure 7. *Observations y and reference lines based on the maximum likelihood estimates* ζ, *the individual and occasion-dependent Bayes estimates* θ, *the estimated individual means* ψ, *and the estimated population mean* β, *respectively, for rat 3, 5*th *dose, group B. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

Although the fit of the estimated population mean was quite bad for rat 3 at 5th dosing occasion (R_β^2 = -4.3) the first observation was identified as outlier using both, the maximum likelihood and the population mean estimate (see figure 7).

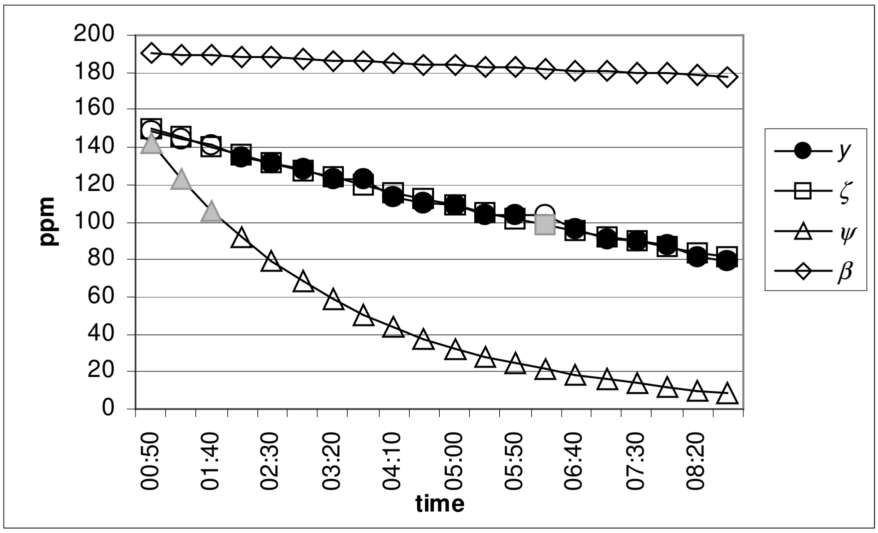

Figure 8. *Observations y and reference lines based on the maximum likelihood estimates ζ, the individual and occasion-dependent Bayes estimates θ, the estimated individual means ψ, and the estimated population mean β, respectively, for rat 4, 4^{th} dose, group B. Observations identified as outliers with respect to at least one of the reference lines are marked by an open circle, the respective reference line is marked grey at that point.*

Figure 8 is an example of a misleading identification of 'outliers' by an inadequate model. Although there is no apparent deviation of the first three observations from the supposed concentration-time curve these observation are identified as outliers by the use of the estimates individual mean β_4 ($R^2_{\beta_4} = -9.0$). A small deviation is identified as outlier by the use of the maximum likelihood estimate ζ_{44}.

Using the approximation of the standardisation $g(20, \alpha_{20}) = 5.82$, $\alpha = .05$, from Davies and Gather (1993) as before yields slightly different lower and upper bounds of the respective outlier region. For all estimators of the reference line but φ_i^* the observations identified as outliers were exactly the same. In case of the individual mean an approximation of the standardisation $g(21, \alpha_{21}) = 5.87$ yields an additional outlier for rat 4, 5^{th} dose, 50 minutes after application of ethylene.

Thus, the approximation of the standardisation seems to have a minor effect on the identifying procedure if the sample size is near enough to the tabled value.

5. Discussion

The present approach provides a flexible tool for outlier detection in data sets which are supposed to be generated by some known processes as it is the case in toxicology, for instance. In case of population models it is further possible to search for outliers without valid estimates for a subset of the data.

Application of this approach to a toxicokinetic study with the chemical ethylene shows the strong dependency of the identified observations from the model. Nevertheless, alternative models (maximum likelihood and Bayes) with similar good fit to the data in terms of R^2 classify almost the same observations as outliers. Of course 'clearly' outlying observations are detected by both identifying rules. However, the suggested identification procedure does not replace the look at the data as the classification may result from observations which differ much or little from the rest of the data as well as from systematic deviations from the model. Hence, we consider the modified Hampel identifier as a powerful screening tool more than as a formal decision rule that separates irregular from regular observations.

The trouble spot of the modified Hampel identifier is the dependency of estimation procedure used for the calculation of the reference line. Maximum likelihood or least squares estimation, for instance, is not robust against outliers. Thus, the irregular observations we wish to detect influence the estimation procedure and as a result the identification rule itself. An alternative would be the application of robust estimators for the parameters like for example M- or S-estimators which presumably have to be adapted to the model situation considered here. Since these estimators are computationally expensive even for simpler models, we expect a much higher computational effort in our case. For the data of the ethylene study already performing the maximum likelihood estimation procedure lasted several weeks which gives a hint on what to expect for robust procedures. Nevertheless, this is the next step and will be done in future research.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. 3rd ed., John Wiley & Sons, Chichester, UK.

Becka, M., Bolt, H. M. and Urfer, W. (1993). 'Statistical evaluation of toxicokinetic data'. *Environmetrics* **4**, 311-322.

Becker, C. and Gather, U. (1999). 'The masking breakdown point of multivariate outlier identification rules'. *Journal of the American Statistical Association* **94**, 947-955.

Bolt, H.M. (1996). 'Quantification of endogenous carcinogens. The ethylene oxide paradox'. *Biochemical Pharmacology* **52**, 1-5.

Bolt, H.M. (1998). 'The Carcinogenic Risk of Ethene (Ethylene)'. *Toxicologic Pathology* **26**, 454-456.

Bolt, H.M. and Filser, J.G. (1987). 'Kinetics and disposition in toxicology. Example: Carcinogenic risk estimate for ethylene'. *Archives of Toxicology* **60**, 73-76.

Bolt, H.M., Filser, J.G. and Störmer, F. (1984). 'Inhalation pharmacokinetics based on gas uptake studies V. Comparative pharmacokinetics of ethylene and 1,3-butadiene in rats'. *Archives of Toxicology* **55**, 213-218.

Bolt, H.M., Leutbecher, M. and Golka, K. (1997). 'A note on the physiological background of the ethylene oxide adduct 7(2-hydroxyethyl)guanine in DNA from human blood'. *Archives of Toxicology* **71**, 719-721.

Davies, L. and Gather, U. (1989). 'The identification of multiple outliers'. *Forschungsbericht Nr. 89/1*, Department of Statistics, University of Dortmund.

Davies, L. and Gather, U. (1993). 'The identification of multiple outliers'. *Journal of the American Statistical Association* **88**, 782-792.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). 'Maximum likelihood from incomplete data via the EM algorithm'. *Journal of the Royal Statistical Society, Series B,* **39**, 1-38.

Filser, J.G. (1992). 'The closed chamber technique – uptake, endogenous production, excretion, steady-state kinetics and rates of metabolism of gases and vapours'. *Archives of Toxicology* **66**, 1-10.

Filser, J.G. and Bolt, H.M. (1983). 'Exhalation of ethylene oxide by rats exposed to ethylene'. *Mutation Research* **120**, 57-60.

Filser, J.G. and Bolt, H.M. (1984). 'Inhalation pharmacokinetics based on gas uptake studies. VI Comparative evaluation of ethylene oxide and butadiene monoxide as exhaled reactive metabolites of ethylene and 1,3-butadiene in rats'. *Archives of Toxicology* **55**, 219-223.

Gather, U. (1990). 'Modelling the occurrence of multiple outliers'. *Allgemeines Statistisches Archiv* **74**, 413-428.

Gather, U. and Becker, C. (1997). 'Outlier identification and robust methods'. In: G.S. Maddala and C.R. Rao (eds.): *Handbook of Statistics, vol. 15*. Elsevier Science B. V., Amsterdam, The Netherlands, 123-143.

Hampel, F.R. (1985). 'The breakdown points of the mean combined with some rejection rules'. *Technometrics* **27**, 95-107.

Hawkins, D.M. (1980). *Identification of Outliers*. Chapman and Hall, London.

Kirkovski, L.I., Lermontov, S.A., Zavorin, S.I., Sukhozhenko, I.I., Zavelsky, V.I., Thier, R. and Bolt, H.M. (1998). 'Hydrolysis of genotoxic methyl-substituted oxiranes: experimental kinetic and semiempirical studies'. *Environmental Toxicology and Chemistry* **17**, 2141-2147.

Lindley, D.V. and Smith, A.F.M. (1972). 'Bayes estimates for the linear model' (with discussion). *Journal of the Royal Statistical Society, Series B,* **34**, 1-42.

Quinke, B., Selinski, S., Golka, K. and Blaszkewicz, M. (2000). 'Population toxicokinetics of ethylene: Calibration and preceding investigations'. *Technical Report 3/2000*, University of Dortmund. [http://www.statistik.uni-dortmund.de/sfb475/sfblit.htm]

Racine-Poon A. (1985). 'A Bayesian Approach to Nonlinear Random Effect Models'. *Biometrics* **41**, 1015-1023.

Racine-Poon A. and Smith A.F.M. (1990). 'Population models'. In: D.A. Berry (ed.): *Statistical Methodology in the Pharmaceutical Science*. Marcel Dekker, New York, USA, 139-162.

Robert, C.P. (1994). *The Bayesian choice: a decision-theoretic motivation*. Springer-Verlag, New York, USA.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Schirm, F.A. and Selinski, S. (2000). 'Interindividual and interoccasion variability of toxicokinetic parameters of uptake, exhalation, and metabolism of ethylene'. *Technical Report 7/2000*, University of Dortmund. [http://www.statistik.uni-dortmund.de/sfb475/sfblit.htm]

Selinski, S. (2000). 'Estimation of toxicokinetic population parameters in a four-stage hierarchical model'. *Technical Report 1/2000*, University of Dortmund. [http://www.statistik.uni-dortmund.de/sfb475/sfblit.htm]

Selinski, S. (2001). *Bayes estimation of toxicokinetic parameters in a four-stage hierarchical model*. Thesis, Department of Statistics, University of Dortmund.

Selinski, S. and Urfer, W. (1998). 'Interindividual and interoccasion variability of toxicokinetic parameters in population models'. *Technical Report 38/1998*, University of Dortmund. [http://www.statistik.uni-dortmund.de/sfb475/sfblit.htm]

Selinski, S., Golka, K., Bolt, H.M. and Urfer, W. (2000). 'Estimation of toxicokinetic parameters in population models for inhalation studies with ethylene'. *Environmetrics* **11**, 479-495.

Urfer, W. and Becka, M. (1996). 'Exploratory and model-based inference in toxicokinetics'. In *Statistics in Toxicology*, ed. B.J.T. Morgan. Oxford University Press, 198-216.