



## Biadditive Mixed Models - Advancing Computational Methods and Applications

Jensen, Sofie Pødenphant

*Publication date:*  
2019

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Jensen, S. P. (2019). *Biadditive Mixed Models - Advancing Computational Methods and Applications*. Technical University of Denmark. DTU Compute PHD-2018 Vol. 487

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Biadditive Mixed Models

- Advancing Computational Methods and Applications

Sofie Pødenphant Jensen

DTU



Kongens Lyngby 2018  
PhD-2018-487

Technical University of Denmark  
Department of Applied Mathematics and Computer Science  
Richard Petersens Plads, building 324,  
2800 Kongens Lyngby, Denmark  
Phone +45 4525 3031  
[compute@compute.dtu.dk](mailto:compute@compute.dtu.dk)  
[www.compute.dtu.dk](http://www.compute.dtu.dk) PhD-2018-487

# Summary (English)

---

Analysis of Variance (ANOVA) is widely used to analyze data in most scientific areas. However, the standard linear ANOVA models are not always adequate to describe the structures in a data set sufficiently. This means that an improved inference and a better insight might be obtained by extending the linear ANOVA models. An example of this, is the multiplicative models also named biadditive models, which arise when one or more multiplicative terms are added to a linear ANOVA model. These models are especially popular within agriculture to analyze genotype-by-environment data, but they are also used in e.g. sensometrics to analyse sensory profile data or in medicine to analyse data from method comparison studies. In general, these models are relevant whenever an interaction between two factors is not completely unstructured, but can be described either fully or partly by a linear regression, where one of the variables in the multiplicative term can be interpreted as the regressor and the other variable as the slope.

In this thesis, the main focus is on a specific version of the multiplicative mixed models, with the multiplicative term being a product of a fixed effect and a random effect, where the fixed effect is a part of the mean structure. This means that this fixed effect enters both the mean and the variance structure, which complicates the estimation of the model. One main goal of this work was to develop a user-friendly open-source software tool to fit this kind of models. For this purpose, R-package `mumm` was created, which is now available on CRAN. The thesis demonstrates how to use the package, which was found to be faster than the commercial alternative. Another aim of this thesis was to investigate the advantages obtained by using the multiplicative mixed model, instead of a simple linear mixed ANOVA model. By simulation studies, it was

demonstrated that the power to detect a significant fixed effect increases by using a multiplicative mixed model instead of a two-way mixed ANOVA model, when the "multiplicative effect" is present in the data.

In sensometrics, a linear approximation to the multiplicative mixed model was presented in [Brockhoff et al. \(2015\)](#). This model is named the Mixed Assessor Model (MAM), and is used to analyze sensory profile data, where a panel of judges assesses different products and scores them in relation to a specific characteristic. In this thesis it was showed that also by using the MAM instead of a standard two-way mixed ANOVA model, an increased power to detect significant product differences can be obtained. The power gain by using the MAM was, however, found to be smaller than when using the multiplicative mixed model. Therefore we have proposed a new F-test, which increases the power of the MAM. In [Brockhoff et al. \(2015\)](#), a method to produce confidence intervals for product differences, based on the MAM, was suggested. This method has been evaluated, and we demonstrated that it works well, resulting in confidence intervals that behave similar to the profile likelihood based confidence intervals estimated by the `mumm` package.

Further, the thesis gives an overview of the different biadditive (mixed) model versions, including a literature review and a description of their applications. Finally, it was demonstrated how these models can be estimated by the R-package TMB (Template Model Builder).

# Summary (Danish)

---

Variationsanalyse (ANOVA) bliver hyppigt anvendt til at analysere data inden for de fleste videnskabelige områder. Standard lineære ANOVA-modeller formår dog ikke altid at beskrive strukturen i data tilstrækkeligt. Det betyder, at man kan opnå forbedret inferens samt en bedre indsigt ved at udvide de lineære ANOVA-modeller. Multiplikative modeller, også kaldet biadditive modeller, er et eksempel på dette, som fremkommer når et eller flere multiplikative termer føjes til en lineær ANOVA-model. Disse modeller er især populære inden for landbrugsvidenskab, hvor de bruges til at analysere genotype-miljø data. De benyttes bl.a. også inden for sensometri til at analysere sensorisk profildata samt inden for medicin til at analysere data fra metodesammenligningsstudier. Generelt er disse modeller relevante i alle situationer, hvor en interaktion mellem to faktorer ikke er fuldstændigt ustruktureret, men i stedet kan beskrives enten helt eller delvist ved en lineær regression, hvor en af de variable fra det multiplikative term kan fortolkes som regressor og den anden variabel som hældningen.

Hovedfokus i denne afhandling er på en specifik version af de multiplikative mixede modeller, hvor det multiplikative term er et produkt af en systematisk effekt (fixed effect) og en tilfældig effekt (random effect), hvoraf den systematiske effekt samtidigt er en del af middelværdistrukturen. Dette betyder, at den systematiske effekt indgår både i middelværdi- og variansstrukturen, hvilket komplicerer estimeringen af modellen. Et hovedformål med dette projekt var derfor at udvikle brugervenlige open-source softwareværktøjer, der kan fitte denne type modeller. Med dette formål, blev R-pakken `mm` implementeret, som nu er tilgængelig på CRAN. Denne afhandling demonstrerer, hvordan man bruger R-pakken, som i øvrigt er hurtigere end det kommercielle alternativ, ifølge de udførte test. Et andet formål med denne afhandling var at undersøge

fordelene ved at benytte den multiplikative mixede model i stedet for en simpel lineær ANOVA-model. Via simulationsstudier demonstrerede vi, at styrken til at finde signifikante systematiske effekter stiger ved at benytte en multiplikativ mixed model i stedet for en lineær tovejs mixed ANOVA-model, når den "multiplikative effekt" er til stede i data.

Inden for sensometri har [Brockhoff et al. \(2015\)](#) præsenteret en lineær approksimation til den multiplikative mixede model. Denne model, som kaldes "the Mixed Assessor Model" (MAM), benyttes til at analysere sensorisk profildata, hvor et dommerpanel bedømmer forskellige produkter og giver dem point på baggrund af specifikke karakteristika. I denne afhandling har vi vist, at man også kan opnå en højere styrke til at finde signifikante systematiske effekter ved at benytte MAM i stedet for en standard tovejs mixed ANOVA-model. Den observerede forbedring i styrke var dog ikke lige så stor, som når den multiplikative model benyttes. Derfor har vi foreslået en ny F-test, som forbedrer styrken for MAM. En metode til at producere konfidensintervaller for produktforskelle, baseret på MAM, er anbefalet i [Brockhoff et al. \(2015\)](#). Denne metode har vi evalueret, og vi demonstrerede, at den fungerer godt og resulterer i konfidensintervaller, som opfører sig som de likelihoodprofil-baserede konfidensintervaller estimeret af `mum`-pakken.

Endvidere giver afhandlingen et overblik over de forskellige biadditive (mixede) modeller. Dette inkluderer et litteraturstudie samt en beskrivelse af modellernes anvendelse. Slutteligt har vi demonstreret, hvordan disse modeller kan estimeres via R-pakken TMB (Template Model Builder).

# Preface

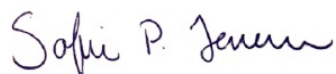
---

This thesis was prepared at the Section of Statistics and Data Analysis at the Department of Applied Mathematics and Computer Science (DTU Compute) at the Technical University of Denmark (DTU) in partial fulfilment of the requirements for acquiring a Ph.D. degree in Applied Statistics. The project was supervised by professor Per Bruun Brockhoff and co-supervised by associate professor Murat Kulahci from DTU Compute. It was further supervised by senior researcher Kasper Kristensen from DTU Aqua. The project was fully funded by DTU.

The thesis deals with multiplicative mixed models in sensometrics, agriculture and medicine. The main focus is on reviewing multiplicative models and their use in science and on developing software tools for user-friendly and time-efficient estimation of the models.

The thesis is based on three research papers and an R-package, documented by its reference manual, written and developed during the period 2015-2018.

Lyngby, 31-August-2018



Sofie Pødenphant Jensen





# List of contributions

---

## Journal papers

- [A] Pødenphant, S., Kristensen, K., & Brockhoff, P. B. (2018). *The multiplicative mixed model with the mumm r-package as a general and easy random interaction model tool*. Submitted to *The American Statistician* (under review).
- [B] Pødenphant, S., Truong, M. H., Kristensen, K., & Brockhoff, P. B. (2018). *The Mixed Assessor Model and the multiplicative mixed model (MAM and mumm)*. Submitted to *Food Quality and Preference* (under review).
- [C] Pødenphant, S., Hadasch, S., Brockhoff, P. B., Piepho, H.-P. (2018). *Biadditive Models and the Template Model Builder - a model overview and estimation methods* (Draft intended for a statistics journal).

## Software

The mumm R-package:

<https://CRAN.R-project.org/package=mumm>

with the following reference manual:

[D] **Sofie Pødenphant** and Per Bruun Brockhoff (2018). *mumm: Multiplicative Mixed Models using the Template Model Builder. R package version 0.2.1.*

## Conference talks

Alexandra Kuznetsova and **Sofie Pødenphant Jensen**, *Mixed Effects Models Using R*, Visionday 2016, May 24, 2016, Technical University of Denmark, Kgs. Lyngby, Denmark. (Joint talk)

**Sofie Pødenphant Jensen**, Kasper Kristensen & Per Bruun Brockhoff, *mumm: An R-package for fitting multiplicative mixed models using the Template Model Builder (TMB)*, UseR!: The R User Conference 2016, June 27-30, 2016, Stanford University, Stanford, California, USA.

**Sofie Pødenphant Jensen**, Kasper Kristensen, Minh H. Truong, Per Bruun Brockhoff, *Investigating the Mixed Assessor Model with use of the Template Model Builder (TMB)*, Sensometrics 2016: the 14th Conference of the Sensometric Society, July 27-29, 2016, Brighton, UK.

**Winner of the student award.**

# Acknowledgements

---

First of all, I would like to thank my supervisor Professor Per Bruun Brockhoff. I am very grateful for all the help, support and ideas you gave me during my PhD, but also for creating a joyful work environment with your very positive and friendly spirit. I also want to thank my co-supervisor Kasper Kristensen. Thank you for always taking your time to help me solving whatever technical problem, I have bumped into during the last three years. Further, I want to thank my second co-supervisor Murat Kulahci for encouraging meetings and guidance, when I needed it the most.

I want to thank Professor Hans-Peter Piepho for enlightening discussions and advice. I am very thankful for all the feedback you have given me on my work and for always having a desk ready for me. Thanks to the rest of the group in Germany for always making me feel welcome in Hohenheim.

I want to send a lot of hugs to my current and former PhD colleagues as a thanks for all the fun moments and for making me feel happy about going to work everyday. Especially I want to thank Azucena Campillo Navarro for keeping me company every evening in the office during the last part of the project, and Agnes Martine Nielsen for being right next to me from the very beginning to the very end of my PhD. I would also like to thank Nina Munkholt Jakobsen for giving me priceless feedback on my work, when it was very needed.

I want to send a special thanks to my family for always being there for me. Lastly, I want to thank my boyfriend and co-author Steffen Hadasch. Thank you for all your love and support. You made my external research stay in Stuttgart unforgettable and you have enriched my days ever since.



# Contents

---

<b>Summary (English)</b>	<b>i</b>
<b>Summary (Danish)</b>	<b>iii</b>
<b>Preface</b>	<b>v</b>
<b>List of contributions</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview of the thesis . . . . .	3
<b>2 Biadditive Models</b>	<b>5</b>
<b>3 The multiplicative mixed model</b>	<b>7</b>
3.1 Estimation of the multiplicative mixed model . . . . .	8
3.1.1 The Template Model Builder (TMB) . . . . .	11
3.2 The multiplicative mixed model in sensometrics . . . . .	16
3.2.1 The Mixed Assessor Model (MAM) . . . . .	20
3.3 The multiplicative mixed model in agriculture . . . . .	21
3.4 The multiplicative mixed model in medicine . . . . .	23
3.5 Properties of the maximum likelihood estimator . . . . .	25
3.6 The multiplicative mixed model and significance testing - how many degrees of freedom? . . . . .	30
3.7 Power to detect significant differences . . . . .	33
3.8 Advantages of the Laplace approximation . . . . .	36

---

<b>4</b>	<b>R-package mumm</b>	<b>39</b>
4.1	The functions in mumm . . . . .	39
4.1.1	Function mumm . . . . .	40
4.1.2	Function ranef . . . . .	41
4.1.3	Function confint . . . . .	43
4.1.4	Function lrt . . . . .	44
4.2	Estimation of extended models . . . . .	45
<b>5</b>	<b>The shifted multiplicative model</b>	<b>47</b>
<b>6</b>	<b>Conclusion</b>	<b>51</b>
6.1	Future Work . . . . .	52
	<b>Bibliography</b>	<b>54</b>
<b>A</b>	<b>Paper A</b>	<b>59</b>
<b>B</b>	<b>Paper B</b>	<b>89</b>
<b>C</b>	<b>Paper C</b>	<b>115</b>
<b>D</b>	<b>Reference Manual</b>	<b>133</b>

## CHAPTER 1

# Introduction

---

100 years ago, Ronald Fisher defined the term *variance* as the squared standard deviation, which he thought was practical since the variance of the sum of independent random variables is simply the sum of the individual variances (Fisher, 1918). A few years later, in 1921, he published an application of analysis of variance (ANOVA) for the first time (Fisher, 1921). Ever since Fisher included ANOVA in his book "Statistical Methods for Research Workers" from 1925 (Fisher, 1925), the method has been widely known. Today, 100 years later, ANOVA models are used to analyze data from most scientific areas.

The purpose of ANOVA models is to examine the influence of categorical variables (factors) on a continuous dependent variable called the response. The response could for instance be the yield of a wheat plant, which we wish to describe by two factors, where the levels of the variables correspond to different genotypes of the plant and different locations. Often the influence of the factors on the response is not additive, meaning that the response is not well described by simply adding the main effects of the two factors. This is due to a potential interaction between the factors. In the wheat plant example, a specific genotype might for instance result in a relatively high yield in general, but be very sensitive to the amount of water, such that it performs very poorly in a dry location, compared to other genotypes. In that case, there is an interaction between the effect of the genotype and the effect of the location. The interaction is often assumed to be completely unstructured, but this is not always a reasonable as-



sumption. In some situations, (a part of) the interaction is better described by a multiplicative term, which models the interaction by linear regression of the response on one of the factors, where the slope depends on the level of another factor. In the example with the wheat plants, it could make sense to regress the yield on the location main effect, where the slope depends on the genotype; the more sensitive a genotype is to the location, the bigger the slope. This approach is known as Finlay-Wilkinson regression and was proposed by Yates and Cochran already in 1938, but popularized in 1963 by Finlay and Wilkinson.

In general, when a two-way ANOVA model is extended by one or more multiplicative terms, we call the resulting model a biadditive model. This definition was proposed by Denis and Gower in 1994. Biadditive models are widely used in several scientific disciplines, including agriculture, as illustrated by the small wheat plant example. Other examples of use can be found in medicine, sensorimetrics and social sciences ([Hawkins and Sharma, 2010](#); [Smith et al., 2003](#); [Gollob, 1968](#)). The motivation for applying a biadditive model instead of a standard two-way ANOVA is to obtain a better understanding and interpretation of the interaction and to achieve improved inference about the effects in the model.

Depending on the aim of the analysis, it might make sense to consider one of the factors as random. This assumption is only reasonable, if the levels of the factor are randomly chosen from a larger population of possible levels. In the wheat plant example, it might for instance be reasonable to consider the locations in the experiment to be randomly chosen from a large population of possible locations. In that case, the effect of location can be considered random, which makes sense if the aim of the analysis is to draw inference about the genotypes in the experiment, which is valid in general and not only for the chosen locations. If one of the factors is considered random, the model belongs to the class of biadditive mixed models.

One aim of this thesis is to give an overview of the different types of biadditive models and their applications. Especially one specific version of the biadditive mixed models will be studied in detail. This version, which will be referred to as the multiplicative mixed model, is particularly interesting due to the expectation of the response being confounded with the variance. The confounding makes the model harder to estimate, and therefore one main goal is to develop a user-friendly open-source software tool to fit the multiplicative mixed model. Throughout this thesis, data is assumed to be balanced.

## 1.1 Overview of the thesis

The thesis consists of six main chapters, three journal papers (A, B and C) and an R-package reference manual.

Chapter 2 gives an introduction to biadditive models and thereby introduces Paper C.

Chapter 3 gives a detailed description of the multiplicative mixed model and how to estimate it by maximum likelihood. In this context, the Template Model Builder (TMB) R-package is introduced, which is used in our R-package `mumm` to facilitate the estimation of the multiplicative mixed model. Further, the chapter includes a section about sensory science, which motivates the use of the multiplicative mixed model for analyzing sensory data. In relation to this, the Mixed Assessor Model (MAM) is introduced, which is a linear approximation of the multiplicative mixed model (Brockhoff et al., 2015). In this chapter, the use of the multiplicative mixed model in agriculture and in method comparison studies in medicine is also motivated. Afterwards, the properties of the maximum likelihood estimator, which is used for estimating the parameters in the multiplicative mixed model, is investigated and compared to the estimator used for estimating the parameters in the MAM. Further, the chapter contains a discussion of hypothesis testing by likelihood ratio tests and a comparison of the power to detect significant fixed effect differences for the multiplicative mixed model, the MAM, and the standard linear two-way mixed model. The chapter ends with a small discussion of the advantages of using the TMB R-package. This chapter stands as an introduction to and a discussion of Paper A and B.

Chapter 4 is a detailed guide on how to use our R-package `mumm` to fit the multiplicative mixed model.

Chapter 5 introduces the shifted multiplicative model and discusses how to estimate it.

Chapter 6 wraps up the thesis with concluding remarks and ideas for future work.

Paper A is included in Appendix A. The paper is entitled "The Multiplicative Mixed Model with the `mumm` R-package as a General and Easy Random Interaction Model Tool" and is under review for publication in the *American Statistician*. The paper gives a detailed description of the multiplicative mixed model and presents our R-package `mumm`, which is the first R-package that can fit this particular model. Further, the model is applied to data from sensory science, agriculture, and method comparison studies in medicine to motivate its

use.

Paper [B](#) is included in [Appendix B](#). It is entitled "The Mixed Assessor Model and the multiplicative mixed model (MAM and mumm)" and is under review for publication in *Food Quality and Preference*. The paper describes the multiplicative mixed model and the Mixed Assessor Model in the context of sensory science. In the paper, the power to detect product differences is compared for the multiplicative mixed model, the MAM and the standard linear two-way mixed model. In [Brockhoff et al. \(2015\)](#), a method to compute product difference confidence intervals is suggested but never validated. In Paper [B](#), this method is evaluated and the resulting confidence intervals are compared to the profile likelihood estimated confidence intervals, obtained by use of our R-package `mumm`. The paper further suggests a new F-test for the MAM for testing the significance of an overall product effect.

Paper [C](#) is included in [Appendix C](#). The paper, entitled "Biadditive Models - a model overview and estimation by the Template Model Builder", is a draft intended for a statistics journal. It gives a review of biadditive (mixed) models regarding their history, their applications and how they can be estimated. The appendix of the paper includes C++-templates, defining the negative joint log-likelihood function, for a few selected models. These C++-templates are needed to estimate the models by the TMB R-package.

[Appendix D](#) is the reference manual of our R-package `mumm`.

# Biadditive Models

---

Biadditive models for two-way tables are two-way ANOVA models extended with one or more multiplicative term(s). They are formally defined in [Denis and Gower \(1994\)](#) and [Denis and Gower \(1996\)](#) as models on the form:

$$y_{ij} = \mu + \alpha_i + \gamma_j + \sum_{m=1}^M \beta_{im} \nu_{jm} + e_{ij}, \quad (2.1)$$

$$i = 1, \dots, I; \quad j = 1, \dots, J; \quad e_{ij} \sim N(0, \sigma^2),$$

where  $\alpha_i$  is the row main effect,  $\gamma_j$  is the column main effect, and  $M$  is the number of multiplicative terms. In this thesis, the definition of biadditive models is extended to also include models in the form of (2.1) with replicates, i.e.

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \sum_{m=1}^M \beta_{im} \nu_{jm} + \delta_{ij} + e_{ijk}, \quad (2.2)$$

$$k = 1, \dots, K.$$

Biadditive models have been widely used in several scientific areas, especially in agriculture, sensometrics and medicine ([Gauch Jr, 1992](#); [Brockhoff, 2003](#); [Hawkins and Sharma, 2010](#)). The models are relevant whenever the interaction is not completely unstructured but has a pattern, which can be (partly) described by linear regressions. The linear regressions are modelled by the multiplicative terms, where the  $\beta_{im}$ s can be interpreted as regression coefficients and the  $\nu_{jm}$ s as latent covariates, or vice versa.

The models can be altered by dropping either the row main effect or the column main effect or both. Further, the general term,  $\mu$ , might be dropped. The result of dropping for example the column main effect is that the first multiplicative term will model a regression on exactly the column main effect (or a scaled version of it). Models altered in this fashion still belong to the class of biadditive models.

Mixed versions of the biadditive models have also been applied to analyze sensory evaluation data and data from agricultural experiments (Smith et al., 2003, 2005), where either the row effect or the column effect is considered random. In agriculture for instance, genotype-by-environment data is often analyzed by a biadditive mixed model, where the genotype effect is considered fixed and the environment effect is considered random (Piepho, 1997, 1999).

An overview of the different biadditive models, their applications, and suggestions on how to estimate them are given in Paper C.

The rest of the thesis mainly deals with a mixed version of model (2.2), where  $i$  is considered random and  $j$  is considered fixed, where the column main effect,  $\gamma_j$ , is dropped from the model and where  $M = 1$ . As written in Chapter 1, this model is particularly interesting because it has several applications, but is troublesome to estimate, since the expectation of  $y_{ijk}$  is confounded with the variance. This is discussed further in Chapter 3. From now on, this model will be referred to as the multiplicative mixed model.

## CHAPTER 3

# The multiplicative mixed model

---

In this thesis, one specific biadditive model will be in focus, which is the following multiplicative mixed model

$$\begin{aligned} y_{ijk} &= \mu + a_i + \nu_j + b_i \nu_j + d_{ij} + \epsilon_{ijk}, \\ i &= 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \\ d_{ij} &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \quad \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \\ (a_i, b_i) &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\right), \end{aligned} \tag{3.1}$$

where the row dependent terms are written with Latin letters to emphasize that they are considered random. By the following small reparameterization the model takes the form of (2.2) (without  $\gamma_j$  and with  $M = 1$ ):

$$\begin{aligned} y_{ijk} &= \mu + a_i + \nu_j + b_i \nu_j + d_{ij} + \epsilon_{ijk} \\ &= \mu + a_i + \tilde{b}_i \nu_j + d_{ij} + \epsilon_{ijk}, \end{aligned}$$

with  $\tilde{b}_i = b_i + 1$ . The multiplicative mixed model (3.1) can be interpreted as linear regression lines, where we regress on the column main effect, plus a random unexplained interaction. The intercepts of the lines are  $\mu + a_i$  and the slopes are  $b_i + 1$ .

**Identifiability constraints:**

Model (3.1) is slightly overparameterized regarding its expectation,  $E(y_{ijk}) = \mu + \nu_j$ , but only one single constraint needs to be imposed to make the model identifiable. The most intuitive choice is to force the  $\nu_j$ s to sum to zero, such that  $\mu$  can be interpreted as the overall mean and  $\nu_j$  as the column dependent deviation from the mean. However, another solution would be to set  $\mu$  equal to zero, such that  $\nu_j$  is the column specific expectation. In that case, one should have  $\nu_j - \bar{\nu}$  in the multiplicative term instead of  $\nu_j$ , to avoid changes in the estimates of the variance components and their interpretation. Yet another possibility is to set  $\nu_1$  to zero, such that  $\mu$  is the expectation of the first column and the rest of the  $\nu_j$ s are the deviations from this. Also in this case the  $\nu_j - \bar{\nu}$  should be used in the multiplicative term to avoid changes in the variance component estimates. The two latter constraining options are very common ways to deal with identifiability issues in R, and both of these options are possible when estimating the multiplicative mixed model by the R-package `mumm` (Pødenphant and Brockhoff, 2018) (see Chapter 4).

**The model without replications:**

In some applications it is not feasible to replicate experiments. When the data is without replicates, the multiplicative mixed model is instead written as

$$\begin{aligned} y_{ij} &= \mu + a_i + \nu_j + b_i \nu_j + \epsilon_{ij}, \\ i &= 1, \dots, I, \quad j = 1, \dots, J, \end{aligned} \tag{3.2}$$

$$(a_i, b_i) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\right), \quad \epsilon_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),$$

where the interaction term,  $d_{ij}$ , has been removed since it is completely confounded with the error.

### 3.1 Estimation of the multiplicative mixed model

The multiplicative mixed model (3.1) can be written in matrix form as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\theta} + \mathbf{Z}(\boldsymbol{\theta})\mathbf{w} + \boldsymbol{\epsilon}, \\ \mathbf{w} &\sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{3.3}$$

where  $\boldsymbol{\theta} = (\mu, \boldsymbol{\nu}^T)^T$  is a vector containing the fixed effects and  $\mathbf{w} = (\mathbf{a}^T, \mathbf{b}^T, \mathbf{d}^T)^T$  is a vector containing the random effects. The matrix  $\mathbf{R} = \sigma^2\mathbf{I}$  is a simple

diagonal-matrix and the matrix  $\mathbf{G}$  is

$$\mathbf{G} = \begin{bmatrix} \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix} \otimes \mathbf{I}_I & \mathbf{0} \\ \mathbf{0} & \sigma_d^2 \mathbf{I}_{IJ} \end{bmatrix},$$

where  $\mathbf{I}_N$  is an identity matrix of size  $N \times N$ .

The form in (3.3) looks very similar to the form of a linear mixed model, except that  $\mathbf{Z}$  contains fixed effects, due to the multiplicative term. Model (3.1) therefore belongs to the class of nonlinear mixed models. The first two moments of  $\mathbf{y}$  are

$$\begin{aligned} E(y_{ijk}) &= \mu + \nu_j \\ \text{Var}(y_{ijk}) &= \sigma_a^2 + \nu_j^2 \sigma_b^2 + 2\nu_j \rho \sigma_a \sigma_b + \sigma_d^2 + \sigma^2, \end{aligned}$$

which can be written in matrix form as

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}\boldsymbol{\theta} \\ \text{Var}(\mathbf{y}) &= \mathbf{V} = \mathbf{Z}(\boldsymbol{\theta})\mathbf{G}\mathbf{Z}(\boldsymbol{\theta})^T + \mathbf{R}, \end{aligned}$$

where it is clear that both the mean and the variance depend on  $\boldsymbol{\nu}$ , which is a result of the non-linearity. This makes the model more difficult to estimate. In the standard procedure for optimization of the likelihood-function of  $\mathbf{y}$ , for a linear mixed model, it is utilized that  $\mathbf{V}$  only contains variance parameters (Gumedze and Dunne, 2011). Since  $\mathbf{V}$  in our case also contains fixed effects, we need another optimization approach.

Let  $\boldsymbol{\psi}$  be a vector containing the variance components of the random effects. The negative marginal log-likelihood of  $\mathbf{y}$  is then given as

$$\ell(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \mathbf{y}) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\mathbf{V}| + \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$$

Assuming that  $\mathbf{y}$  is arranged according to the  $ijk$  subscripts, i.e.

$$\mathbf{y} = (y_{111}, \dots, y_{11K}, y_{121}, \dots, y_{12K}, \dots, y_{IJK})^T,$$

then  $\mathbf{V}$  is a block diagonal matrix

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 & 0 & 0 \\ 0 & \mathbf{V}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{V}_I \end{bmatrix},$$



with

$$\begin{aligned} \mathbf{V}_1 = \mathbf{V}_2 = \dots = \mathbf{V}_I &= \sigma_a^2 \mathbf{I}_{JK} + \sigma_b^2 \boldsymbol{\nu} \boldsymbol{\nu}^T \otimes \mathbf{1}_{K \times K} \\ &\quad + \rho \sigma_a \sigma_b (\boldsymbol{\nu} \mathbf{1}_{1 \times J} + \mathbf{1}_{J \times 1} \boldsymbol{\nu}^T) \otimes \mathbf{1}_{K \times K} \\ &\quad + \sigma_d^2 \mathbf{I}_J \otimes \mathbf{1}_{K \times K} + \sigma^2 \mathbf{I}_{JK}, \end{aligned}$$

where  $\mathbf{1}_{M \times N}$  is an  $M \times N$  matrix with ones.

By removing  $\frac{n}{2} \log(2\pi)$ , since it does not affect the optimization, and by using the block structure of  $\mathbf{V}$ , the negative marginal log-likelihood of  $\mathbf{y}$  can be written as

$$\begin{aligned} \ell(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \mathbf{y}) &= \frac{1}{2} \log \left( \prod_{i=1}^I |\mathbf{V}_i| \right) + \frac{1}{2} \sum_{i=1}^I (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta})^T \tilde{\mathbf{V}}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}) \\ &= \frac{1}{2} I \log(|\tilde{\mathbf{V}}|) + \frac{1}{2} \sum_{i=1}^I (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta})^T \tilde{\mathbf{V}}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}), \end{aligned} \tag{3.4}$$

with  $\tilde{\mathbf{V}} = \mathbf{V}_1 = \mathbf{V}_2 = \dots = \mathbf{V}_I$ . It should be possible to optimize  $\ell(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \mathbf{y})$  with respect to the fixed effects and the variance components, but this is expected to be a slow process, due to the inversion of  $\tilde{\mathbf{V}}$ . Even though, due to the above simplification of the likelihood function, we do not need to find the inverse of  $\mathbf{V}$ , but only of  $\tilde{\mathbf{V}}$ , the problem will expand quickly with the number of products and replicates. In [Piepho \(1999\)](#) maximum likelihood estimates of model (3.1) were found by Fisher scoring. The author however notes that the algorithm "... took many iterations, made slow progress at the last iterations, and appeared to be crisscrossing a ridge of the log likelihood in a hemstiching pattern ...". In this project, another method is therefore pursued, which is based on optimization of the Laplace approximation of the marginal log-likelihood.

The Laplace approximation of the marginal log-likelihood of  $\mathbf{y}$  in (3.1) is given as

$$\ell_{LA}(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \mathbf{y}) = h(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \tilde{\mathbf{w}}, \mathbf{y}) - \frac{1}{2} \log \left( \left| \frac{-\mathbf{H}(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}, \tilde{\mathbf{w}})}{2\pi} \right| \right),$$

where  $\tilde{\mathbf{w}}$  is the maximizer of the joint log-likelihood function,  $h(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}; \mathbf{w}, \mathbf{y})$ , and  $\mathbf{H}(\boldsymbol{\theta}, \sigma, \boldsymbol{\psi}, \mathbf{w})$  is the Hessian of the joint log-likelihood function with respect to  $\mathbf{w}$ . Since the random effects in (3.1) enter the model linearly,  $\mathbf{y}$  is normally distributed, and the Laplace approximation is therefore exact ([Wolfinger and Lin, 1997](#)).

The Laplace approximation of the marginal log-likelihood can be optimized by use of the R-package TMB (Template Model Builder), which will be described in the following section.

Paper [A](#) covers most of the theory described in this section and includes examples on how to apply the multiplicative mixed model.

### 3.1.1 The Template Model Builder (TMB)

The R-package TMB (Template Model Builder) ([Kristensen et al., 2016](#)) enables optimization of the Laplace approximation of the marginal log-likelihood function of nonlinear mixed models. The user needs to define the joint log-likelihood function of the model in a C++ template, but the rest of the user-coding is done in R (see Section [3.1.1.2](#)). By use of Automatic Differentiation (AD) through the high-performance C++-library CppAD, R-package TMB computes the derivatives, up to order three, of the joint log-likelihood function. These derivatives are then used to compute the Laplace approximation of the marginal log-likelihood function and its gradient. In this sense, TMB resembles the C++-package ADMB (AD Model Builder) ([Fournier et al., 2012a](#)), but [Kristensen et al. \(2016\)](#) shows that TMB is faster for models with random effects. The speed-ups were shown to be as big as 100 for large problems. The superior performance of TMB can be partly explained by its use of high-performance libraries but also its ability to take advantage of the sparsity structure of the Hessian matrix of the joint log-likelihood function (see Section [3.8](#)).

The following section explains the concept of Automatic Differentiation, while Section [3.1.1.2](#) gives a short guide on how to use the TMB R-package to fit the multiplicative mixed model.

#### 3.1.1.1 Automatic Differentiation (AD)

Automatic differentiation is a chain rule based technique for computing accurate derivatives of a function. It utilizes that all calculations done by a computer can be broken down into a long sequence of simple operations (+, -, log, sin,  $\sqrt{\cdot}$ , ...). The computer keeps track of all the operations used to evaluate the function in question and builds a representation of it, also known as a *tape* or a *computational graph* ([Kristensen et al., 2016](#)). There exist two main types of automatic differentiation referred to as *forward mode* and *reverse mode*. In reverse mode the program runs the tape backwards, beginning with the model function, and uses the chain rule to update the derivatives one operation at a time. In forward mode, the derivatives are calculated in a similar fashion by the chain rule, beginning with the independent variables ([Madsen and Thyregod, 2010](#)). Thus, we need to run through the tape  $n$  times, with  $n$  being the number of independent variables, when using forward mode. In contrast, only one single

run is required in reverse mode when the model function is a scalar function. Further, it can be proven that in forward mode the cost of evaluating the gradient of a function is less than  $4n$  times the cost of evaluating the function itself, with the cost being measured in the number of simple operations required. On the other hand, the cost of evaluating the gradient in reverse mode is less than 4 times the cost of evaluating the model function (Griewank and Walther, 2008; Fournier et al., 2012b). Consequently, reverse mode is most often superior to forward mode. In comparison, the cost of finding the gradient of a function by the finite difference method is  $n+1$  times the cost of evaluating the function itself, and that method is considerably less accurate than automatic differentiation methods, which emphasizes the benefits of using the latter (Griewank, 1989; Skaug and Fournier, 2006).

There is, however, one practical concern with automatic differentiation, which is the memory usage. All the intermediate variables and derivatives need to be stored, and if the tape contains a large number of simple operations, the memory requirements become immense (Madsen and Thyregod, 2010). However, the tapes can be optimized by smart implementation as shown in Kristensen et al. (2016), where the advantages of the CppAD-library for automatic differentiation in C++ is described.

First order derivatives are usually sufficient for the maximization of likelihood functions. The Laplace approximation to the marginal log-likelihood cannot, however, be described by a straightforward sequence of simple operations due to the inner optimization in the approximation, and since the approximation itself contains up to second order derivatives, up to third order derivatives are necessary for the optimization of it. This is facilitated by TMB by clever application of automatic differentiation.

### 3.1.1.2 How to use TMB

The user needs to write a C++-template, which defines the negative joint log-likelihood function of the mixed model. An example of a C++-template, which defines the negative joint log-likelihood function of the multiplicative mixed model in (3.1) is shown below:

**Listing 3.1:** The C++-template for the multiplicative mixed model in (3.1).

```
#include <TMB.hpp>

template<class Type>
Type objective_function<Type>::operator() ()
{
  /* Data Section*/
```

```

DATA_FACTOR(fixed);
DATA_FACTOR(random);
DATA_FACTOR(randomfixed);
DATA_VECTOR(y);

/* Parameter Section */
PARAMETER(mu);
PARAMETER_VECTOR(a);
PARAMETER_VECTOR(b);
PARAMETER_VECTOR(d);
PARAMETER_VECTOR(nu);
PARAMETER(log_sigma);
PARAMETER(log_sigma_a);
PARAMETER(log_sigma_b);
PARAMETER(log_sigma_d);
PARAMETER(transf_rho);

/* Parameters on natural scale */
Type sigma = exp(log_sigma);
Type sigma_a = exp(log_sigma_a);
Type sigma_b = exp(log_sigma_b);
Type sigma_d = exp(log_sigma_d);
Type rho = transf_rho / sqrt(1. + transf_rho*transf_rho);

/* Include the parameters on natural scale in the report */
ADREPORT(sigma);
ADREPORT(sigma_a);
ADREPORT(sigma_b);
ADREPORT(sigma_d);
ADREPORT(rho);

/* Restricting nu to sum to zero.
 * Resulting vector is nu0
 */
vector<Type> nu0(nu.size()+1);
for(int i=0; i<nu.size(); i++){
    nu0[i] = nu[i];
}
nu0[nu.size()] = -nu.sum();

/* Negative joint log-likelihood function */
Type nll = 0;
int nobs = y.size();
for(int i=0; i<nobs; i++){
    nll -= dnorm(y[i],
                mu + a[random[i]] + nu0[fixed[i]] +
                b[random[i]]*nu0[fixed[i]] +
                d[randomfixed[i]],
                sigma, true);
}

nll -= dnorm(d, Type(0), sigma_d, true).sum();

/* Covariance matrix for (a,b) */

```

```

matrix<Type> Sigma(2,2);
Sigma(0,0)=sigma_a*sigma_a;
Sigma(1,1)=sigma_b*sigma_b;
Sigma(0,1)=rho*sigma_a*sigma_b;
Sigma(1,0)=Sigma(0,1);

for(int i=0; i<a.size(); i++) {
  vector<Type> ab(2); ab(0)=a(i); ab(1)=b(i);
  nll += density::MVNORM(Sigma)(ab);
}

return nll;
}

```

The compiled C++-template, together with the data and starting values for the model parameters, are then given as input to the `MakeADFun` function from the TMB R-package, which outputs functions to calculate the likelihood function and its gradient.

```

library(TMB)
library(lmerTest)
data = TVbo
data$fixed = factor(data$TVset:data$Picture)
data$random = data$Assessor
data$randomfixed = data$random:data$fixed
data$y = data$Dimlasseffect

setwd("C:/sofp/PhD/PhDThesis/Rcode")
compile("model.cpp")

## Note: Using Makevars in C:/Users/sofp/Documents/.R/Makevars
## [1] 0

dyn.load(dynlib("model"))

obj <- MakeADFun(
  data = data,
  parameters = list(
    mu = 0,
    nu = rep(0,nlevels(data$fixed)-1),
    a = rep(0,nlevels(data$random)),
    b = rep(0,nlevels(data$random)),
    d = rep(0,nlevels(data$randomfixed)),
    log_sigma = 1,
    log_sigma_a = 1,
    log_sigma_b = 1,

```

```

    log_sigma_d = 1,
    transf_rho = 0
  ),
  random = c("a", "b", "d"),
  DLL = "model",
  silent = TRUE,
  hessian = TRUE
)

```

Finally, the log-likelihood function is optimized by `nlm` from the `stats` R-package (R Core Team, 2018), which takes as input the starting values and the two functions to calculate the likelihood function and its gradient, and returns the estimated model parameters.

```

opt = nlm(obj$par, obj$fn, obj$gr);
opt

## $par
##      mu      nu      nu      nu      nu
## 3.01666309 -0.79078023 -1.03722107 0.79495350 -0.98933487
##      nu      nu      nu      nu      nu
## 0.16065825 -1.01621601 2.43523930 -0.58116371 0.72948829
##      nu      nu      log_sigma log_sigma_a log_sigma_b
## -0.96222086 2.18169250 0.29012880 0.15697293 -0.09249552
## log_sigma_d transf_rho
## -9.94917280 1.34500720
##
## $objective
## [1] 348.0131
##
## $convergence
## [1] 0
##
## $iterations
## [1] 82
##
## $evaluations
## function gradient
##      112      83
##
## $message
## [1] "relative convergence (4)"

```

The coding of the C++-template can be troublesome for users without C++-experience or without a statistical background. To fit the multiplicative mixed

model without any C++ coding, our R-package `mumm`, which is a wrapper of TMB, can be used instead (see Chapter 4).

## 3.2 The multiplicative mixed model in sensometrics

Sensometrics is the analysis of sensory and consumer data, which is data that stems from experiments, where a panel of human assessors evaluates products by their senses. The common aim is to draw inference about the tested products, which can be anything from food, to lotions, to headphones etc. One classical type of sensory data is sensory profile data, where the assessors rate each product in regards to different attributes, e.g. sweetness, bitterness etc., depending on how he/she perceives the intensity of the attribute.

This type of data, where the response is the scores given by the assessors, with  $J$  products,  $I$  assessors and  $K$  replicates, is often analyzed by a standard two-way ANOVA model (Brockhoff et al., 2015)

$$y_{ijk} = \mu + \alpha_i + \nu_j + \gamma_{ij} + \epsilon_{ijk},$$

$$\epsilon_{ijk} \sim N(0, \sigma^2),$$

where  $\alpha_i$  is the assessor main effect,  $i = 1, \dots, I$ ,  $\nu_j$  is the product main effect,  $j = 1, \dots, J$ ,  $\gamma_{ij}$  is the assessor-by-product interaction and  $\epsilon_{ijk}$  is the random error,  $k = 1, \dots, K$ .

When the aim of the analysis is to draw inference about the products and not about the specific assessors in the panel, it is common to consider the assessor effect as random; assuming that the assessors in the panel are drawn randomly from a large population of assessors. This corresponds to using a two-way mixed ANOVA model

$$y_{ijk} = \mu + a_i + \nu_j + g_{ij} + \epsilon_{ijk},$$

$$a_i \sim N(0, \sigma_a^2), g_{ij} \sim N(0, \sigma_g^2), \epsilon_{ijk} \sim N(0, \sigma^2),$$

where the assessor dependent effects are written with Latin letters, now that they are considered random.

The assessor-by-product interaction models the deviation from additivity, which is present when the assessors perceive the products differently. However, the interaction can also be due to the assessors using the scoring scale differently, in the way that some assessors use a bigger part of the scale than others to

distinguish between the products. This effect is illustrated in Figure 3.1, which shows the scores given by three assessors in a visual test of televisions. The data stems from the *TVbo* data set from the R-package *lmerTest* (Kuznetsova et al., 2016, 2017). Eight assessors tested televisions from Bang & Olufsen, which were characterized by two design factors, *Picture* and *TVset*, with four and three levels, respectively. The two design factors were crossed, yielding 12 products in total. Each assessor evaluated each product twice and the evaluation was based on 15 attributes, of which *Dim glass effect* is chosen for this illustration. The *Dim glass effect* for a television is big if parts of the picture seems dim/dull. In Figure 3.1 the scores given by Assessor 1, 2 and 5 are plotted. We see that the three assessors all agree that product 11 has a larger *Dim glass effect* than most of the other products, but assessor 1 only rates it slightly higher than the remaining products whereas assessor 5 makes the distinction more clear. The figure also reveals that Assessor 1 does not detect any difference between the rest of the products, whereas the two other assessors agree that product 3, 5, 7, and 9 are in the higher end. Further, it seems that assessor 2 has a high scoring baseline.

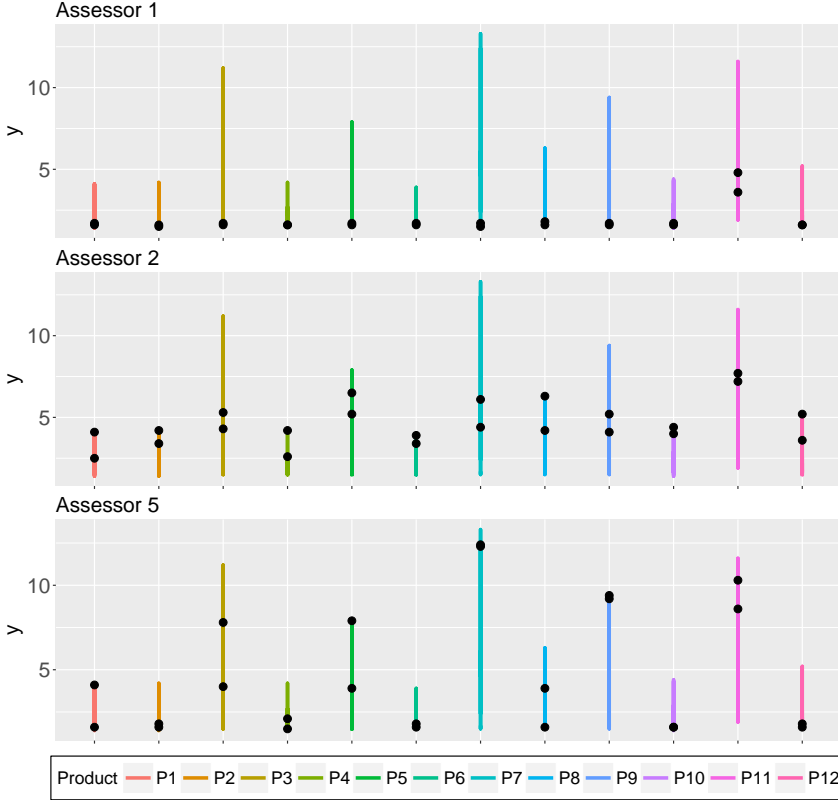
The different baselines of the assessors are captured by the assessor main effect,  $a_i$ . However, the variation due to some assessors using a bigger part of the scale than others, called the scaling effect, is normally pooled into the unstructured and unexplained interaction part, instead of being modelled in a proper way. More specifically, the scaling effect can be captured by a multiplicative term, which models a linear regression on the product effects. The slopes of the regression lines are assessor specific, such that assessors with a big range of scale use will have a steep regression line.

In other words, the product-by-assessor interaction can be divided into a scaling effect and an unexplained disagreement effect. Thereby we end up with the multiplicative mixed model in (3.1), where  $d_{ij}$  is the disagreement term and  $b_i$  is the assessor specific scaling coefficient, resulting in a regression line with slope  $b_i + 1$ .

Table 3.1 shows the parameter estimates for the multiplicative mixed model when fitted to the *TVbo* data set with *Dim glass effect* as the response variable, and Table 3.2 shows the estimates of the random main effect of assessor and the random scaling coefficient. As expected from Figure 3.1, Assessor 1 has a low scaling coefficient and Assessor 5 has a high coefficient. Further, Assessor 2 has a large main effect.

Figure 3.2 illustrates the fitted regression lines for Assessor 1, 2 and 5 plotted together with the corresponding data points. The figure also shows a plot of the fitted regression lines for all of the eight assessors, which emphasizes how different the assessors use the scale. When using the multiplicative mixed model





**Figure 3.1:** The scores given by Assessor 1, 2 and 5.

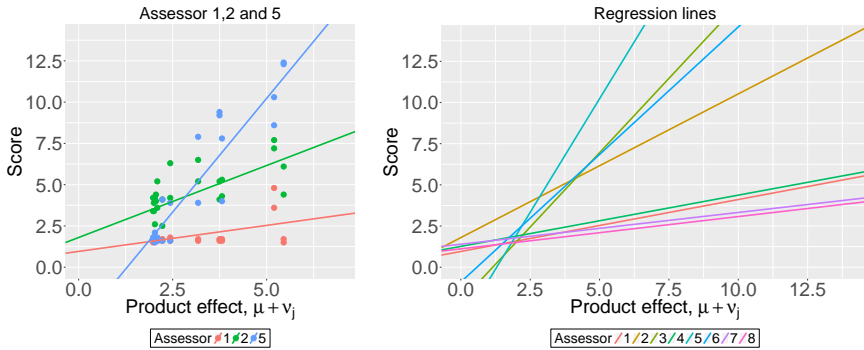
$\mu + \nu_1$	$\mu + \nu_2$	$\mu + \nu_3$	$\mu + \nu_4$	$\mu + \nu_5$	$\mu + \nu_6$
2.2259	1.9794	3.8116	2.0273	3.1773	2.0004
$\mu + \nu_7$	$\mu + \nu_8$	$\mu + \nu_9$	$\mu + \nu_{10}$	$\mu + \nu_{11}$	$\mu + \nu_{12}$
5.4519	2.4355	3.7462	2.0544	5.1984	2.0916
$\sigma$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\rho$	
1.3366	1.1700	0.9117	$6.62 \cdot 10^{-5}$	0.8025	

**Table 3.1:** Estimates of the fixed effects and the variance components.

instead of a standard two-way mixed model, we obtain a more nuanced interpretation of the assessor-by-product interaction, if the scaling effect is present. In a meta study in [Brockhoff et al. \(2015\)](#) with 8619 attributes from 369 sen-

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$
-1.1091	1.4139	0.3557	-0.8133	1.7436	0.7580	-1.0391	-1.3098
$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$
-0.6847	-0.1275	0.8053	-0.6889	1.7586	0.5484	-0.8072	-0.8042

**Table 3.2:** Estimates of the main effects and scaling effects of the assessors.



**Figure 3.2:** The fitted regression lines.

sory profile data sets, the scaling effect was significant in 45% of all cases. This indicates that it is not reasonable to assume that such an effect is not present. Besides a more nuanced interpretation of the interaction, improved statistical inference might also be obtained by the use of the multiplicative mixed model. One of the main goals in the analysis of sensory profile data is to detect potential product differences. Therefore, Paper B investigates if the power to detect product differences increases when using a multiplicative mixed model instead of a two-way mixed model.

As described in Section 3.1, the multiplicative mixed model is not straightforward to estimate. In Brockhoff et al. (2015) a linear approximation to (3.1) is proposed to simplify the estimation. The approximate model, which is named the Mixed Assessor Model (MAM), will be described in the following section.

### 3.2.1 The Mixed Assessor Model (MAM)

The mixed assessor model, proposed by [Brockhoff et al. \(2015\)](#), is a linear approximation to the multiplicative mixed model (3.1) and is written as

$$\begin{aligned} y_{ijk} &= \mu + a_i + \nu_j + \beta_i x_j + d_{ij} + \epsilon_{ijk}, \\ a_i &\sim N(0, \sigma_a^2), d_{ij} \sim N(0, \sigma_d^2), \epsilon_{ijk} \sim N(0, \sigma^2), \end{aligned} \quad (3.5)$$

where  $x_j$ s are the centered product averages computed directly from the raw means ( $x_j = \bar{y}_{.j} - \bar{y}_{...}$ ). Thus,  $x_j$  acts as a known covariate, and the MAM is therefore a simple linear mixed model. The scaling coefficient is in the MAM considered fixed, even though the assessor main effect is considered random. If the  $\beta_i$ s were considered random, the model would be a random coefficient model and [Brockhoff et al. \(2015\)](#) argues that such a model would not be appropriate for this kind of analysis. Even though the scaling coefficient is considered fixed instead of random, the MAM results in valid hypothesis tests for overall product differences and for post hoc product difference testing.

[Brockhoff et al. \(2015\)](#) shows that the hypothesis test for an overall product difference improves when using the MAM compared to using the standard two-way mixed model. However, the MAM fails to produce valid product difference confidence intervals, if the variance contribution from the scaling term is not taking into account. In [Brockhoff et al. \(2015\)](#) a method to produce appropriate product difference confidence intervals is proposed, which takes this variance into account.

In the MAM, the product effects,  $\nu_j$ , are for balanced data estimated as simple averages, and the variance components are estimated by the usual expected mean squares approach. Even though the scaling coefficients are considered fixed, [Brockhoff et al. \(2015\)](#) writes up a formula for estimating  $\sigma_b^2$ , which is used in the confidence interval estimation method, to account for the variance contribution from the scaling term. The method is implemented in the R-package `SensMixed` ([Kuznetsova et al., 2018](#)).

Paper [B](#) investigates whether the suggested procedure in [Brockhoff et al. \(2015\)](#) results in proper confidence intervals for the product differences. This is done by simulation studies, where the coverage probabilities of the intervals are calculated. Further, these confidence intervals are compared to profile likelihood based confidence intervals for the multiplicative mixed model.

Paper [B](#) also includes a study of the power to detect product differences when using the MAM instead of the multiplicative mixed model or the standard two-way mixed model.

### 3.3 The multiplicative mixed model in agriculture

Multiplicative models have been widely used to analyze data from agriculture, particularly genotype-by-environment data. Genotype-by-environment data stems from experiments, where different varieties (plants with different genotypes) are grown in different environments. For plant breeders, the key aim of these experiments is to compare varieties across different environments, to know which variety they should breed for commercial purposes. Another aim is to explore the genotype-by-environment interaction in order to identify genotypes that are well adapted to certain environments or to identify genotypes with a stable performance across all environments. For farmers, the key aim might be to find out how much they will benefit from growing a new variety instead of the current one, or to find out what they can expect from growing varieties, recommended by the plant breeders, in their environment.

Already in 1938, Yates and Cochran proposed a multiplicative model to analyse genotype-by-environment data (Yates and Cochran, 1938). This model was later named after Finlay and Wilkison, who promoted the model in 1963 in Finlay and Wilkinson (1963). The Finlay-Wilkinson model is given by

$$y_{ij} = \mu + \alpha_i + \beta_i \nu_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \quad (3.6)$$

where  $\alpha_i$  is the main effect of genotype  $i$ ,  $\nu_j$  is the main effect of environment  $j$  and  $\beta_i$  is a genotype specific regression coefficient, which can be interpreted as a measure of the sensitivity of genotype  $i$  to the environments. In other words, using this model, genotype-specific regressions on the environmental quality is performed.

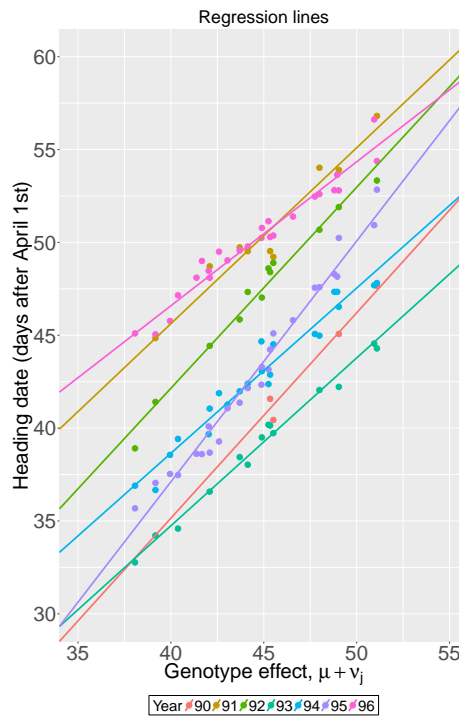
Another well-known multiplicative model for the analysis of genotype-by-environment data is the GGE model (genotype main effects and genotype-by-environment interaction model) introduced by Gauch and Zobel et al. in 1988 in Gauch Jr (1988) and Zobel et al. (1988), and named by Yan in 2000 in Yan et al. (2000). The GGE model often has multiple multiplicative terms, but in the case of only one multiplicative term, the GGE is similar to (3.6), except that  $i$  denotes the environment and  $j$  the genotype. This means that the model performs environment-specific regressions on the genotype effect.

It has been heavily discussed whether the genotype effect and/or the environment effect should be considered fixed or random (Smith et al., 2005). Following Smith et al. (2005) the decision should depend on the aim of the analysis. If the aim is to rank the genotypes as precisely as possible, it might make sense to regard the genotype effects as random and environment effects as fixed. On the

other hand, if the aim is to estimate the difference between varieties, the genotype effects should be regarded as fixed since the BLUP (best linear unbiased prediction) of a difference is biased. If the inference about the varieties should be valid in general across different environments and if the environments in the experiment can be assumed to be randomly chosen from a larger population, it will be reasonable to regard environment effects as random.

If the genotype effect is considered as random and the environment effect as fixed, the Finlay-Wilkinson regression is identical to (3.2). This situation is considered in Paper A. If the genotypes are instead regarded as fixed and environments as random, the GGE model with one multiplicative term is identical to (3.2). This situation is considered in the rest of this section.

In a special type of genotype-by-environment data, the investigated trait is the heading date of grass varieties, where the environment factor often will denote different years instead of different locations. The GGE model with one multiplicative term has shown to be appropriate for the analysis of this type of data (Piepho, 1999). This is explained by the fact that in a year with a late spring, the heading dates of different varieties tend to fall close to each other, whereas the heading dates in a year with an early spring are more widespread. Hence, a linear regression on the genotype effect with year specific slopes makes sense to describe the heading dates; an early spring will result in steep line and a late spring will result in a relatively flat line. In Piepho (1999) the genotypes are regarded as fixed and the environments (years) are regarded as random. This makes sense, since we are typically not interested in a comparison of the specific years but in the heading date of specific varieties in years to come. Thus, model (3.2) is an appropriate choice for this type of data. To illustrate this, model (3.2) is fitted to a data set from Piepho (1999). The data set contains the heading dates of 25 grass varieties collected in year 1990 to 1996. The heading dates are measured in the number of days after April 1st. The fitted regression lines are shown in Figure 3.3. It seems that the spring was early in 1995, since the line for that year is relatively steep. It should further we noticed how nicely the regression lines fit the data points.



**Figure 3.3:** The fitted regression lines.

### 3.4 The multiplicative mixed model in medicine

In medicine, a lot of different measurements are taken on patients, e.g. blood pressure, blood sugar level, cardiac stroke volume, oxygen uptake and infection count. Often several methods to take a specific type of measurements exist, and a comparison between these methods might be wished for. Such investigations are called method comparison studies. To our knowledge, the multiplicative mixed model in (3.1) has never been applied to analyze data from method comparison studies. In this section though, it will be argued why the model is relevant for this kind of analysis.

[Altman and Bland \(1983\)](#) propose an approach to compare two measurement methods. They suggest to check if the differences between the methods correlate with the size of the measurements, by plotting the the differences against the averages. In case of a significant correlation they recommend to transform the data to see if this correlation can be removed. If not, they suggest to describe

the difference between the methods by regressing the differences on the averages. In [Bland and Altman \(1999\)](#) they use so-called *limits of agreement* (LoA) to assess whether or not two methods can be used interchangeably. Limits of agreement define an interval where most differences between measurements by the two methods will lie within, and if this interval only contains differences small enough to not be clinically important, it does not matter which method is used. However, if there is an association between the differences and the size of the observations, the limits of agreement will be wider apart than necessary in some regions of the measurement scale but too narrow in other regions. Therefore they again suggest to regress the difference between methods on the average and adjust the limits of agreements accordingly, such that the limits depend on the size of the measurements. A similar approach is used in [Carstensen \(2010\)](#), where the aim is to produce a conversion between the methods, when the difference between the methods is not constant. When the difference is not constant, the most obvious extension of the standard two-way ANOVA model is to include a multiplicative term such that a measurement on patient  $j$  by method  $i$  depends linearly on the true mean of patient  $j$ . The model can then be written as

$$\begin{aligned} y_{ij} &= a_i + \beta_i \nu_j + \epsilon_{ij}, \\ \epsilon_{ij} &\sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \end{aligned} \tag{3.7}$$

where  $a_i$  is the main effect of method and  $\nu_j$  is the main effect of patient/item. [Carstensen \(2010\)](#) shows how to use the model parameters to predict a measurement on a new patient by method 1, and to find the prediction limits, when measurements on that patient has only been taken by method 2. He further shows, how the parameters can be estimated by regressing the differences on the means.

[Hawkins and Sharma \(2010\)](#) consider method comparison studies with multiple methods, and consider model (3.7) for the analysis. They argue that the model is useful, when the methods do not have the same linear calibration.

In [Altman and Bland \(1983\)](#), [Bland and Altman \(1999\)](#) and [Carstensen \(2010\)](#) all of the effects are considered fixed. [Ekstrøm and Carstensen \(2017\)](#) argue, however, that sometimes the measurement methods in an experiment can be regarded as a random sample from a larger population of possible methods, meaning that the method effects should be regarded as random. This is especially relevant for situations where each method represents a doctor, who evaluates the patients. In this case, method comparison studies suddenly have very much in common with analysis of sensory data, described in Section 3.2. The aim of such method comparison studies will often be to investigate how well medical doctors in general agree on the evaluation of a patient's condition and not to investigate how well specific doctors agree, which stresses why it makes

sense to regard the method effect as random instead of fixed. In [Ekstrøm and Carstensen \(2017\)](#), however, they do not consider to regress on the patient effect like in (3.7), and they therefore use a standard two-way mixed ANOVA model. It would make sense, though, to combine the regression model in (3.7) with the idea of regarding the method effects as random, and thereby end up with the multiplicative mixed model in (3.2).

Paper [A](#) includes an example where the multiplicative mixed model is fitted to data from method comparison studies and shows how the limits of agreement change when the method effects are regarded as random.

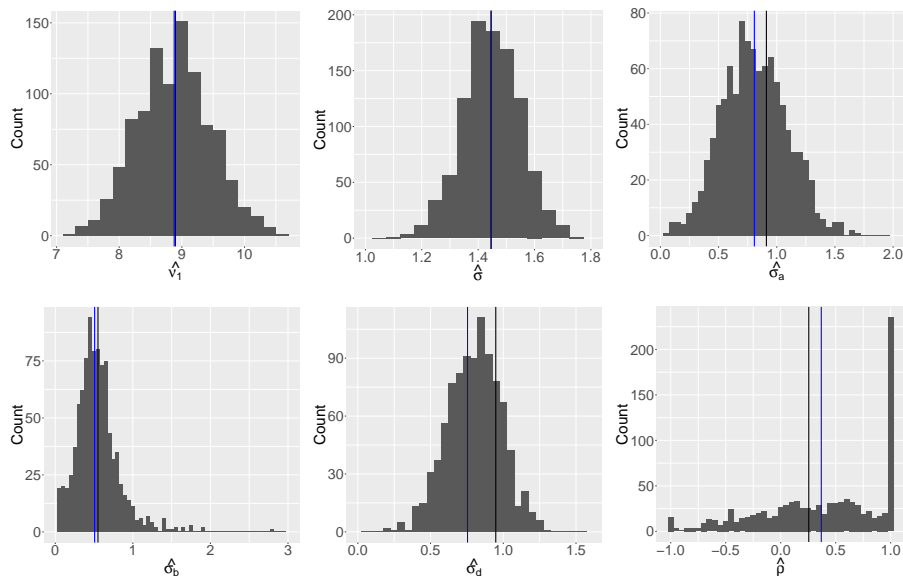
### 3.5 Properties of the maximum likelihood estimator

The parameters in the multiplicative mixed model are, as mentioned, estimated by optimization of the likelihood function. In this section, the most important properties of the maximum likelihood estimator will be investigated, i.e. unbiasedness, consistency, and efficiency. These properties will be explored by simulation studies, such that the true values of the parameters are known. The data is simulated from the multiplicative mixed model (3.1), where the parameters are set equal to the estimates obtained by fitting (3.1) to the *TVbo* data set described in Section 3.2, with *Colourbalance* as the response variable. These parameter estimates are shown in Table 3.3. The properties of the maximum likelihood estimator will be compared to the properties of the estimator used for fitting the parameters in the Mixed Assessor Model (MAM) (recall that the MAM is a linear approximation of the multiplicative mixed model, see Section 3.2). The latter estimator will from now on be denoted as "the MAM estimator".

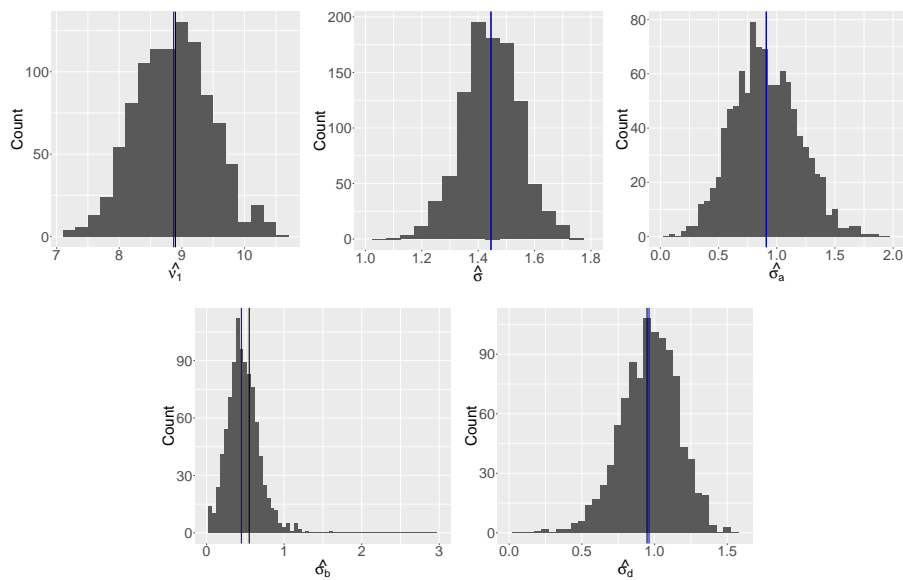
#### Unbiasedness

1000 data sets were simulated and fitted by the multiplicative mixed model and the Mixed Assessor Model. Figure 3.4 shows histograms of the parameter estimates for  $\nu_1$  and for the variance components in the multiplicative mixed model. Similar plots were made for  $\nu_2, \dots, \nu_{12}$ , but they are left out since they resemble the histogram of  $\hat{\nu}_1$ . It seems that the maximum likelihood estimator is unbiased for  $\nu_1$  and  $\sigma$ , but downward biased for  $\sigma_a, \sigma_b$ , and especially  $\sigma_d$ . This is not surprising, since maximum likelihood estimated variance components are known to be underestimated. We further see that  $\rho$  is difficult to estimate, resulting in an upward bias.





**Figure 3.4:** Parameter estimates for model (3.1). The black line marks the true value and the blue line marks the empirical average of the estimates.



**Figure 3.5:** Parameter estimates for the MAM. The black line marks the true value and the blue line marks the empirical average of the estimates.

$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$
8.8967	7.6620	8.3710	8.2482	6.0336	7.1020
$\nu_7$	$\nu_8$	$\nu_9$	$\nu_{10}$	$\nu_{11}$	$\nu_{12}$
7.2425	5.9809	8.8400	9.7595	8.9727	9.2600
$\sigma$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\rho$	
1.4454	0.9110	0.5503	0.9483	0.2548	

**Table 3.3:** Estimated parameters for the multiplicative mixed model (3.1), used to simulate data. Note that  $\mu$  is set equal to zero.

Figure 3.5 shows the corresponding histograms for the parameters in the Mixed Assessor Model, where the fixed effects are estimated as the raw product means and the variance components are estimated by the expected mean squares approach. Even though the slopes (the  $\beta_i$ s) are assumed to be fixed in the MAM, Brockhoff et al. (2015) shows how  $\sigma_b$  could be estimated from the mean squares in the random case. The correlation,  $\rho$ , is not estimated for the MAM. The MAM estimator appears to be unbiased, except for  $\sigma_b$  which is underestimated. Also  $\sigma_d$  might be slightly biased. In general the parameters seem to be less biased when estimated by the linear approximation method instead of using the full likelihood approach.

### Consistency

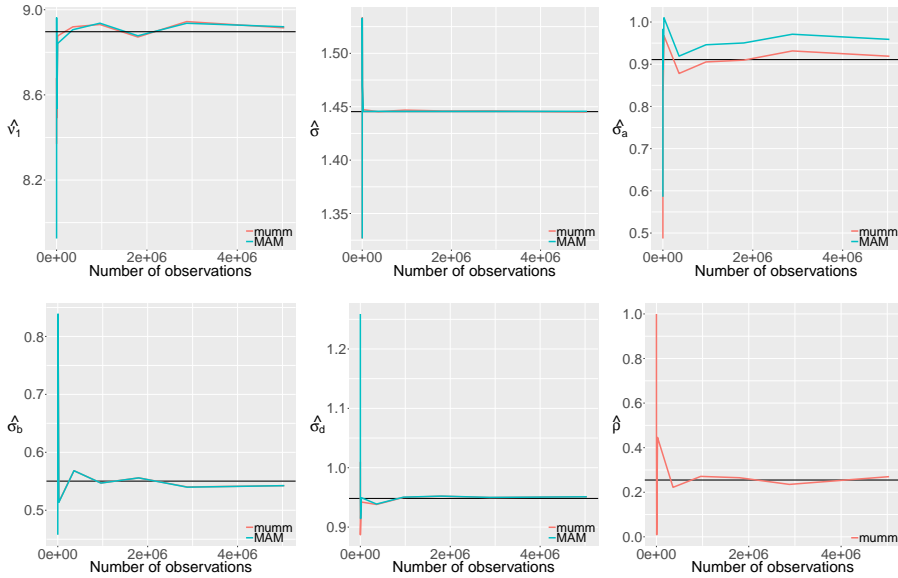
To investigate the consistency of the estimators, nine data sets with an increasing number of observations were simulated. Table 3.4 shows the number of assessors and replicates in the data sets. The parameter estimates for both models are plotted in Figure 3.6, where it, for both models and all the parameters, seems that the estimates converge towards the true value, when the number of observations in the data set increases.

$I$	8	20	50	100	1000	2000	3000	4000	6000
$K$	2	4	10	20	30	40	50	60	70
$N$	192	960	6000	24000	360,000	960,000	$1.8 \cdot 10^6$	$2.88 \cdot 10^6$	$5.04 \cdot 10^6$

**Table 3.4:** The number of assessors, replicates and observations in the data sets used to investigate the consistency.

### Efficiency

The efficiency of the estimators is measured in the mean squared error of the



**Figure 3.6:** The consistency of the estimators.

parameter estimates. 1000 data sets (with 8 assessors and 2 replicates) were simulated and the parameters in the two models were estimated for each data set. The MSE as a function of the number of data sets is plotted in Figure 3.7. Further, the variance and the squared bias of the estimates, as a function of the number of data sets, are plotted. This is to explore if the differences in the MSE are mainly due to differences in the bias or in the variance of the estimators ( $MSE = variance + bias^2$ ) (Sijbers and Den Dekker, 2004). We see that for the fixed effect,  $\nu_1$ , the MSE is highest for the MAM (the same is seen for the rest of the  $\nu$ s). The bias for both models is very close to zero, so the difference in the MSE is explained by the lower variance of the maximum likelihood estimator. The estimators of  $\sigma$  also have a very low bias, but this time they have a very similar variance, which results in a very similar size of the MSE. Looking at the MSE for the estimators of  $\sigma_a$ , the picture changes. This time the MSE is highest for the maximum likelihood estimator, which is due to the large bias. The difference between the MSE is larger when estimating  $\sigma_a$ , due to the very large bias and large variance of the maximum likelihood estimator. The difference in the MSE is lower for the estimators of  $\sigma_b$ , where MAM results in the highest bias but the lowest variance, resulting in a lower MSE. To sum up, the efficiency of the maximum likelihood estimator appears to be lower for the variance components but higher for the fixed effects, compared to the MAM estimator.

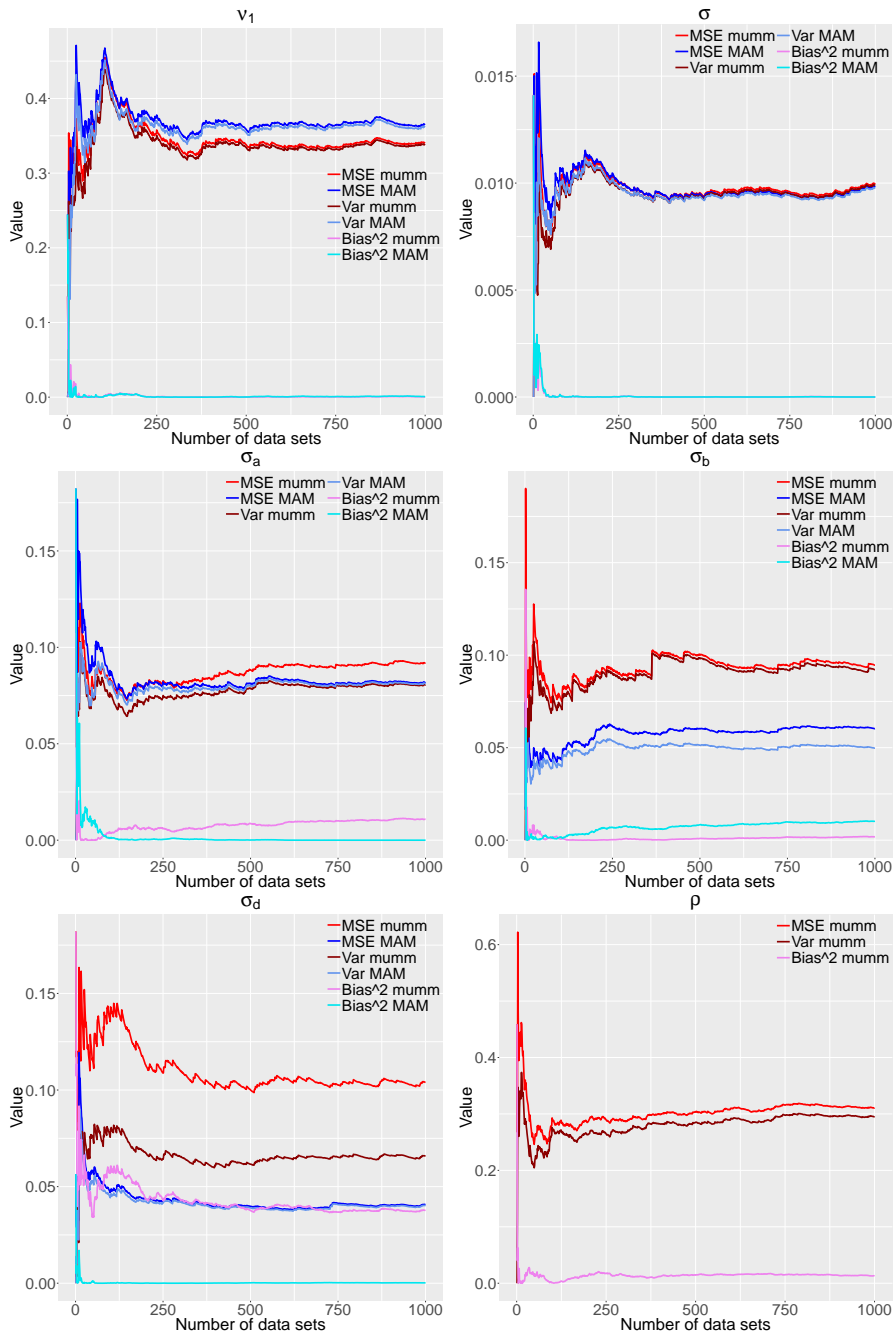


Figure 3.7: The efficiency of the estimators.

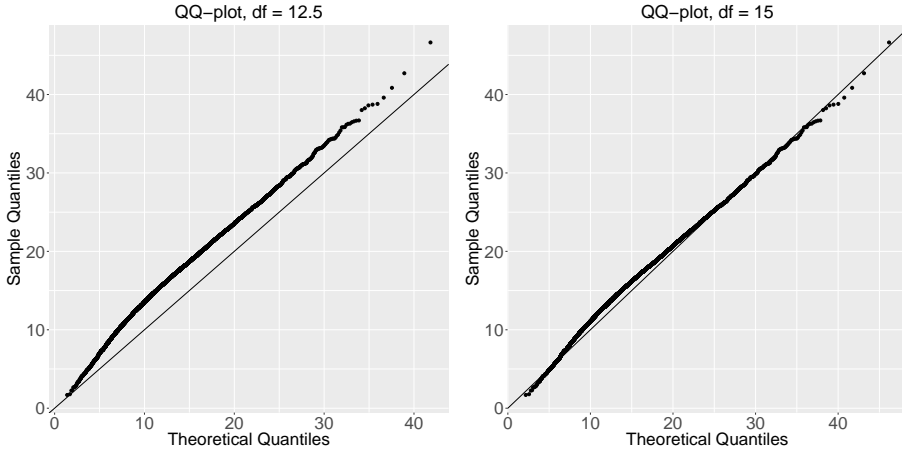
### 3.6 The multiplicative mixed model and significance testing - how many degrees of freedom?

A common method for testing the significance of a model term is to perform a likelihood ratio test, where the likelihood of the null model (without the term) is compared to the likelihood of the alternative model (including the term). The likelihood ratio test statistic, for testing the null hypothesis of the null model being adequate, can be written as

$$W = -2\log\left(\frac{L_{null}}{L_{alternative}}\right) = 2(\ell_{null} - \ell_{alternative}),$$

where  $L_{null}$  and  $L_{alternative}$  are the likelihood of the null model and alternative model, respectively, and  $\ell_{null}$  and  $\ell_{alternative}$  denote the corresponding values of the negative log-likelihood. Wilks's theorem states that, under the null hypothesis,  $W$  approximately follows the  $\chi^2$ -distribution with  $df_{alternative} - df_{null}$  degrees of freedom, where  $df_{alternative}$  and  $df_{null}$  are the degrees of freedom in the two models (Wilks, 1938). However, Wilks's theorem only holds under certain conditions, which for example are not met, when testing the significance of a random effect term. Testing a random effect is equivalent to setting a variance parameter in the null model to zero, which is on the boundary of a the parameter space. In this case the test statistic often approximately follows an equal mixture of a  $\chi_0^2$ -distribution and a  $\chi_1^2$ -distribution, which means that the degrees of freedom in the test should be set equal to 1/2 instead of 1 to overcome this boundary issue (Self and Liang, 1987; Stram and Lee, 1994). Pinheiro and Bates (2000), however, shows that this adjustment is not always successful. Pinheiro and Bates (2000) further warns about using likelihood ratio tests for testing the significance of fixed effect terms in mixed models. The reason is that the test often ends up being "anticonservative", i.e. the p-values are underestimated, when assuming that  $W$  follows a  $\chi_{p-1}^2$ -distribution, where  $p$  is the number of levels in the fixed effect. So to sum up, testing random effects and/or fixed effects can be troublesome by likelihood ratio tests.

When testing the significance of the fixed effect,  $\nu_j$ , in the multiplicative mixed model, not only the  $\nu_j$ s are removed in the null model, since the multiplicative term (and thereby the scaling effect) automatically also vanishes. This means that the null model contains neither,  $\nu$ ,  $\sigma_b$  nor  $\rho$ . Thus, the number of degrees of freedom to use in the test is far from obvious. The intuitive and maybe naive approach would be so assign  $J - 1 + 1/2 + 1 = J + 1/2$  degrees of freedom, but a  $\chi_{J+1/2}^2$ -distribution might be a bad approximation of the actual distribution of  $W$ . A possible solution is to use simulation studies to investigate the empirical distribution of  $W$  under the null hypothesis, i.e finding the p-value by parametric bootstrapping.



**Figure 3.8:** QQ-plots of the likelihood ratio test statistics from the simulation study. The theoretical distributions are  $\chi^2_{12.5}$  (left) and  $\chi^2_{15}$  (right).

If we for example want to use parametric bootstrapping to test the significance of the product effect in the *TVbo* data set described in Section 3.2, we first fit the data to the null model

$$\begin{aligned}
 y_{ijk} &= \mu + a_i + d_{ij} + \epsilon_{ijk}, \\
 i &= 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \\
 a_i &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_a^2), \quad d_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \quad \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),
 \end{aligned} \tag{3.8}$$

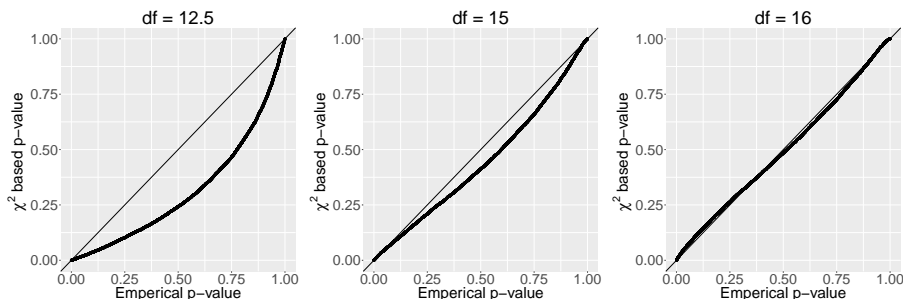
to obtain the maximum likelihood estimated model parameters. These model parameters are then used to simulate data sets from the null model. Afterwards, each data set is fitted by the multiplicative mixed model (alternative model) and the null model, such that the likelihood ratio test statistic,  $W$ , can be calculated. Thereby we obtain an empirical distribution of the test statistic under the null hypothesis. A simulation study has been conducted, where 10.000 data sets are simulated from the null model (3.8), where the parameter estimates are set equal to the estimates obtained by fitting (3.8) to the *TVbo* data set, with *Flickering movement* as the response variable. The parameter estimates are shown in Table 3.5. For comparison, Table 3.6 shows the parameter estimates from fitting the multiplicative mixed model to the data set. Figure 3.8 shows QQ-plots of

$\mu$	$\sigma$	$\sigma_a$	$\sigma_d$
9.8172	3.0533	1.8037	0

**Table 3.5:** Estimated parameters for the null model (3.8).

$\mu + \nu_1$	$\mu + \nu_2$	$\mu + \nu_3$	$\mu + \nu_4$	$\mu + \nu_5$	$\mu + \nu_6$
10.4476	10.5600	8.9988	10.2447	9.8624	10.4506
$\mu + \nu_7$	$\mu + \nu_8$	$\mu + \nu_9$	$\mu + \nu_{10}$	$\mu + \nu_{11}$	$\mu + \nu_{12}$
8.2857	9.7151	10.2484	10.3889	8.3039	10.3004
$\sigma$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\rho$	
2.8091	1.8202	1.0803	0.0020	0.0330	

**Table 3.6:** Estimated parameters for the multiplicative mixed model (3.1).



**Figure 3.9:** Plots of p-values found from the likelihood ratio test statistics from the simulation study. The  $\chi^2$  based p-values are plotted against the p-values obtained by use of the empirical test statistic distribution. The number of degrees of freedom in the  $\chi^2$ -distribution is (from left to right) 12.5, 15 and 16.

the resulting 10.000 likelihood ratio test statistics, where the quantiles from the empirical distribution are plotted against the quantiles from a  $\chi^2_{12.5}$ -distribution and a  $\chi^2_{15}$ -distribution. We see that the intuitive number of degrees of freedom (12.5) is not enough in this case, and that 15 degrees of freedom would be more appropriate, even though the points deviate systematically from the line in one part of the interval. From the test statistics, p-values can be computed either by assuming that  $W$ , under the null hypothesis, follows a  $\chi^2$ -distribution or by using the empirical distribution of the test statistics. Figure 3.9 shows the p-values found from  $\chi^2$ -distributions with 12.5, 15 and 16 degrees of freedom, respectively, plotted against the p-values found by making use of the empirical distribution. The plots show that the p-values are heavily underestimated when assuming that  $W$  follows a  $\chi^2_{12.5}$ -distribution, but also that they are slightly underestimated when using a  $\chi^2_{15}$ -distribution. The p-values found by using a  $\chi^2_{16}$ -distribution, however, seem to coincide fairly well with the p-values found from the empirical distribution.

By using the empirical distribution, we can test the significance of the product effect in our example data set. Model (3.8) and (3.1) are fitted to the data and the test statistic is computed, which gives us the parametric bootstrap p-value equal to 0.1506. This result is quite different from using a  $\chi^2_{12.5}$ -distribution, which gives us a p-value equal to 0.0566. Using a  $\chi^2_{15}$  or a  $\chi^2_{16}$ -distribution gives us more reasonable results with p-values equal to 0.1288 and 0.1688. This shows that the number of degrees of freedom assigned can have a substantial impact, and that one should be careful with "naively" counting the degrees of freedom in the test.

In Paper A, the number of degrees of freedom used in the likelihood ratio test, to test the significance of the fixed effect, is simply set equal to  $J - 1 + 1/2 + 1 = J + 1/2$ , which is not important for the scope of the paper. In Paper B, however, we have chosen to adjust the number of degrees of freedom used in the likelihood ratio test, such that the false discovery rate is close to the significance level. This adjustment is very important for the aim of Paper B, which is to compare different models in regards to their power to detect product differences. For a fair comparison, all the tests should have similar false discovery rates.

### 3.7 Power to detect significant differences

In sensory profile studies, one main goal is to discover potential product differences. Also when analyzing genotype-by-environment data or data from method comparison studies in medicine, we wish to find out if the fixed effect has a significant influence on the response; do some varieties of grass flourish earlier than others or is the difference non-significant? In this section, it is investigated whether or not the use of the multiplicative mixed model results in a higher power to detect significant differences, compared to using a standard two-way mixed model. For this purpose, a simulation study was conducted, where data was simulated from the multiplicative mixed model in (3.1). The variance components were set equal to the estimates from Table 3.1 throughout the whole study, while the size of the product effect varied. The size of the product effect was measured in the standard deviation of  $\nu$ ; when the standard deviation increases from zero, the products go from being completely equal to being more and more different. In the study, 200 data sets were simulated for each value of  $std(\nu)$ , which ranged from 0 to 1. The values of  $\nu$  were drawn from a normal distribution with the desired standard deviation, and were kept constant for all of the 200 data sets. The power to detect a significant product difference is equivalent to the proportion of times the p-value is below the significance level. This proportion was therefore calculated for the multiplicative mixed model, the Mixed Assessor Model and the standard two-way mixed model. The p-value for



the multiplicative mixed model was calculated from likelihood ratio tests, as shown in the previous section, where the number of degrees of freedom in the test was set to 16. The p-values for the MAM and the two-way mixed model were calculated by F-tests. Two different F-tests were used for the MAM: the F-test suggested in Brockhoff et al. (2015) and the new F-test suggested in Paper B. The different tests are described in detail in Paper B, but to get a quick overview, the three different F-statistics are written here:

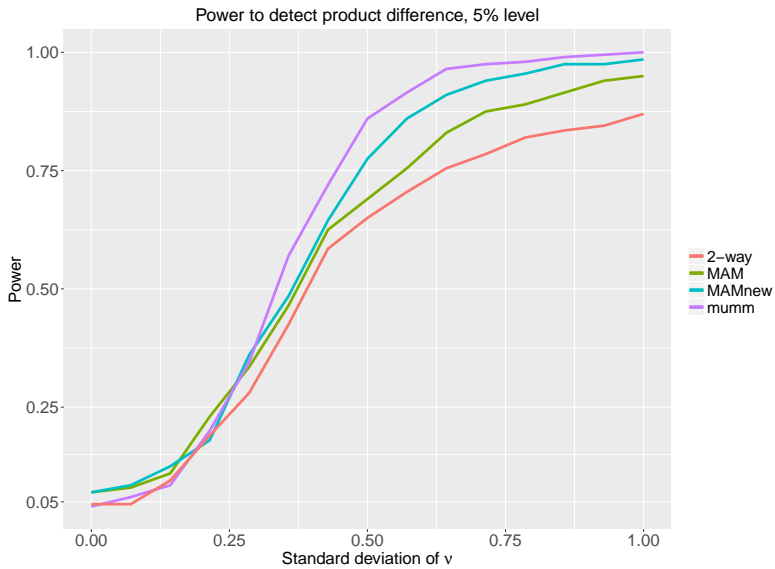
$$F_{2way} = \frac{MS_{product}}{MS_{interaction}}, \quad F_{MAM} = \frac{MS_{product}}{MS_{disagreement}},$$

$$F_{MAMnew} = \frac{MS_{product+scaling}}{MS_{disagreement}},$$

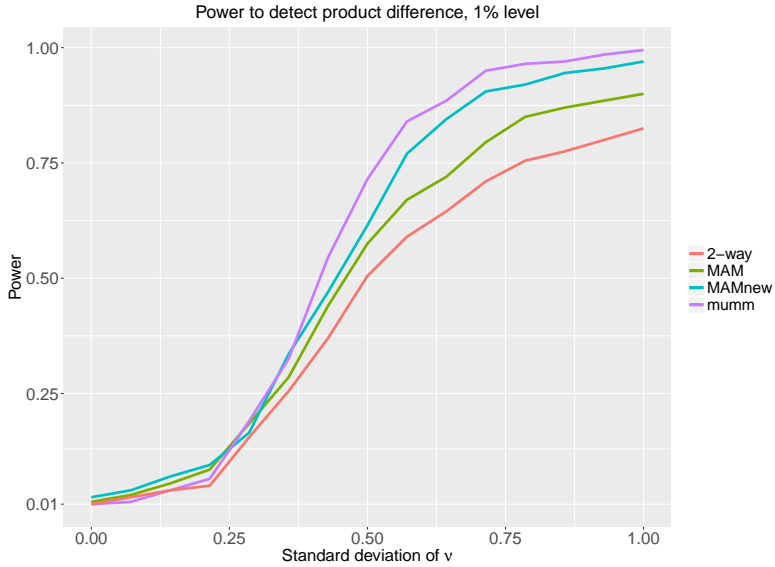
where  $MS_{interaction}$  is the mean square for the product-by-assessor interaction and  $MS_{disagreement}$  is the mean square for the part of the interaction not explained by scaling differences. The idea in the new F-statistic,  $F_{MAMnew}$ , is that a scaling effect can only be present in data if the products are different. Therefore one can argue that the numerator should be a joint mean square for the scaling effect and the product effect, such that the scaling effect contributes to the evidence of an overall product effect. This stands as a big contrast to  $F_{2way}$ , where the scaling effect instead contributes to the error. In  $F_{MAM}$  the scaling effect is removed from the error, but it is not added to the evidence.

Figure 3.10 and 3.11 show the power curves from the simulation study, where the significance level is set at 5% and 1%, respectively. In Figure 3.10, we see that the power to detect a significant product effect is lowest for the 2-way model and highest for the multiplicative mixed model, in most of the interval. We further see that by using the new F-test,  $F_{MAMnew}$ , instead of  $F_{MAM}$ , we obtain an increased power, which is fairly close to the power of the multiplicative model. It should be noticed, though, that the false discovery rate (the "power" when the products are equal) is a bit too high (above 0.05) for both of the MAM F-tests. Looking at Figure 3.11, we get the same picture, except that this time only the new MAM F-test seems to have a slightly too high false discovery rate. The false discovery rate for the multiplicative mixed model seems to be a bit below the significance level, which shows that it was a reasonable choice to assign 16 degrees of freedom in the test.

The results from this section are consistent with the results from Paper B, where we found that by using the multiplicative mixed model or the MAM we get an increased power to detect product differences, when the scaling effect is significant, compared to using a standard two-way mixed ANOVA model. In Paper B, it was also found that by using the MAM with the novel F-test, we obtain a power similar to the power obtained by using the multiplicative mixed model.



**Figure 3.10:** Power curves for testing the significance of an overall product effect, when the significance level is 5%.

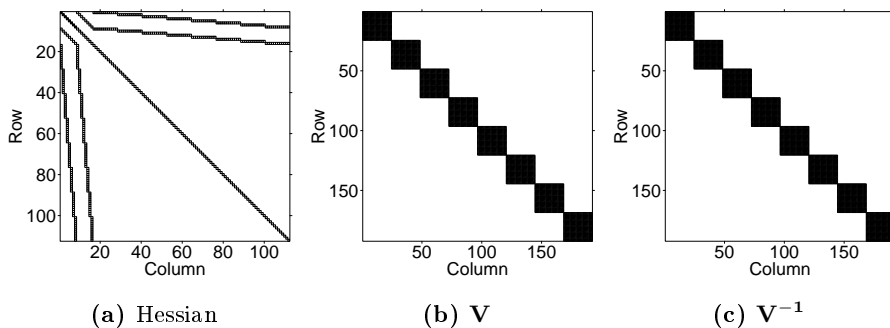


**Figure 3.11:** Power curves for testing the significance of an overall product effect, when the significance level is 1%.

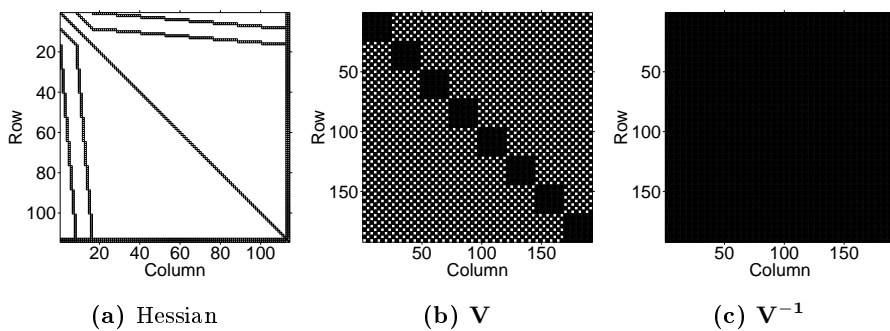
### 3.8 Advantages of the Laplace approximation

In Section 3.1, the log-likelihood function for model (3.1) was written, and it was shown that an inversion of the full covariance matrix,  $\mathbf{V}$ , is not needed in the calculation, due to its block diagonal structure. The block diagonal structure of  $\mathbf{V}$ , means that also  $\mathbf{V}^{-1}$  has a block diagonal structure. In other words, only the inverse of the small blocks,  $\tilde{\mathbf{V}}^{-1}$ , is needed to calculate the log-likelihood function. The pattern of  $\mathbf{V}$  and  $\mathbf{V}^{-1}$  is shown in Figure 3.12 for the *TVbo* data set. The size of the fully dense matrices,  $\tilde{\mathbf{V}}$  and  $\tilde{\mathbf{V}}^{-1}$  increases, though, with the number of products and replicates, which can make the estimation burdensome. The estimation gets even more troublesome, if the model is expanded to e.g. include an effect of replicates, which would introduce a correlation between observations from different assessors. Figure 3.13 shows how the structure of  $\mathbf{V}$  and  $\mathbf{V}^{-1}$  would look like in that case, where we see that  $\mathbf{V}$  is very dense and  $\mathbf{V}^{-1}$  fully dense, which would make the optimization of the log-likelihood function unreasonable burdensome.

However, when optimizing the Laplace approximation of the log-likelihood function instead, we saw in Section 3.1, that only the Hessian of the joint log-likelihood function with respect to the random effects,  $\mathbf{w}$ , is needed. Figure 3.12 and Figure 3.13 show the Hessian for the two different model cases, when the models are fitted to the *TVbo* data set. It is seen that the matrix is sparse in both situations. The sparseness of the Hessian makes it possible to estimate the models fast by R-package TMB, and thereby also mumm, even when the model contains a large number of random effects (Kristensen et al., 2016). Hence, by utilizing the Laplace approximation through TMB, the speed of the optimization is less sensitive to the covariance structure of the data and to model expansions.



**Figure 3.12:** The Hessian of the joint log-likelihood function, the covariance matrix of  $\mathbf{y}$  and its inverse, for model (3.1).



**Figure 3.13:** The Hessian of the joint log-likelihood function, the covariance matrix of  $\mathbf{y}$  and its inverse, for model (3.1) with an added block effect of replicates.



# R-package mumm

---

One of the main goals of this project was to develop user-friendly software tools for the estimation of multiplicative mixed models. The R-package `mumm` (Pødenphant and Brockhoff, 2018) has been developed to facilitate user-friendly and time-efficient estimation of the multiplicative mixed model in (3.1). No other R-package exists, which can fit this type of biadditive mixed models directly. The model can be fitted by the NLMIXED procedure in SAS (SAS Institute Inc., 2009), but as described in Paper A, this procedure has certain disadvantages. Further, a license is needed to use the procedures in SAS. The lack of applicable open-source tools might be the main reason that multiplicative models of the type in (3.1) are not used as much as they could and should be, which is the key motivation for the development of R-package `mumm`.

## 4.1 The functions in `mumm`

This section illustrates how to use R-package `mumm`, by working through a short analysis of the *TVbo* data set presented in Section 3.2. First, the `mumm` package is loaded. Secondly, the R-package `lmerTest` is loaded to get access to the data set, which is thereafter prepared for the analysis.

```
library(mumm)
library(lmerTest)
data = TVbo
data$Product = factor(data$TVset:data$Picture)
data$y = data$Dimlasseffect
```

### 4.1.1 Function mumm

The main function in R-package `mumm` is also named `mumm`, and features estimation of the model parameters in the multiplicative mixed model in (3.1). The syntax is very similar to the syntax of the well-known `lmer` function in R-package `lme4` to estimate linear mixed effects models, which makes function `mumm` very user-friendly. To fit (3.1) to the *TVbo* data set, the `mumm` function is called in the following way

```
fit <- mumm( y ~ 1 + Product + (1|Assessor) + (1|Product:Assessor) +
            mp(Assessor,Product), data = data)

fit

## Multiplicative mixel model fit by ML
## Formula:y ~ 1 + Product + (1 | Assessor) + (1 | Product:Assessor) +
##      mp(Assessor, Product)
## Data: data
## Log-likelihood at convergence: -348.0131
## Random effects:
## Groups          Std.Dev.
## Product:Assessor 5.035067e-05
## Assessor         1.169963e+00
## mp Assessor:Product 9.116466e-01
## Residual         1.336599e+00
## Correlation: 0.8025014
## Number of obs: 192
## Fixed Effects:
## (Intercept) ProductTV1:2 ProductTV1:3 ProductTV1:4
## 2.2258870 -0.2464437 1.5857506 -0.1985680
## ProductTV2:1 ProductTV2:2 ProductTV2:3 ProductTV2:4
## 0.9514391 -0.2254456 3.2260412 0.2096170
## ProductTV3:1 ProductTV3:2 ProductTV3:3 ProductTV3:4
## 1.5202755 -0.1714423 2.9724735 -0.1343194
```

which shows that the input as well as the output resemble those of `lmer`. In the input, the multiplicative term is simply added to the linear mixed model part by

addition of the expression `mp(Assessor, Product)`. As seen, the layout of the output is also almost identical to the output from `lmer`. The biggest difference is that the output from `mumm` shows the correlation between the scaling coefficient and the random main effect. Function `mumm` returns an object of class `mumm`, whereas `lmer` returns an object of class `lmerMod`, or class `lmerModLmerTest` if the `lmerTest` package is loaded.

As seen, the default to handle identifiability issues regarding the fixed main effect, is to set the estimate for the first level,  $\nu_1$ , to zero. Another choice is to set  $\mu$  to zero, i.e. removing the intercept:

```
fit2 <- mumm( y ~ -1 + Product + (1|Assessor) + (1|Product:Assessor) +
             mp(Assessor,Product), data = data)
fit2

## Multiplicative mixed model fit by ML
## Formula:y ~ -1 + Product + (1 | Assessor) + (1 | Product:Assessor) +
##      mp(Assessor, Product)
## Data: data
## Log-likelihood at convergence: -348.0131
## Random effects:
## Groups          Std.Dev.
## Product:Assessor 6.619757e-05
## Assessor         1.169974e+00
## mp Assessor:Product 9.116508e-01
## Residual         1.336599e+00
## Correlation: 0.8025012
## Number of obs: 192
## Fixed Effects:
## ProductTV1:1 ProductTV1:2 ProductTV1:3 ProductTV1:4
##      2.225889      1.979446      3.811649      2.027323
## ProductTV2:1 ProductTV2:2 ProductTV2:3 ProductTV2:4
##      3.177331      2.000432      5.451930      2.435518
## ProductTV3:1 ProductTV3:2 ProductTV3:3 ProductTV3:4
##      3.746183      2.054428      5.198354      2.091577
```

This choice makes it easier to compare the product effects directly, and is a necessity when estimating product contrast confidence intervals.

### 4.1.2 Function ranef

Due to an S3 method implemented in R-package `mumm`, the generic function `ranef` prints the estimated random effect parameters, BLUPs (Best Linear Un-



biased Predictions), of an object of class mumm:

```
ranef(fit)
```

```
## $`Product:Assessor`
##      TV1:1:1      TV1:1:2      TV1:1:3      TV1:1:4
## -2.325227e-11 -1.250669e-09 -1.262338e-09 -7.302998e-10
##      TV1:1:5      TV1:1:6      TV1:1:7      TV1:1:8
##  7.695454e-10  2.833162e-10  2.341217e-09 -2.895973e-10
##      TV1:2:1      TV1:2:2      TV1:2:3      TV1:2:4
## -8.651606e-11  7.786820e-10  1.422737e-10 -1.080300e-09
##      TV1:2:5      TV1:2:6      TV1:2:7      TV1:2:8
## -5.648431e-10  8.920094e-11 -7.826710e-11 -1.526228e-10
##      TV1:3:1      TV1:3:2      TV1:3:3      TV1:3:4
## -1.442409e-09 -9.201945e-10  5.087414e-09 -2.130674e-09
##      TV1:3:5      TV1:3:6      TV1:3:7      TV1:3:8
## -2.989515e-09  1.261196e-09 -1.648581e-09  1.383388e-09
##      TV1:4:1      TV1:4:2      TV1:4:3      TV1:4:4
##  1.254643e-11 -4.751396e-10  3.226987e-10 -5.549447e-10
##      TV1:4:5      TV1:4:6      TV1:4:7      TV1:4:8
## -6.558635e-10  5.883414e-10  4.631665e-10 -1.792323e-10
##      TV2:1:1      TV2:1:2      TV2:1:3      TV2:1:4
## -8.747364e-10  3.630623e-09 -3.440966e-09  1.835289e-09
##      TV2:1:5      TV2:1:6      TV2:1:7      TV2:1:8
##  1.976764e-09 -1.343678e-09 -8.757247e-10 -8.184112e-10
##      TV2:2:1      TV2:2:2      TV2:2:3      TV2:2:4
##  1.785088e-10  3.009594e-10  3.468608e-11 -3.893007e-10
##      TV2:2:5      TV2:2:6      TV2:2:7      TV2:2:8
## -7.292450e-10  7.064619e-10 -7.993007e-10 -3.062021e-10
##      TV2:3:1      TV2:3:2      TV2:3:3      TV2:3:4
## -3.052285e-09 -3.704855e-09  3.352754e-09 -7.410430e-10
##      TV2:3:5      TV2:3:6      TV2:3:7      TV2:3:8
##  2.474186e-09 -3.251227e-09 -2.546241e-09 -1.940736e-09
##      TV2:4:1      TV2:4:2      TV2:4:3      TV2:4:4
## -6.893916e-11  3.764689e-09 -6.581754e-11  5.036723e-10
##      TV2:4:5      TV2:4:6      TV2:4:7      TV2:4:8
## -1.155447e-09 -3.540833e-10  2.397846e-10 -2.641948e-10
##      TV3:1:1      TV3:1:2      TV3:1:3      TV3:1:4
## -1.383813e-09 -1.183784e-09 -5.078337e-09 -5.118599e-10
##      TV3:1:5      TV3:1:6      TV3:1:7      TV3:1:8
##  7.172892e-09 -4.978846e-09 -1.187024e-09 -9.926644e-10
##      TV3:2:1      TV3:2:2      TV3:2:3      TV3:2:4
##  1.301789e-10  1.728225e-09 -3.839192e-10  2.826904e-09
##      TV3:2:5      TV3:2:6      TV3:2:7      TV3:2:8
## -1.435875e-09  1.036764e-09 -8.288544e-10 -1.943089e-10
##      TV3:3:1      TV3:3:2      TV3:3:3      TV3:3:4
##  4.553884e-09  3.167025e-09 -1.592013e-09  9.019669e-10
```

```
##          TV3:3:5          TV3:3:6          TV3:3:7          TV3:3:8
## -3.771222e-09  6.519558e-09  3.694589e-09  1.464092e-09
##          TV3:4:1          TV3:4:2          TV3:4:3          TV3:4:4
## -4.495244e-11  2.203931e-09  4.192335e-10 -4.697741e-10
##          TV3:4:5          TV3:4:6          TV3:4:7          TV3:4:8
## -1.442709e-09  4.478925e-10  1.441891e-10 -2.149420e-10
##
## $Assessor
##          1          2          3          4          5
## -1.1091233  1.4139454  0.3556883 -0.8133025  1.7436472
##          6          7          8
##  0.7579927 -1.0390939 -1.3097716
##
## $`mp Assessor:Product`
##          1          2          3          4          5
## -0.6846758 -0.1275042  0.8052836 -0.6888492  1.7586094
##          6          7          8
##  0.5484491 -0.8071797 -0.8041680
```

The parameters were already estimated during the call to function `mumm`, but printed nicely by `ranef`. As seen, the BLUPs for the random main effect and the scaling coefficient are identical to the parameters shown in Table 3.2.

### 4.1.3 Function `confint`

A method belonging to the generic function `confint`, for an object of class `mumm`, has also been implemented in the `mumm` package. It can be called in the following way, to estimate the 95%-confidence intervals for all of the fixed effect parameters and variance components:

```
confint(fit2)

##          lower      upper
## ProductTV1:1    1.4854856  3.03973153
## ProductTV1:2    1.2104200  2.73796655
## ProductTV1:3    2.3653676  5.39677224
## ProductTV1:4    1.2682576  2.79466899
## ProductTV2:1    2.1265989  4.40074286
## ProductTV2:2    1.2354320  2.76251981
## ProductTV2:3    2.8657047  8.09581074
## ProductTV2:4    1.6819457  3.31501645
## ProductTV3:1    2.3439964  5.29582349
## ProductTV3:2    1.2981953  2.82414595
```

```
## ProductTV3:3      2.7871934 7.68060812
## ProductTV3:4      1.3389080 2.86893614
## Product:Assessor      NA 0.06151043
## Assessor           0.7393064 2.13878777
## mp Assessor:Product 0.4852908 4.63969500
## Residual           1.2082256 1.48943116
## Correlation        0.2591708 1.65228212
```

We see that the lower confidence interval limit cannot be found for  $\sigma_d$ , meaning that  $\sigma_d$  is not significantly different from 0. The `confint` function has a long running time (almost 6 mins for this example), since the intervals are estimated from likelihood profiles. For each knot in the interpolated likelihood profile curve, the multiplicative mixed model needs to be fitted, which makes likelihood profiles time-consuming to compute. Therefore it is made possible in `confint` to estimate the confidence intervals for user-selected variables instead of all variables. Further, it is possible to achieve e.g. the 90%-confidence intervals instead of the 95%-confidence intervals:

```
confint(fit2, parm = c("ProductTV1:1", "Assessor"), level= 0.90)
```

```
##              lower      upper
## ProductTV1:1 1.615029 2.888415
## Assessor      0.790864 1.914699
```

To compute confidence intervals for product contrasts, a vector with a linear combination of the parameters can be given as the `parm` input. The following call to `confint` computes the confidence interval for the difference between the first two products:

```
confint(fit2, parm = c(1,-1,rep(0,15)))
```

```
##              lower      upper
## parameter -0.4903666 1.118951
```

Several contrast confidence intervals can be computed at the same time, by giving a matrix as input, where each row is a linear combination of the parameters.

#### 4.1.4 Function `lrt`

Function `lrt` has been implemented in R-package `mumm` to facilitate likelihood ratio tests for comparing a null model against a multiplicative mixed model.

The null model can either be another (reduced) multiplicative mixed model or a linear mixed model. To test the significance of the multiplicative term in (3.1), the function can be called as shown in the following example:

```
library(lme4)
fit3 <- lmer(y ~ 1 + Product + (1|Assessor) + (1|Product:Assessor),
            data = data)
lrt(fit,fit3)

## Data: data
## Models:
## Object:y ~ -1 + Product + (1 | Assessor) + (1 | Product:Assessor) +
##      mp(Assessor, Product)
## ...1 :y ~ 1 + Product + (1 | Assessor) + (1 | Product:Assessor)
##      Df logLik Chisq ChiDf pvalue
## Object 17 -348.0
## ..1   15 -384.6 73.15     2 1.11e-16
```

As seen in the output, the difference in degrees of freedom for the two models is  $17 - 15 = 2$ , since the null model contains two parameters less ( $\sigma_b$  and  $\rho$ ). It can be argued though, that the degrees of freedom for variance components should be  $1/2$  instead  $1$ , as mentioned in Paper A and in Section 3.6. For function `lrt` we chose, however, to go with the same choice as in the well-known `anova` function from the `stats` package, for objects of class `lmerMod` and class `lmerModLmerTest`, and thereby assigning 1 degree of freedom to variance components.

## 4.2 Estimation of extended models

R-package `mumm` can not only handle the multiplicative mixed model in (3.1), but also extended versions of the model, where one or more main effects and/or interactions between them are added. Model (3.1) can for example be extended with the random main effect of replications,  $r_k$ , which gives us:

$$\begin{aligned}
 y_{ijk} &= \mu + a_i + \nu_j + b_i \nu_j + d_{ij} + r_k + \epsilon_{ijk}, \\
 i &= 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \\
 r_k &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_r^2), \quad d_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \quad \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \quad (4.1) \\
 (a_i, b_i) &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\right).
 \end{aligned}$$

Model (4.1) can be fitted by the following call:

```
fit4 <- mumm( y ~ 1 + Product + (1|Assessor) + (1|Product:Assessor) +
             mp(Assessor,Product) + (1|Repeat), data = data)
fit4

## Multiplicative mixel model fit by ML
## Formula:y ~ 1 + Product + (1 | Assessor) + (1 | Product:Assessor) +
##      mp(Assessor, Product) + (1 | Repeat)
## Data: data
## Log-likelihood at convergence: -347.96
## Random effects:
## Groups          Std.Dev.
## Product:Assessor 6.868756e-05
## Assessor         1.171830e+00
## Repeat           9.314227e-02
## mp Assessor:Product 9.117640e-01
## Residual         1.334326e+00
## Correlation: 0.8011325
## Number of obs: 192
## Fixed Effects:
## (Intercept) ProductTV1:2 ProductTV1:3 ProductTV1:4
## 2.2259968 -0.2465157 1.5857504 -0.1985807
## ProductTV2:1 ProductTV2:2 ProductTV2:3 ProductTV2:4
## 0.9511817 -0.2255015 3.2261046 0.2094075
## ProductTV3:1 ProductTV3:2 ProductTV3:3 ProductTV3:4
## 1.5202397 -0.1715767 2.9721848 -0.1344515
```

## CHAPTER 5

# The shifted multiplicative model

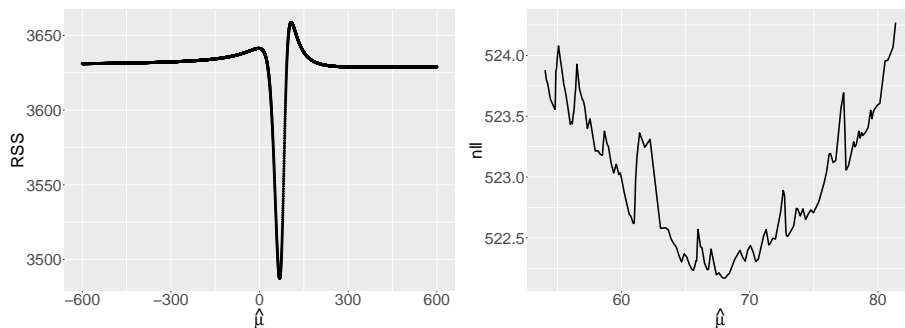
---

The shifted multiplicative model is a special type of biadditive model, where both the row and column main effect are dropped. It can be written as

$$y_{ij} = \mu + \sum_{m=1}^M \beta_{im} \nu_{jm} + e_{ij}, \quad (5.1)$$
$$i = 1, \dots, I; \quad j = 1, \dots, J; \quad e_{ij} \sim N(0, \sigma^2).$$

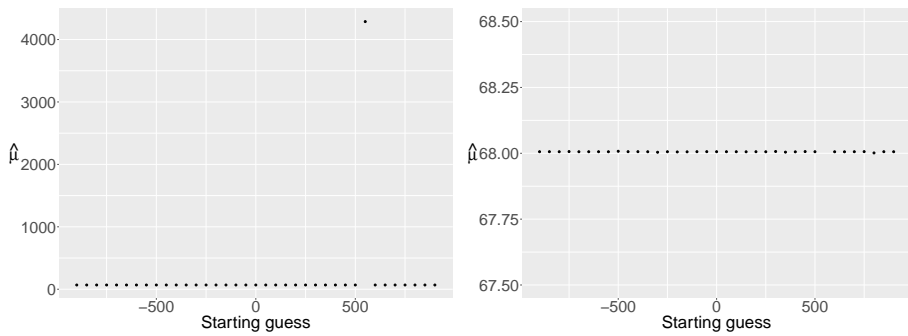
This model stands out from the rest of the biadditive models, in the way that  $\mu$  cannot be interpreted as the overall mean; instead it is called the shift parameter. In Paper C, it is described that the fixed versions of the biadditive model in general can be fitted by first estimating  $\mu$  and the main effects from raw means, and afterwards estimating the multiplicative terms by singular value decomposition. This is, however, not possible for the shifted multiplicative model, since the maximum likelihood fit (or least squares fit) of  $\mu$  is not equal to the raw overall mean. In Paper C, we describe that the shift parameter is often found by minimization of the residual sum of squares as a function of  $\hat{\mu}$ ,  $RSS(\hat{\mu})$ , but that this is not always straightforward due to the shape of  $RSS(\hat{\mu})$ . Often  $RSS(\hat{\mu})$  has several local minima and maxima, and asymptotes when  $\hat{\mu}$  approaches  $-\infty$  or  $\infty$ , which makes it hard to find the global minima.

In this chapter, the shifted multiplicative model with two multiplicative terms



**Figure 5.1:** The residual sum of squares as a function of  $\hat{\mu}$  for the shifted multiplicative model with two multiplicative terms (note that there is a local minimum around  $\hat{\mu} = 400$ ) (left), and the negative log-likelihood as a function of  $\hat{\mu}$  (right).

will be fitted to an example data set. The data set is given in [Piepho \(1998\)](#) and stems from the Bundessortenamt (Hannover, Germany). The observations in the data are yield measurements for 20 varieties of fodder beet at 9 locations. Figure 5.1 shows the restricted sum of squares as a function of the shift parameter. We see that there are two local maxima and two local minima, whereof the local minima at  $\hat{\mu} = 68.01$  is also the global minimum. In [Piepho \(1998\)](#) an algorithm, which needs two starting values, is used to optimize the RSS. With this algorithm, the global minimum is found in 50 out of 89 starting value combinations. In this thesis, the negative log-likelihood is optimized instead, by the TMB R-package. The profile likelihood for  $\mu$  is shown in Figure 5.1, where it is seen that the curve is very "spiky". This is most likely due to the model not being fully identifiable with the chosen constraints. The used C++-template is shown in Paper C, which reveals that no orthogonality restriction is imposed on the multiplicative terms (a standard restriction for models with multiple multiplicative terms). This is not very important as long as the shift parameter can be identified; after  $\mu$  has been estimated, the parameters in the multiplicative term can be found from performing a singular value decomposition of  $\mathbf{Y} - \hat{\mu}\mathbf{1}_{I \times J}$ . Figure 5.2 shows the value of the estimated shift parameter for 37 starting guesses, equally spaced between -900 and 900, and we see that the global minimum is found for all but 1 starting guesses. The method therefore seems relatively robust to the starting guess, compared to the algorithm in [Piepho \(1998\)](#). Optimizing the likelihood function, therefore seems like a good alternative to minimizing the residual sum of squares, especially when the speed of TMB is taken into consideration. In Paper C it was found that the optimization is done in less than 50 ms. The method might get even more robust, with proper identifiability constraints, which would also make it possible to estimate all of the model parameters at once. This will be left for future work.



**Figure 5.2:** The fitted value of  $\mu$  for different starting values, when the likelihood function is optimized by TMB (left). The right plot is zoomed in around  $\hat{\mu} = 68$ .





# Conclusion

---

In this work, the R-package `mumm`, for fitting multiplicative mixed models, was implemented and presented. The package was made user-friendly by imitating the syntax of the heavily used `lmer` function for fitting linear mixed models. In Paper A, the package was further shown to be up to 40 times faster than `NLMIXED` in SAS.

In Paper B, the multiplicative mixed model and the Mixed Assessor Model were compared. The method to compute product contrast confidence intervals for the MAM, suggested by Brockhoff et al. (2015), was shown to perform well, resulting in confidence intervals that emulate the behavior of the profile likelihood estimated confidence intervals. Further, the power to detect an overall product difference was compared for the different models. Using the multiplicative mixed model together with likelihood ratio tests, we obtained the greatest power, when a scaling effect is present. The MAM was also shown to be superior to the standard two-way mixed model. However, to improve the power when using the MAM, we proposed a new F-test. Using the MAM with this new F-test, we obtained a power to detect product differences, which is close to the power obtained when using the multiplicative mixed model. This result was further confirmed in Section 3.7. In the light of this and of MAM having a very low computation time, regarding both the estimation of the model parameters and the product difference confidence intervals, we therefore concluded in paper B that the MAM is a good alternative to the multiplicative mixed model.

In Section 3.5, we saw that the estimator of the variance components in the MAM is less prone to bias and has a higher efficiency. This also speaks to the advantage of using the MAM. However, in Section 3.5, it was also shown that the maximum likelihood estimator of the fixed effect parameters has the highest efficiency. In other words, there is a trade-off between estimator efficiency for the fixed effects and for the variance components, which should be taken into consideration when choosing between the two models, depending on the aim of the analysis. It is further worth mentioning that, due to the advantages of the MAM, it could be relevant to apply the model outside the field of sensometrics, e.g. in agriculture.

In Paper C, we gave an overview of the different biadditive (mixed) model versions, including a literature review. It was further shown how these models can be fitted by use of the R-package TMB. This was found to be especially relevant for the estimation of the shifted multiplicative model, which is otherwise troublesome to estimate. Also concerning the biadditive mixed models with a factor analytic covariance structure, the use of the TMB package is highly relevant, since, to our knowledge, these models could, until now, only be estimated by commercial software like SAS or ASReML.

In general, it was found that multiplicative mixed models and biadditive (mixed) models have various applications across different scientific fields, where the insight and inference can be improved by use of these models, compared to using a standard two-way mixed model. By use of our `mumm` R-package, the multiplicative mixed model is now easy to handle, and we showed that by use of the already existing TMB R-package, the rest of the biadditive models can also be handled.

## 6.1 Future Work

So far, the R-package `mumm` can only handle the multiplicative mixed model in (3.1), and versions of (3.1), where the linear part is expanded. Therefore it would be relevant to further develop the package, such that e.g. biadditive mixed models with a factor analytic covariance structure can be handled as well. This would spare the user from C++-programming, which is needed when using the TMB package.

In this thesis, we have only considered models with one multiplicative term. When analyzing agricultural data, fixed versions of the biadditive models with several multiplicative terms are often applied (Gauch Jr, 1992). Therefore it would be highly relevant to expand the theory from this thesis to include bi-

---

additive mixed models with multiple multiplicative terms, and to make options for multiple multiplicative terms available in `mumm`.

In Section 3.6 it was investigated, how many degrees of freedom should be used in a likelihood ratio test, when testing the significance of the fixed effect in the multiplicative mixed model. However, the number was not obvious and no general rule was found. It would therefore be interesting to dive deeper into this issue. If a general rule could be found, it could be used in the `lrt`-function in `mumm`, for performing likelihood ratio tests.

In Section 3.5 we saw that the maximum likelihood estimator of the variance components in the multiplicative mixed model is biased. Therefore it would be beneficial to use restricted maximum likelihood instead to estimate the model parameters. This is, however, not straightforward due to the confounding of the mean and the variance of the response in the model. Alternatively, it could be investigated if the variance components can be adjusted, after they are estimated by maximum likelihood, to circumvent the bias.

So far, we have only performed power comparisons for situations with replicated experiments. It would be interesting to make similar investigations for data without replicates, i.e. for models without the interaction term, like model (3.2). In relation to this, it would be relevant to develop and evaluate a version of the MAM without the "disagreement effect". Such a version of the MAM could especially be relevant outside the field of sensometrics, e.g. in medicine for method comparison studies, where replicated experiments are rare.

The data sets used for comparing the different models in this thesis are all balanced. Therefore the properties of the estimators and the power to detect significant differences should be investigated for unbalanced data as well, to see how the results in this thesis generalize.



# Bibliography

---

- Altman, Douglas G and Bland, J Martin. Measurement in medicine: the analysis of method comparison studies. *The statistician*, pages 307–317, 1983.
- Bland, J Martin and Altman, Douglas G. Measuring agreement in method comparison studies. *Statistical methods in medical research*, 8(2):135–160, 1999.
- Brockhoff, Per Bruun. Statistical testing of individual differences in sensory profiling. *Food Quality and Preference*, 14(5):425–434, 2003.
- Brockhoff, Per Bruun; Schlich, Pascal, and Skovgaard, Ib. Taking individual scaling differences into account by analyzing profile data with the mixed assessor model. *Food Quality and Preference*, 39:156–166, 2015.
- Carstensen, Bendix. Comparing methods of measurement: Extending the loa by regression. *Statistics in medicine*, 29(3):401–410, 2010.
- Denis, Jean-Baptiste and Gower, John C. Biadditive models. *Biometrics*, 50: 310–311, 1994.
- Denis, Jean-Baptiste and Gower, John C. Asymptotic confidence regions for biadditive models: interpreting genotype-environment interactions. *Applied Statistics*, pages 479–493, 1996.
- Ekstrøm, Claus Thorn and Carstensen, Bendix. Statistical models for assessing agreement in method comparison studies with heterogeneous random raters and replicate measurements. *Annals of Applied Statistics*, In press, 2017.
- Finlay, KW and Wilkinson, GN. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, 14(6):742–754, 1963.

- Fisher, Ronald A. Xv.—the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1918.
- Fisher, Ronald A. On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32, 1921.
- Fisher, Ronald A. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd, 1925.
- Fournier, David A; Skaug, Hans J; Ancheta, Johnnoel; Ianelli, James; Magnusson, Arni; Maunder, Mark N; Nielsen, Anders, and Sibert, John. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249, 2012a.
- Fournier, David A; Skaug, Hans J; Ancheta, Johnnoel; Ianelli, James; Magnusson, Arni; Maunder, Mark N; Nielsen, Anders, and Sibert, John. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2):233–249, 2012b.
- Gauch Jr, H. G. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, 1992.
- Gauch Jr, Hugh G. Model selection and validation for yield trials with interaction. *Biometrics*, pages 705–715, 1988.
- Gollob, Harry F. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33(1):73–115, 1968.
- Griewank, Andreas. On automatic differentiation. *Mathematical Programming: recent developments and applications*, 6(6):83–107, 1989.
- Griewank, Andreas and Walther, Andrea. *Evaluating derivatives: principles and techniques of algorithmic differentiation*, chapter 3, pages 31–59. SIAM, 2008.
- Gumedze, FN and Dunne, TT. Parameter estimation and inference in the linear mixed model. *Linear Algebra and its Applications*, 435(8):1920–1944, 2011.
- Hawkins, Douglas M and Sharma, Abha. Comparison of measurements by multiple methods or instruments. *Journal of biopharmaceutical statistics*, 20(5): 965–984, 2010.
- Kristensen, Kasper; Nielsen, Anders; Berg, Casper W.; Skaug, Hans, and Bell, Bradley M. TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5):1–21, 2016. doi: 10.18637/jss.v070.i05.

- Kuznetsova, Alexandra; Brockhoff, Per Bruun, and Christensen, Rune Haubo Bojesen. *lmerTest: Tests in Linear Mixed Effects Models*, 2016. URL <https://CRAN.R-project.org/package=lmerTest>. R package version 2.0-32.
- Kuznetsova, Alexandra; Brockhoff, Per Bruun, and Christensen, Rune Haubo Bojesen. lmerTest package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, In press, 2017.
- Kuznetsova, Alexandra; Bruun Brockhoff, Per, and Haubo Bojesen Christensen, Rune. *SensMixed: Analysis of Sensory and Consumer Data in a Mixed Model Framework*, 2018. URL <https://CRAN.R-project.org/package=SensMixed>. R package version 2.1-0.
- Madsen, Henrik and Thyregod, Poul. *Introduction to general and generalized linear models*. CRC Press, 2010.
- Piepho, H-P. Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics*, 53:761–766, 1997.
- Piepho, Hans-Peter. An algorithm for fitting the shifted multiplicative model. *Journal of Statistical Computation and Simulation*, 62(1-2):29–43, 1998.
- Piepho, Hans-Peter. Fitting a regression model for genotype-by-environment data on heading dates in grasses by methods for nonlinear mixed models. *Biometrics*, 55(4):1120–1128, 1999.
- Pinheiro, José C and Bates, Douglas M. *Mixed-effects Models in S and S-PLUS*. Springer, 2000.
- Pødenphant, Sofie and Brockhoff, Per Bruun. *mumm: Multiplicative Mixed Models using the Template Model Builder*, 2018. URL <https://CRAN.R-project.org/package=mumm>. R package version 0.2.1.
- R Core Team, . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL <https://www.R-project.org/>.
- SAS Institute Inc., . SAS/STAT 9.2 user’s guide, second edition. *Cary, NC: SAS Institute Inc.*, pages 4378–4379, 2009.
- Self, Steven G and Liang, Kung-Yee. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.
- Sijbers, Jan and Den Dekker, AJ. Maximum likelihood estimation of signal amplitude and noise variance from mr data. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 51(3):586–594, 2004.



- Skaug, Hans J and Fournier, David A. Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis*, 51(2):699–709, 2006.
- Smith, AB; Cullis, Brian R, and Thompson, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science*, 143(6):449–462, 2005.
- Smith, Alison; Cullis, Brian; Brockhoff, Per, and Thompson, Robin. Multiplicative mixed models for the analysis of sensory evaluation data. *Food Quality and Preference*, 14(5):387–395, 2003.
- Stram, Daniel O and Lee, Jae Won. Variance components testing in the longitudinal mixed effects model. *Biometrics*, pages 1171–1177, 1994.
- Wilks, Samuel S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.
- Wolfinger, Russell D and Lin, Xihong. Two taylor-series approximation methods for nonlinear mixed models. *Computational Statistics & Data Analysis*, 25(4): 465–490, 1997.
- Yan, Weikai; Hunt, LA; Sheng, Qinglai, and Szlavnic, Zorka. Cultivar evaluation and mega-environment investigation based on the gge biplot. *Crop Science*, 40(3):597–605, 2000.
- Yates, F and Cochran, WG. The analysis of groups of experiments. *The Journal of Agricultural Science*, 28(4):556–580, 1938.
- Zobel, Richard W; Wright, Madison J, and Gauch, Hugh G. Statistical analysis of a yield trial. *Agronomy journal*, 80(3):388–393, 1988.

APPENDIX A

The Multiplicative Mixed  
Model with the mumm  
R-package as a General and  
Easy Random Interaction  
Model Tool

---

**Pødenphant, S.**, Kristensen, K., & Brockhoff, P. B. (2018). *The multiplicative mixed model with the mumm r-package as a general and easy random interaction model tool*. Submitted to The American Statistician (under review).

# The Multiplicative Mixed Model with the mumm R-package as a General and Easy Random Interaction Model Tool

Sofie Pødenphant\*

Section for Statistics and Data Analysis, DTU Compute,  
Technical University of Denmark, Kongens Lyngby, Denmark  
(e-mail: sofpo@dtu.dk)

and

Kasper Kristensen

Section for Marine Living Resources, DTU Aqua,  
Technical University of Denmark, Kongens Lyngby, Denmark  
and

Per B. Brockhoff

DTU Compute,  
Technical University of Denmark, Kongens Lyngby, Denmark

August 3, 2018

## Abstract

Multiplicative mixed models can be applied in a wide range of scientific disciplines, since they are relevant in every situation where an interaction between a fixed effect and a random effect is present. Until now, no R-package has been published, which can fit this type of models. The lack of user friendly open source tools to fit these models, is the main reason that the models are not used as often as they could or should be. In this paper we introduce the user friendly R-package `mumm` for fitting multiplicative mixed models in a time-efficient manner. To illustrate the interpretation of the multiplicative term, we provide four data analysis examples, where the model is fitted to data sets that stem from studies in sensometrics, agriculture and medicine. With these examples it is shown that the statistical inference can be improved by using a multiplicative mixed model, instead of a linear mixed model which is usually employed.

*Keywords:* Genotype-by-environment data, Method comparison studies, Multiplicative interaction, Sensory profile data, Template Model Builder

---

\*The conduct of the research is fully funded by the Technical University of Denmark (DTU).

---

# 1 Introduction

Linear mixed models are commonly used in a wide variety of disciplines in biological, medical, social, sensory and physical science. They are, however, not always able to describe the complex structure in data sufficiently, and in some cases the data analysis can be improved by extending the linear mixed model with one or more multiplicative term(s), resulting in a so-called multiplicative mixed model (Piepho, 1997, 1999; Brockhoff et al, 2015; Smith et al, 2001, 2003, 2005). A multiplicative term is in the current work considered to be a product of a random effect and a fixed effect, i.e. a term that models a part of the interaction as a random coefficient model based on linear regression on the fixed effect.

This paper presents three examples on how the multiplicative mixed model can be applied in different fields. The first example is from the field of sensory science, where the objective is to compare certain products, based on scores given to these products by a panel of assessors. For this type of data, the use of the multiplicative mixed model results in a more nuanced modelling and interpretation of the panelist-by-product interaction and improved inference about the product effect. This is possible because the multiplicative term explains the part of the interaction that arises from the panelists using the scoring scale differently. In Brockhoff et al (2015) only a linear approximation to the model is fitted, whereas the exact model is fitted in this paper by optimization of the likelihood function.

The second application is in agriculture, where the multiplicative term in the model accounts for genotypes having different sensitivities to the cultivation environment. The last example in this paper deals with method comparison studies in medicine, where the multiplicative mixed model improves the analysis when the medical methods do not have the same linear calibration. Often in method comparison studies, limits of agreement (LoA) are estimated to assess the agreement between two methods. This paper provides a formula for estimating the LoA under the multiplicative mixed model, together with a simulation study that illustrates the difference between the LoA estimated under the multiplicative mixed model and under the standard mixed two-way ANOVA model.

In general the multiplicative mixed model can be useful in every situation where we have an interaction between a fixed and a random effect, e.g. in randomized block designs.

This implies that there are a lot of applications, where a more correct inference might be possible to achieve. Further, the multiplicative mixed model can be used as a model validation of the assumption of completely unstructured and independent random effects.

However, until now, no open source tools to fit this kind of models exists. In this paper, the user-friendly R-package `mumm` (Pødenphant and Brockhoff, 2016), for fitting multiplicative mixed models is presented, which finds the maximum likelihood estimated model parameters quickly by making use of R-package `TMB` (Template Model Builder) (Kristensen et al, 2016). The syntax of `mumm` is very similar to the syntax of the heavily used functions `lm` and `lmer` (R Core Team, 2016; Bates et al, 2015) for fitting linear and mixed linear models, which makes it straightforward to use.

The following section gives a detailed formulation of the model and describes how it is estimated by `mumm`. Section 3 presents the three examples of applications and section 4 includes a small discussion about computation time, where `mumm` and the `NLMIXED` procedure in SAS are compared. Finally, concluding remarks are stated in section 5.

## 2 Method

The version of the multiplicative mixed model studied in this paper can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\theta} + \mathbf{Z}(\boldsymbol{\theta})\mathbf{w} + \boldsymbol{\epsilon}, \\ \mathbf{w} &\sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{1}$$

where  $\boldsymbol{\theta}$  is a vector containing fixed effects, and  $\mathbf{w}$  is a vector containing the random effects. Moreover  $\mathbf{R} = \sigma^2\mathbf{I}_n$ , where  $n$  is the total number of observations.

We will limit our focus to multiplicative models with only one multiplicative term, and only consider models where the fixed effect in the multiplicative term is part of the mean structure, such that  $\mathbf{Z}$  contains parameters from  $\boldsymbol{\theta}$ , meaning that a fixed effect from the mean structure enters the variance structure.

In this paper we will mainly use the following two-way model

$$\begin{aligned}
 y_{ijk} &= \mu + a_i + \nu_j + b_i \nu_j + d_{ij} + \epsilon_{ijk}, \\
 i &= 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \\
 d_{ij} &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \quad \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \\
 (a_i, b_i) &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\right),
 \end{aligned} \tag{2}$$

where we regress on the fixed effect,  $\nu_j$ , and the regression slopes  $(b_i + 1)$  have mean 1.

In situations where the fixed effect from the multiplicative term is *not* a part of the mean structure, the mean and the variance are not confounded and the covariance structure has a factor-analytic form (Piepho, 1997; Piepho et al, 2015). Such models can be fitted by e.g., the MIXED procedure in SAS, and will not be considered further in this paper.

## 2.1 Model Estimation

We want to minimize the negative log-likelihood which is given by

$$\ell(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}; \mathbf{y}) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| + \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}),$$

where  $\boldsymbol{\Psi}$  is a vector containing the variance components of the random effects, and where

$$\begin{aligned}
 \mathbf{V}(\boldsymbol{\theta}) &= \text{Cov}(\mathbf{y}) \\
 &= \mathbf{Z}(\boldsymbol{\theta}) \mathbf{G} \mathbf{Z}(\boldsymbol{\theta})^T + \mathbf{R}.
 \end{aligned}$$

It is important to notice that the covariance matrix contains the fixed effects, since it influences the optimization of the likelihood function.

The standard approach for finding the minimum when working with linear mixed models is to first profile the likelihood so it is a function of the variance components only. For the multiplicative mixed model, however, the fixed effect parameters cannot be profiled out of the likelihood function, since they, as shown above, are part of the covariance matrix.

Alternatively, the multiplicative mixed model can be estimated by maximizing the Laplace approximation to the log-likelihood, which is the approach we will follow in this

paper. The Laplace approximation can be written as:

$$\ell_{LA}(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}; \mathbf{y}) = h(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}; \tilde{\mathbf{w}}, \mathbf{y}) - \frac{1}{2} \log \left( \left| \frac{-\mathbf{H}(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}, \tilde{\mathbf{w}})}{2\pi} \right| \right),$$

where  $h(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}; \mathbf{w}, \mathbf{y})$  is the joint log-likelihood function,  $\mathbf{H}(\boldsymbol{\theta}, \sigma, \boldsymbol{\Psi}, \mathbf{w})$  is the Hessian of the joint log-likelihood function with respect to  $\mathbf{w}$ , and  $\tilde{\mathbf{w}}$  is the maximizer of  $h$ .

The optimization of  $\ell_{LA}$  for the multiplicative mixed model can be done by making use of the R-package `mumm`.

## 2.2 R-package `mumm`

The R-package `mumm` (Pødenphant and Brockhoff, 2016) makes it possible to fit multiplicative mixed models. The package provides a function where the user only needs to give a model formula and the data set as input to get the estimated model fit and standard model summaries as output. The syntax for the model formula is very similar to the syntax in the `lmer`-function for fitting linear mixed models by the `lme4` R-package, which makes package `mumm` very user friendly. For the optimization part, the package makes use of the TMB package, which will be described in the following section.

### 2.2.1 The Template Model Builder (TMB)

The Template Model Builder (TMB) is a recently developed R-package (Kristensen et al, 2016) that enables fast maximization of the Laplace approximation to the marginal log-likelihood function of nonlinear mixed models. The user needs to define the negative joint log-likelihood function in a C++-template, while the remaining code is written in R. The package uses Automatic Differentiation (AD) to compute the derivatives of the joint log-likelihood function with respect to the random effect coefficients, which are used to build the Laplace approximation and its gradient. First order derivatives are usually sufficient for the maximization of likelihood functions, but since the Laplace approximation involves up to second order derivatives, up to third order derivatives are necessary to compute its gradient. This is facilitated by TMB by clever application of automatic differentiation. The Laplace approximation and its gradient are then given as input to a minimizer in R, e.g. the R-function `nlminb` (R Core Team, 2016), which optimizes the Laplace likelihood and returns the maximum likelihood estimates of the parameters.

---

By using R-package `mumm` to fit the multiplicative mixed model, the user exploits the speed of TMB but avoids the coding of C++-templates.

## 3 Applications

In this section, we provide four data analysis examples, where the `mumm` package is used to fit the multiplicative mixed model to data from sensometrics, agriculture and medicine.

### 3.1 Sensory Science

In the field of sensory science, the use of models with multiplicative terms has been proposed and discussed in several publications (see, e.g., Brockhoff et al, 2015; Brockhoff and Skovgaard, 1994; Smith et al, 2003). Sensory profile data, where  $I$  assessors scored  $J$  products in  $K$  replicates, is frequently analysed by a linear mixed two-way ANOVA model. The model contains the overall mean, the random effect of assessor, a fixed product effect, a random panelist-by-product disagreement effect and the random residual error. It has been widely discussed in the sensory literature whether to consider the assessor effect as random or fixed, but as argued in Brockhoff et al (2015), the usual purpose of the experiment is to draw inference about the tested products that may generalize to a larger setting than the assessors entering the panel, and a common approach is thus to consider the effects related to the assessors as random.

The panelist-by-product interaction models the differences in perception of the products but also the differences in the assessors individual ranges of scale use. To account for these individual ranges of scale use, Brockhoff and Skovgaard (1994) introduces a multiplicative term as an extra term in the two-way ANOVA model without interaction. The assessors are in this approach considered to be a fixed effect, which simplifies the estimation of the model. In Brockhoff et al (2015) however, the model is introduced in an extended version where the standard interaction is included and where the assessor dependent effects are considered to be random. The purpose of this model is to partition the panelist-by-product interaction into an assessor dependent scaling effect and the actual disagreement



effect that is due to the differences in perception of the products. The formal multiplicative mixed model, which we are fitting, is not fitted in Brockhoff et al (2015). Instead a linear approximate approach is suggested based on the so called Mixed Assessor Model (MAM):

$$\begin{aligned} y_{ijk} &= \mu + a_i + \nu_j + \beta_i x_j + d_{ij} + \epsilon_{ijk}, \\ a_i &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_a^2), \quad d_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \\ \epsilon_{ijk} &\sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \end{aligned} \tag{3}$$

where  $x_j = \bar{y}_{.j} - \bar{y}_{...}$  is the centered product averages for product  $j$ , included as a covariate, and the regression coefficient,  $\beta_i$ , is the individual scaling slope for assessor  $i$ , with  $\sum_{i=1}^I \beta_i = 0$ . The MAM is a simple linear mixed model since the  $x_j$ s are found directly from data before estimating the model and are therefore considered known. This is justified by the fact that this model still produces valid hypothesis tests for product differences. However, this model will in general not produce valid product difference confidence intervals, since the  $\beta$ s are considered fixed instead of random, even though they are assessor dependent.

The formal multiplicative mixed model can be written as (2), where  $a_i$  is the main effect of assessor,  $\nu_j$  is the product effect,  $b_i$  is the assessor specific scaling effect and  $d_{ij}$  is the panelist-by-product disagreement effect. By including the multiplicative term, we perform a regression for each assessor against the centered product effects - the larger the range of scale use, the larger the regression coefficient ( $1+b_i$ ). It is assumed that observations across assessors are independent and that  $\epsilon_{ijk}$  is independent of the random effects. Under this model, profile likelihood based confidence intervals for the product differences can easily be obtained with R-package `mumm`, together with the model parameter estimates.

### 3.1.1 Data Example: B&O TV Data

The example data set is acquired from the `lmerTest`-package in R (Kuznetsova et al, 2016a, 2017) and it stems from a sensory evaluation of a series of Bang & Olufsen televisions. The televisions are assessed based on 15 characteristics such as *colour balance*, *colour saturation*, *noise*, *sharpness*, *cutting* etc., but in this example we restrict ourselves to one

single characteristic - namely *cutting*. The term cutting refers to how much of the picture that has been cut off on the screen, and the scale goes from nothing to much.

The purpose is to test the products, which are specified by two attributes *TVset* and *Picture* having three and four levels respectively. These attributes are crossed, which gives us one combined product factor with 12 levels ( $J = 12$ ).

The 12 products were assessed by a panel of eight assessors ( $I = 8$ ), scoring each product in two replicates ( $K = 2$ ) - yielding a total of 192 measurements. It is assumed that the replications are completely randomized, hence we assume that no block effect is present.

Using R-package `mumm`, the multiplicative mixed model is fitted to the data by the following code, where `mp(Assessor, Product)` is the syntax for the multiplicative term.

```
fit = mumm(Cutting ~ 1 + Product + (1|Assessor) +
           (1|Assessor:Product) + mp(Assessor, Product), data = B0)
```

Appendix A contains the full code used to analyse this data set.

The parameter estimates and the BLUPs are shown in Table 1, where it is seen that assessor 7 have the largest scaling coefficient ( $b_7$ ) and that assessor 2 has the second most negative scaling coefficient. These coefficients can be interpreted as the deviation from the consensus slope equal to one. The fitted regression lines for these two assessors versus the consensus product pattern are plotted together with the corresponding data points in Figure 1. The slopes of the lines are  $b_i + 1$  and the intercepts are  $a_i - \mu b_i$ . These lines clearly illustrate the scale range differences between the two assessors. Figure 2 shows a plot of the fitted regression lines for all of the eight assessors, where the difference in the slope of the lines is very clear, which justifies the use of the multiplicative model.

Table 2 shows the results of likelihood ratio tests of the significance of the models terms. Following Self and Liang (1987) the test statistic, when testing the significance of a variance parameter, follows an equal mixture of a  $\chi_0^2$ - and  $\chi_1^2$ -distribution, approximately. Therefore the degrees of freedom in the test is chosen to be 1/2. Similarly, when testing the significance of a variance parameter and a covariance parameter at the same time, the degrees

Table 1: The estimated model parameters and the BLUPs.

$\mu + \nu_1$	$\mu + \nu_2$	$\mu + \nu_3$	$\mu + \nu_4$	$\mu + \nu_5$	$\mu + \nu_6$
7.1057	8.5980	7.6681	6.9428	6.5361	6.7693
$\mu + \nu_7$	$\mu + \nu_8$	$\mu + \nu_9$	$\mu + \nu_{10}$	$\mu + \nu_{11}$	$\mu + \nu_{12}$
5.6357	6.7778	4.3898	4.2358	4.0981	4.0615

$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$
-2.9305	-1.4115	1.2018	-2.4946	3.2196	1.1976	-0.0986	1.3161

$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$
-0.2183	-0.4802	0.4141	-0.6071	-0.0742	0.0569	0.8508	0.0581

$\sigma$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\rho$
1.2100	2.0096	0.4692	0.2428	0.4188

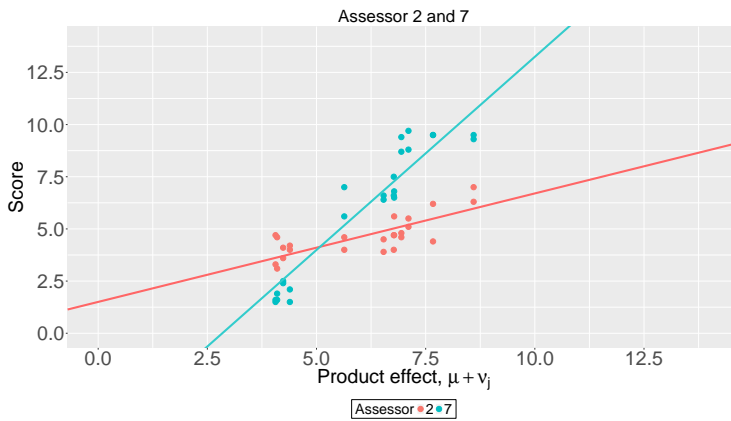


Figure 1: The data points and the fitted regression lines for Assessor 2 and 7 in the B&amp;O TV data set.

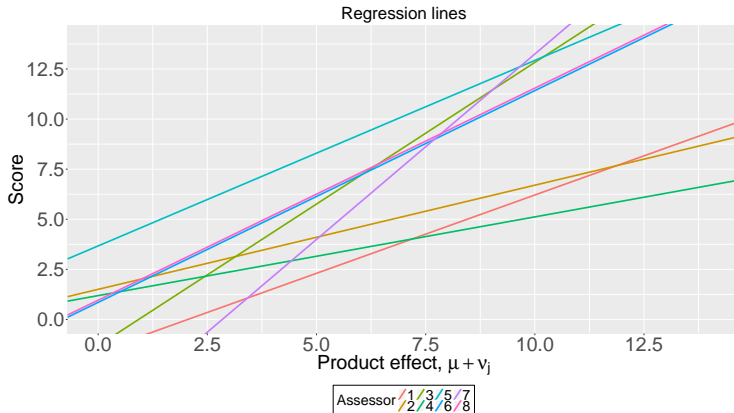


Figure 2: The fitted regression lines for the B&O TV data set. The slopes of the lines are  $b_i + 1$  and the intercepts are  $a_i - \mu b_i$ .

of freedom in the test is chosen to be  $3/2$ , since Stram and Lee (1994) argues that the test statistic in such a test approximately follows an equal mixture of a  $\chi_1^2$ - and  $\chi_2^2$ -distribution. Not surprisingly, it is seen that the scaling effect *is* significant. The disagreement effect is not significant, meaning that there is no significant difference in the assessors perception of the products; the entire panelist-by-product interaction can be explained by the assessors individual ranges of scale use.

### Product Differences

The variance of the product contrast in the multiplicative mixed model given by (2) is

$$\text{var}(\bar{y}_{.1.} - \bar{y}_{.2.}) = \frac{1}{I} \sigma_b^2 (\nu_1 - \nu_2)^2 + \frac{2}{I} \sigma_d^2 + \frac{2}{KI} \sigma^2, \quad (4)$$

whereas the variance in MAM given by (3) is

$$\text{var}(\bar{y}_{.1.} - \bar{y}_{.2.}) = \frac{2}{I} \sigma_d^2 + \frac{2}{KI} \sigma^2 \quad (5)$$

(Brockhoff et al, 2015)

We see that (4) has an extra term that makes the variance of the product difference depend on the actual size of the difference itself. Under the null hypothesis of no product difference, however, the term disappears and the variance expressions for the two models

Table 2: Likelihood Ratio Test of the significance of the model terms.  $M_0$  and  $M_1$  are the null model and the alternative model, respectively.

Effect	$M_0$	$M_1$	$\chi^2$	DF	p-value
Disagreement, $d$	(2) without $d$	(2)	0.13	1/2	$4.49E - 01$
Scaling, $b$	2-way ANOVA	(2)	28.23	3/2	$3.07E - 07$
Assessor, $a$	(2) without $a$	(2)	133.77	3/2	$0.00E + 00$
Product, $\nu$	(2) without $\nu$ and $b$	(2)	120.29	12.5	$0.00E + 00$

are equal. This means that the null-hypothesis test under the MAM is also valid under the multiplicative mixed model.

In (5) the main part of error stems from the disagreement variation, and not from the interaction, which is the case in a mixed two-way ANOVA model. Therefore the F-statistic

$$F_{Product} = \frac{MS_{Product}}{MS_{Disagreement}}, \quad (6)$$

will follow an F-distribution with  $(J - 1, (I - 1)(J - 2))$  degrees of freedom under the null hypothesis and will be a valid test for product difference under the multiplicative mixed model.

Note that the denominator in the F-statistic is  $MS_{Disagreement}$  and not  $MS_{Interaction}$  as in the mixed two-way ANOVA model, which means that the error used for making inference about product differences in (6) is now cleaned out for potential scaling structure. This results in greater power for detecting product differences compared to the standard mixed two-way model, which will be illustrated in the following example.

To see the actual effect of the increased power, the same data set as before will be used but this time with the characteristic *Sharpness of movement* as the response variable. The term *Sharpness of movement* refers to how sharp the picture is during movement or panning and the scale goes from sharp to unsharp.

Table 3 shows the computed p-values for testing the significance of the product effect under the three different models. In the likelihood ratio test the null model is the same as

Table 3: Computed p-values for testing the significance of an overall product difference with *Sharpness of movement* as the response variable.

Method	Model	Effect	F	$\chi^2$	DF	p-value
F-test	2-way ANOVA	Product	3.74	-	11	$2.67E - 4$
F-test <sup>a</sup>	MAM (3)	Product	3.86	-	11	$2.28E - 4$
LRT	MMM (2)	Product	-	38.60	12.5	$1.70E - 4$

<sup>a</sup> The results are found by R-package `SensMixed` (Kuznetsova et al, 2016b).

in Table 2, i.e.

$$\begin{aligned}
 M_0 : \quad & y_{ijk} = \mu + a_i + d_{ij} + \epsilon_{ijk}, \\
 & a_i \sim \text{i.i.d. } \mathcal{N}(0, \sigma_a^2), \quad d_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_d^2), \\
 & \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),
 \end{aligned}$$

and the alternative model is the full multiplicative mixed model (2).

It is seen that both the likelihood ratio test and the F-test under MAM result in a lower p-value compared to the F-test under the mixed two-way model, indicating an increased power for detecting product differences.

### 3.2 Genotype-by-environment Data

Genotype-by-environment data is frequently analysed by a two-way ANOVA model (see, e.g., Malosetti et al, 2014; van Eeuwijk et al, 2005; Piepho, 1997; Perkins and Jinks, 1968), which contains the overall mean of the phenotypic response, the effect of the genotype, the effect of the environment, the genotype-by-environment interaction and the random residual error.

It might though result in a more correct inference to describe the interaction more detailed, and take into consideration that some genotypes are more sensitive to the environment than

others. In Piepho (1997) a multiplicative model is proposed for the analysis of this type of data to account for this and in Malosetti et al (2014) it is argued that genotypes, depending on the aim of the analysis, can be regarded as a random sample from a larger population.

This leads us to the multiplicative mixed model in (2), where a part of the genotype-by-environment interaction is modelled by a genotype-specific linear regression on the environmental effects to account for the varying sensitivities of the different genotypes. In this setting,  $a_i$  and  $b_i$  is the main effect and the sensitivity of genotype  $i$ , respectively. Further,  $\nu_j$  is the fixed main effect of environment  $j$  and  $d_{ij}$  is the part of the interaction that is not explained by the linear regression.

### 3.2.1 Data Example: The Height of Wheat

A data set from the international maize and wheat improvement center (Global Wheat Program et al, 2017) will now be analysed by the multiplicative mixed model in (2). This data set contains the measured height of 50 wheat plants with unique genotypes grown in 45 different environments and there are two replicates. Hence, the phenotypic response in the model is in this case the height of the plant.

The multiplicative mixed model is fitted to the data and the variance component estimates are shown in Table 4. The data points for two of the 50 genotypes are plotted together with the corresponding fitted regression lines in Figure 3, where it is seems evident that the two wheat genotypes have different sensitivities to the environment.

Table 4: The estimated variance components.

$\sigma$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\rho$
5.1958	2.4242	0.0691	3.1660	0.5844

Table 5 shows the results of likelihood ratio tests of the significance of the terms, where it seen that all of the model terms are significant, including the the sensitivity effect.



Figure 3: The fitted regression lines and the data points for genotype 1 and 2. The slopes of the lines are  $b_i + 1$  and the intercepts are  $a_i - \mu b_i$ .

Table 5: Likelihood Ratio Test of the significance of the model terms.

Effect	$\chi^2$	DF	p-value
Unexplained interaction, $d$	168.10	1/2	0.00E + 00
Sensitivity, $b$	50.86	3/2	3.25E - 12
Genotype, $a$	386.82	3/2	0.00E + 00
Environment, $\nu$	5039.37	45.5	0.00E + 00

### 3.3 Method Comparison Studies - agreement between random methods

Method comparison studies are studies that are designed to compare different medical instruments or different methods of clinical measurement. The usual approach when comparing multiple measurement methods is to apply the two-way ANOVA model (Hawkins and Sharma, 2010):

$$y_{ij} = \mu_j + \alpha_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2), \quad (7)$$



where  $y_{ij}$  is the measured value obtained from using method  $i$  on item  $j$ ,  $\mu_j$  is the true value for item  $j$  and  $\alpha_i$  is the relative bias related to method  $i$ . This is also the underlying model in the so-called Bland-Altman setup, which is often used when only two measurement methods are compared and only one measurement by each method is carried out on the items (Carstensen, 2004). However, if replicate measurements are available, the model can easily be expanded with an interaction term that separates the item-method interaction from the error term. In the rest of this section we will, however, only consider data sets without replicates.

Model (7) is not capable of handling situations where the methods do not have the same linear calibration. In such cases, parts of the interaction can be modelled by regressing on the item values (Carstensen, 2004; Hawkins and Sharma, 2010). The model is therefore expanded to:

$$y_{ij} = \mu_j + \alpha_i + \beta_i \mu_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),$$

and where the  $\beta_i$ 's are the method specific regression coefficients.

In some experiments, the used measurement methods can be considered a random sample from a larger population of possible measurement methods. In Ekstrøm and Carstensen (2017) such a situation is considered. In that case, it is more reasonable to consider the method-specific effects as random and thus to fit the following multiplicative mixed model:

$$y_{ij} = \mu_j + \tilde{a}_i + b_i \mu_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),$$

$$(\tilde{a}_i, b_i) \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} \sigma_{\tilde{a}}^2 & \tilde{\rho} \sigma_{\tilde{a}} \sigma_b \\ \tilde{\rho} \sigma_{\tilde{a}} \sigma_b & \sigma_b^2 \end{bmatrix} \right),$$

where the method dependent parameters are now written in Latin letters to clarify that they are considered random.

This multiplicative model is written without a term accounting for the overall mean, but when we allow the method specific random effects,  $a_i, b_i$ , to correlate, this model is similar

to (2), if we discard  $d_{ij}$  and index  $k$ , due to the lack of replicate measurements. In this setting,  $\mu_j = \mu + \nu_j$  and  $\tilde{a}_i = a_i - b_i\mu$ .

### 3.3.1 Data Examples

#### Example 1

In the first data set (kindly supplied by Prof. Douglas Hawkins) (Hawkins and Sharma, 2010), six methods for measuring the concentration of hepatitis virus in blood samples are compared. The number of blood samples (items) in the experiment is 51. The experiment is not replicated and the data is incomplete, resulting in only 248 observations instead of 306.

The multiplicative mixed model in (2) is fitted to the data, and Table 6 shows the estimated variance components in the model and Table 7 shows the estimated method specific random coefficients.

Table 6: The estimated variance components.

$\sigma$	$\sigma_a$	$\sigma_b$	$\rho$
0.2092	0.1248	0.0275	-0.5349

Table 7: The estimated random coefficients.

$i$	capsl1	capsl2	capsld	cobas	hps	versant
$a_i$	-0.1107	0.0358	-0.1044	-0.1220	0.1146	0.1868
$b_i$	0.0047	-0.0244	-0.0018	0.0492	-0.0158	-0.0120

We see that the *cobas* method has the steepest regression line and *capsl2* the flattest. Figure 4 shows the fitted regression lines for these two methods. Table 8 shows the likelihood ratio tests of the significance of the model terms, where it is seen that everything is significant.

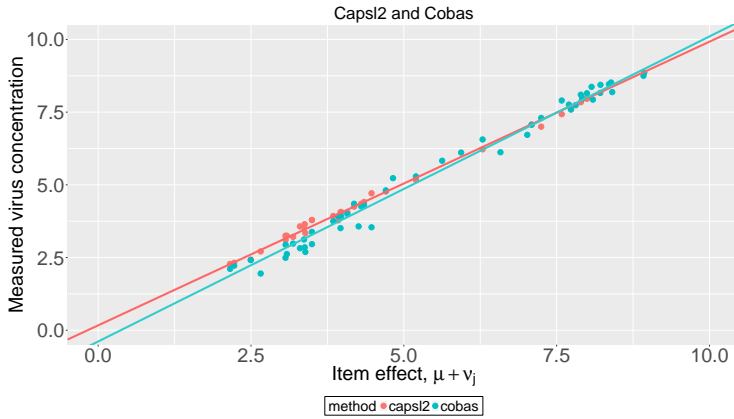


Figure 4: The fitted regression lines and the data points for method Capsl2 and Cobas. The slopes of the lines are  $b_i + 1$  and the intercepts are  $a_i - \mu b_i$ .

Table 8: Likelihood Ratio Test of the significance of the model terms.

Effect	$\chi^2$	DF	p-value
Scaling, $b$	11.01	3/2	$2.08E - 03$
Method, $a$	62.01	3/2	$1.18E - 14$
Item, $\nu$	1100.04	51.5	$0.00E + 00$

### Example 2

The second dataset `glucose` is from the R-package `MethComp` (Carstensen et al, 2015). The data is unbalanced and consists of data from 74 persons (items) and four method types. Each person had blood sampled at 0, 30, 60 and 120 min after a 75 g glucose load and the glucose concentration was measured, but we only consider the data sampled after 120 min, with 328 observations.

Table 9 shows the estimated variance components for the multiplicative mixed model fit where the correlation between the random effects seems to be quite strong. Table 10 shows the estimated method specific random parameters, where it is seen that the slope differs the most between *plasma* and *capil*. This difference is quite clear in Figure 5, showing the

data points and the four regression lines.

Table 9: The estimated variance components.

$\sigma$	$\sigma_a$	$\sigma_b$	$\rho$
0.4363	0.4698	0.1156	0.7381

Table 10: The estimated random coefficients.

$i$	blood	plasma	serum	capil
$a_i$	-0.7530	0.4742	0.2704	0.0084
$b_i$	-0.1007	0.1335	0.0829	-0.1157

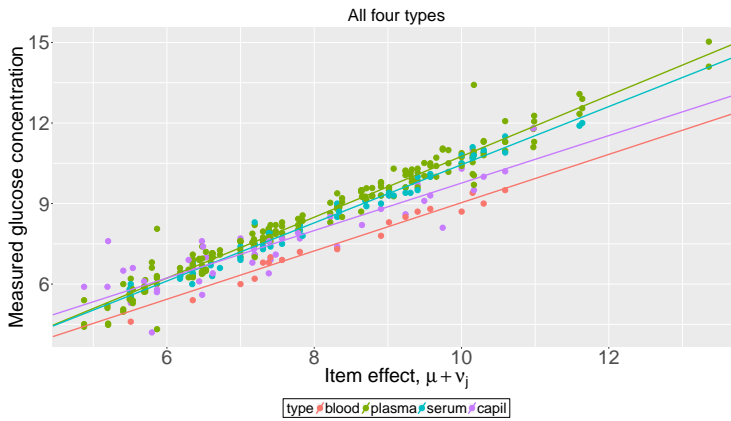


Figure 5: The fitted regression lines and the data points for the glucose data set. The slopes of the lines are  $b_i + 1$  and the intercepts are  $a_i - \mu b_i$

In Table 11 we see again that everything is significant.

Table 11: Likelihood Ratio Test of the significance of the model terms.

Effect	$\chi^2$	DF	p-value
Scaling, $b$	29.02	3/2	$2.06E - 07$
Method, $a$	113.49	3/2	$0.00E + 00$
Item, $\nu$	951.91	74.5	$0.00E + 00$

### 3.3.2 Limits of Agreement

Agreement between methods are often assessed by estimating the limits of agreement (LoA), which give a prediction interval for the difference between measurements from two different methods on the same item. These LoAs depend on the model used and in this section it is emphasized that the intervals might become too narrow for some values of the item effect, if the multiplicative term is ignored.

If the simple model (7) is assumed to hold, the limits of agreement are

$$\begin{aligned}
 \text{LoA} &= \alpha_i - \alpha_{i'} \pm z \sqrt{\text{var}(y_{ij} - y_{i'j})} \\
 &= \alpha_i - \alpha_{i'} \pm z \sqrt{\text{var}(y_{ij}) + \text{var}(y_{i'j})} \\
 &= \alpha_i - \alpha_{i'} \pm z \cdot \sqrt{2\sigma^2},
 \end{aligned}$$

where  $z$  is the quantile in the standard normal distribution for the desired level of the prediction interval.

In Ekstrøm and Carstensen (2017), they consider the effect of method as random, as mentioned above. Hence they consider a mixed two-way ANOVA model, and they thus get the following limits of agreement, which is now the prediction interval for the difference

between measurements from two randomly chosen methods on a new item

$$\begin{aligned}
 \text{LoA} &= 0 \pm z \sqrt{\text{var}(y_{ij} - y_{i'j})} \\
 &= 0 \pm z \sqrt{\text{var}(y_{ij}) + \text{var}(y_{i'j})} \\
 &= 0 \pm z \cdot \sqrt{2 \cdot (\sigma_a^2 + \sigma^2)}
 \end{aligned} \tag{8}$$

This results in a wider prediction interval, since the variation between the random methods is taken into account.

If the multiplicative mixed model is assumed to hold, the limits of agreement for the difference between two random methods on a new item is:

$$\begin{aligned}
 \text{LoA} &= 0 \pm z \sqrt{\text{var}(y_{ij} - y_{i'j})} \\
 &= 0 \pm z \sqrt{\text{var}(y_{ij}) + \text{var}(y_{i'j})} \\
 &= 0 \pm 1.96 \cdot \sqrt{2 \cdot (\sigma_a^2 + \nu_j^2 \sigma_b^2 + 2 \nu_j \rho \sigma_a \sigma_b + \sigma^2)},
 \end{aligned} \tag{9}$$

which gives us "trumpet shaped" prediction intervals.

### Simulation Study

The difference between the LoA in (8) and (9) is illustrated by a small simulation study, which simulates the following scenario: For each of 120 patients two random methods are used to measure the glucose concentration in their blood. The measurements are simulated according to the multiplicative mixed model, where the variance components are set equal to the estimates from Example 2. The method specific random effects are drawn from a multivariate normal distribution by the function `mvrnorm` from R-package `MASS` (Venables and Ripley, 2002). The patients' "true" concentration ranges from 0.1 to 12.

Figure 6 shows the simulated data plotted together with the estimated limits of agreement, found by using (8) and (9) on the data set from Example 2. It is clearly seen that the limits are too narrow when using (8) for large item effects. It is also seen that for some values of the item effect, the limits are actually vaguely more narrow when using (9) instead of (8).

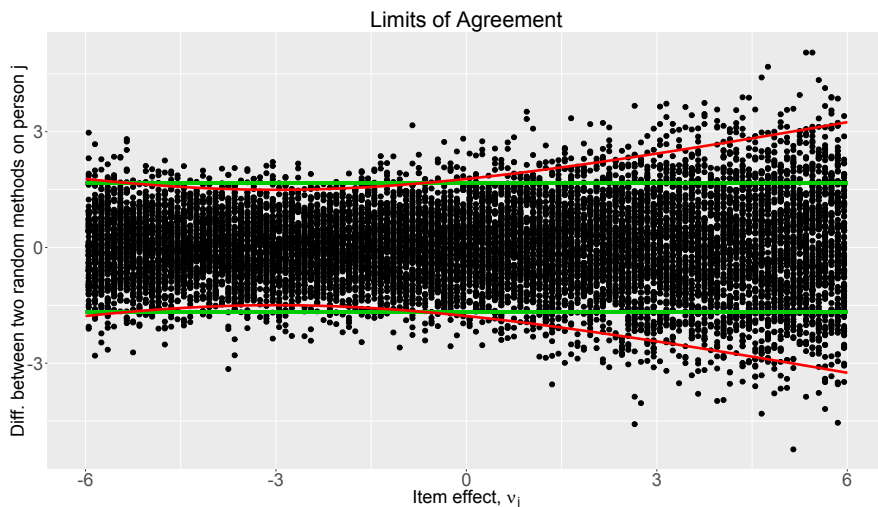


Figure 6: "Bland-Altman" plot for simulated data and the estimated limits of agreement. We have 120 patients and 100 simulated differences for each. The green and the red lines are the limits of agreement calculated from equation (8) and equation (9), respectively.

## 4 R-package mumm vs. NLMIXED in SAS

The multiplicative mixed model can also be fitted by the NLMIXED procedure in SAS (Littell et al, 2007). This might however be way more time consuming than fitting the model by `mumm`, if the model contains a lot of parameters. Table 12 shows the computation time for the two methods when fitting the multiplicative mixed mode to three different data sets. It should be noted that PROC NLMIXED is run with the statement options: `method=fir` `technique=nrridg`. The computer used has 32 GB RAM, an Intel Core i7-4790 processor and runs under the operating system Windows 8.1 Enterprise. The version of R is 3.1.3, and the SAS version is 9.4.

It is seen that R-package `mumm` has a lower computation time, and that the difference gets bigger when the number of parameters in the model increases or when the starting guess is far away from the optimum. Further it is seen that NLMIXED in one case did not succeed in finding the optimum; it stopped with the message "NRRIDG Optimization cannot be completed. Optimization routine cannot improve the function value." It is also

Table 12: Computation time for mumm and NLMIXED.

Data set ( $N_{fixed}, N_{random}$ ) <sup>a</sup>	Starting guess						CPU time (sec.)	
	$\mu + \nu_j$	$\sigma_a$	$\sigma_b$	$\sigma_d$	$\sigma$	$\rho$	mumm	NLMIXED
B&O TV (12,8)	0	1	1	1	1	0	<b>0.50</b>	<b>0.68</b>
B&O TV (12,8)	5	0.1	1	0.5	1	0.2	<b>0.38</b>	<b>0.48</b>
Wheat height (45,50)	0	1	1	-	1	0	<b>9.39</b>	— <sup>c</sup>
Wheat height (45,50)	40	1	0.08	-	4	0	<b>3.29</b>	<b>155.07</b>
Wheat height (45,50)	89	2	0.05	-	6	0.3	<b>1.41</b>	<b>47.87</b>
Hepatitis (51,6)	0	1	1	-	1	0	<b>1.34</b>	<b>64.71</b>
Hepatitis (51,6)	8/4 <sup>b</sup>	0.1	0.05	-	0.2	-0.5	<b>0.83</b>	<b>12.30</b>

<sup>a</sup>  $N_{fixed}, N_{random}$  : the number of levels of the fixed effect and the random effect.

<sup>b</sup> The starting guess is 8 for  $j = 1, \dots, 18$  and 4 for  $j = 19, \dots, 51$ .

<sup>c</sup> SAS stops with the error "Optimization cannot be completed."

worth noting that the "Wheat height" data set is fitted by a multiplicative mixed model without the interaction term,  $d_{ij}$ , even though replicates are present in the data. This is due to the fact that multilevel nonlinear mixed models are not accommodated by the NLMIXED procedure (SAS Institute Inc., 2009). Certain nested random effect structures can though be specified, meaning that model (2), including the random interaction term  $d_{ij}$ , can be fitted by NLMIXED (Littell et al, 2006). However, if the number of levels of the fixed effect is large, the implementation becomes unreasonably burdensome.

#### 4.1 Confidence Intervals for Fixed Effect Contrasts

In equation (4) in Section 3.1.1 it was shown that the variance of the product difference depends on the actual size of the difference itself, with increasing variance for increasing contrast. This means that proper confidence intervals for product contrasts are asymmetric; they are wider "away from zero" than "towards zero" (Brockhoff et al, 2015). This result is obviously valid in the other applications as well. However, the NLMIXED procedure only provides symmetrical t-distribution based confidence intervals, whereas R-package



---

`mumm` makes it possible to obtain profile likelihood based confidence intervals, which will be asymmetric due to the nature of the model.

## 5 Concluding Remarks

In every situation where an interaction between a fixed effect and a random effect is present, improved inference might be possible by extending the usual linear mixed model with a multiplicative term. The purpose of this term is to obtain a more nuanced modelling and interpretation of the interaction, leading to a more correct inference of the model effects. The version of the multiplicative mixed model studied in this paper was shown to have multiple applications, and with the newly developed R-package `mumm`, this non-linear mixed model is easily fitted to data.

## 6 Conflict of Interest

The authors declare that they have no conflict of interest.

## A R-code for B&O TV Data

Listing 1: R-code used to produce the results in Section 3.1.1.

```

library(SensMixed); library(lmerTest); library(mumm)

# ..... Loading The BO TV Data .....
DATA = TVbo
DATA$Product = as.factor(DATA$TVset:DATA$Picture)

BO = data.frame(DATA$Product)
names(BO) = "Product"
BO$Assessor = DATA$Assessor
BO$Cutting = DATA$Cutting
BO$Sharpmove = DATA$Sharpnessofmovement

# ..... Fitting the multiplicative mixed model with mumm .....
fit = mumm(Cutting ~ 1 + Product + (1|Assessor) +
           (1|Assessor:Product) + mp(Assessor,Product), data = BO)

a = mumm::ranef(fit)$Assessor
b = mumm::ranef(fit)$`mp Assessor:Product`

# ..... Likelihood ratio-tests .....
fit_no_d = mumm(Cutting ~ 1 + Product + (1|Assessor) +
                mp(Assessor,Product), data = BO)
fit_no_a = mumm(Cutting ~ 1 + Product + (1|Assessor:Product) +
                mp(Assessor,Product), data = BO)
fit_no_b = lmer(Cutting ~ 1 + Product + (1|Assessor) +
                (1|Assessor:Product), data = BO, REML = FALSE)
fit_no_v = lmer(Cutting ~ 1 + (1|Assessor) + (1|Assessor:Product),

```

```

        data = B0, REML = FALSE)

#The chi^2 values:
X2 = 2*c(-fit$objective-(-fit_no_d$objective), # Test of disagreement.
        -fit$objective-logLik(fit_no_b),      # Test of scaling.
        -fit$objective-(-fit_no_a$objective), # Test of assessor.
        -fit$objective-logLik(fit_no_v)      # Test of product.
        );

pvalues = 1-pchisq(X2, df = c(1/2,3/2,3/2,12.5));

# ..... Detecting product differences.....
# Linear 2-way mixed model:
fit_2way = lmer(Sharpmove ~ 1 + Product + (1|Assessor) +
               (1|Assessor:Product), data = B0, REML = TRUE)
anova(fit_2way)

# Mixed Assessor Model:
fit_MAM = sensmixed(c("Sharpmove","Sharpmove"),
                   c("Product"),c("Assessor"),
                   data = B0, product_structure = 1,
                   error_structure = "ONLY-ASS", MAM = TRUE,
                   control = sensmixedControl(calc_post_hoc = TRUE,
                                               MAM_balanced = TRUE,
                                               MAM_adjusted = FALSE))

fit_MAM[[3]]

# Multiplicative mixed model:
fit2 = mumm(Sharpmove ~ 1 + Product + (1|Assessor) +
           (1|Assessor:Product) + mp(Assessor,Product), data = B0)
fit2_nov = lmer(Sharpmove ~ 1 + (1|Assessor) + (1|Assessor:Product),

```

```

      data = B0, REML = FALSE)
X2 = 2*(-fit2$objective-logLik(fit2_nov))
1-pchisq(X2, df = 12.5);

```

## References

- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48, DOI 10.18637/jss.v067.i01
- Brockhoff PB, Schlich P, Skovgaard I (2015) Taking individual scaling differences into account by analyzing profile data with the mixed assessor model. *Food Quality and Preference* 39:156–166
- Brockhoff PM, Skovgaard IM (1994) Modelling individual differences between assessors in sensory evaluations. *Food quality and preference* 5(3):215–224
- Carstensen B (2004) Comparing and predicting between several methods of measurement. *Biostatistics* 5(3):399–413
- Carstensen B, Gurrin L, Ekstrom C, Figurski M (2015) MethComp: Functions for Analysis of Agreement in Method Comparison Studies. URL <https://CRAN.R-project.org/package=MethComp>, r package version 1.22.2
- van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional anova models to eco-physiological qtl models. *Crop and Pasture Science* 56(9):883–894
- Ekstrøm CT, Carstensen B (2017) Statistical models for assessing agreement in method comparison studies with heterogeneous random raters and replicate measurements. *Annals of Applied Statistics* In press
- Global Wheat Program, IWIN Collaborators, Singh R, Payne T (2017) 18th semi-arid wheat yield trial. DOI 11529/10170, URL <http://hdl.handle.net/11529/10170>

- Hawkins DM, Sharma A (2010) Comparison of measurements by multiple methods or instruments. *Journal of biopharmaceutical statistics* 20(5):965–984
- Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM (2016) TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software* 70(5):1–21, DOI 10.18637/jss.v070.i05
- Kuznetsova A, Brockhoff PB, Christensen RHB (2016a) lmerTest: Tests in Linear Mixed Effects Models. URL <https://CRAN.R-project.org/package=lmerTest>, r package version 2.0-32
- Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R (2016b) SensMixed: Analysis of Sensory and Consumer Data in a Mixed Model Framework. URL <http://CRAN.R-project.org/package=SensMixed>, r package version 2.0-9
- Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13)
- Littell RC, Stroup WW, Milliken GA, Wolfinger RD, Schabenberger O (2006) SAS for mixed models, SAS institute, chap 15.4
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2007) SAS for mixed models. SAS institute
- Malosetti M, Ribaut JM, van Eeuwijk FA (2014) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Drought phenotyping in crops: From theory to practice* 4(44):53
- Perkins JM, Jinks J (1968) Environmental and genotype-environmental components of variability. *Heredity* 23(3):339–356
- Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* pp 761–766
- Piepho HP (1999) Fitting a regression model for genotype-by-environment data on heading dates in grasses by methods for nonlinear mixed models. *Biometrics* 55(4):1120–1128

- 
- Piepho HP, Madden LV, Williams ER (2015) Multiplicative interaction in network meta-analysis. *Statistics in medicine* 34(4):582–594
- Pødenphant S, Brockhoff PB (2016) mumm: Multiplicative Mixed Models using Template Model Builder. URL <https://CRAN.R-project.org/package=mumm>, r package version 0.2.0
- R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- SAS Institute Inc (2009) SAS/STAT 9.2 user’s guide, second edition. Cary, NC: SAS Institute Inc pp 4378–4379
- Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82(398):605–610
- Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57(4):1138–1147
- Smith A, Cullis B, Brockhoff P, Thompson R (2003) Multiplicative mixed models for the analysis of sensory evaluation data. *Food quality and preference* 14(5):387–395
- Smith A, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science* 143(6):449–462
- Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* pp 1171–1177
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York, URL <http://www.stats.ox.ac.uk/pub/MASS4>, ISBN 0-387-95457-0



APPENDIX B

The Mixed Assessor Model  
and the multiplicative  
mixed model (MAM and  
mumm)

---

**Pødenphant, S.**, Truong, M. H., Kristensen, K., & Brockhoff, P. B. (2018).  
*The Mixed Assessor Model and the multiplicative mixed model (MAM and mumm)*.  
Submitted to Food Quality and Preference (under review).



## The Mixed Assessor Model and the multiplicative mixed model (MAM and mumm)

Sofie Pødenphant<sup>a,\*</sup>, Minh H. Truong<sup>a</sup>, Kasper Kristensen<sup>b</sup>, Per B. Brockhoff<sup>a</sup>

<sup>a</sup>*DTU Compute, Section for Statistics and Data Analysis, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kongens Lyngby, Denmark*

<sup>b</sup>*DTU Aqua, Section for Marine Living Resources, Technical University of Denmark, Kemitorvet, Building 202, DK-2800 Kongens Lyngby, Denmark.*

---

### Abstract

A novel possibility for easy and open source based analysis of sensory profile data by a formal multiplicative mixed model with fixed product effects and random assessor effects is presented by means of the generic statistical R-package *mumm*. The package is using likelihood principles and is utilizing newer developments within Automatic Differentiation by means of the Template Model Builder R-package. We compare such formal likelihood based analysis with the Mixed Assessor Model (MAM) analysis suggested by Brockhoff et al. (2015). We use real sensory data as examples together with simulated data. We found that the formal *mumm* approach for hypothesis testing more resembles the MAM than the standard 2-way mixed model, and that both the *mumm* approach and the MAM give a higher power to detect product differences than the 2-way mixed model, when a "scaling effect" is present. We also validated that the novel contrast confidence limit method suggested in Brockhoff et al. (2015) performs well and in line with the formal likelihood based confidence intervals of the *mumm*. Finally, the likelihood based *mumm* approach suggests that the more proper test for product difference would be a test that has a "joint product and scaling effect" interpretation.

*Keywords:* Sensory profile data, Analysis of variance, Multiplicative mixed

---

\*Corresponding author  
Email address: [sofp@dtu.dk](mailto:sofp@dtu.dk) (Sofie Pødenphant)

## 1. Introduction

Sensory profile data, where  $I$  assessors scored  $J$  products in  $K$  replications, is frequently analysed by a 2-way mixed analysis of variance (ANOVA) corresponding to the following model

$$\begin{aligned} Y_{ijk} &= \mu + a_i + \nu_j + g_{ij} + \epsilon_{ijk} \\ a_i &\sim N(0, \sigma_{PAN}^2), g_{ij} \sim N(0, \sigma_g^2), \epsilon_{ijk} \sim N(0, \sigma^2), \end{aligned} \quad (1)$$

where  $a_i$  is the assessor main effect,  $i = 1, \dots, I$ ,  $\nu_j$  is the product main effect,  $j = 1, \dots, J$ ,  $g_{ij}$  is the assessor-by-product interaction and  $\epsilon_{ijk}$ ,  $k = 1, \dots, K$ , is the random residual error.

However, the assessor-by-product interaction will often not only consist of real deviations in perception of product differences (disagreement effect), but also of scale range differences between assessors (scale effect). Scale range differences appear when some assessors use a larger part of the scale than others, when scoring the products. In Brockhoff et al. (2015) a meta study of 8619 attributes from 369 profile data sets showed that such scaling heterogeneity was significantly present in 45% of all the attributes. Thus, it will not be valid in general to assume that scale range differences are not present. To account for the scale effect, Brockhoff et al. (2015) uses the multiplicative model approach suggested in Brockhoff & Skovgaard (1994) and combines it with the general mixed model approach. As a result, the Mixed Assessor Model (MAM) was introduced:

$$\begin{aligned} Y_{ijk} &= \mu + a_i + \nu_j + \beta_i x_j + d_{ij} + \epsilon_{ijk} \\ a_i &\sim N(0, \sigma_{PAN}^2), d_{ij} \sim N(0, \sigma_D^2), \epsilon_{ijk} \sim N(0, \sigma^2), \end{aligned} \quad (2)$$

where  $x_j = \bar{y}_{.j} - \bar{y}_{...}$  are the centered product averages inserted as a covariate, implying that the  $\beta$ s are the individual scaling slopes; the bigger the scale range, the larger the slope. The purpose of including the term  $\beta_i x_j$  is to model

the assessors' individual ranges of scale use, such that the interaction,  $d_{ij}$ , captures the disagreement effect and not the scale effect. Consequently, the scaling heterogeneity between the assessors is removed from the assessor-by-product interaction, when used for hypothesis testing of product differences. Since the  $x_{js}$  are calculated directly from data, MAM is a linear mixed model and should be seen as an approximation of the following more properly specified mixed model (Brockhoff et al., 2015)

$$\begin{aligned}
 Y_{ijk} &= \mu + a_i + \nu_j + b_i \nu_j + d_{ij} + \epsilon_{ijk} & (3) \\
 (a_i, b_i) &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_{PAN}^2 & \rho \sigma_{PAN} \sigma_{SCALE} \\ \rho \sigma_{PAN} \sigma_{SCALE} & \sigma_{SCALE}^2 \end{bmatrix}\right), \\
 d_{ij} &\sim N(0, \sigma_D^2), \epsilon_{ijk} \sim N(0, \sigma^2),
 \end{aligned}$$

5 where the covariate is the true product effects,  $\nu_j$ , and where the scaling slopes,  $b_i$ , are modelled as random effects. It is worth noting that in Brockhoff et al. (2015), the model is written without the correlation between  $a_i$  and  $b_i$ . But the model is in that paper only used to express product contrast variances and these contrasts are not affected by this correlation, so everything in Brockhoff et al  
 10 (2015) is unchanged based on the model including the correlation stated here. However, since it impacts the optimization of the likelihood function for the model and the resulting parameters, we allow for a correlation between them. Due to the multiplicative term,  $b_i \nu_j$ , this model is a so-called multiplicative mixed model and does not belong to the class of linear mixed models. The  
 15 model is not straightforward to estimate, since the unknown product values,  $\nu_j$ , enter both the expectation and the variance structure. Therefore, the theory and the computations become simplified when using the MAM.

In Smith et al. (2003) a multiplicative mixed model is also applied for the analysis of sensory profile data, but since they use the model to examine asses-  
 20 sor performances, they assume assessor effects to be fixed and product effects random. Under that assumption, the model has a factor analytic covariance structure, which simplifies the estimation. The assumption of random product

---

effects and fixed assessor effects is, however, not reasonable in our case, since our main focus is on the comparison of specific products, and not on the particular  
25 assessors in the panel.

Even though the MAM is an approximation, Brockhoff et al. (2015) justify that the model produces valid hypothesis tests for overall product differences and also for post-hoc product difference testing. They further show that MAM increases the power to detect product differences, compared to a standard 2-  
30 way mixed model. The MAM is also used in Peltier et al. (2014) for monitoring assessor and panel performance, based partly on ideas from Brockhoff (2003). In (Brockhoff & Belmonte, 2018) a full overview of the use of the MAM is given, together with a demonstration of how to fit the model by the R-package *SensMixed*, (Kuznetsova et al., 2016b; R Core Team, 2017).

35 However, the MAM will, in general, fail to produce valid post-hoc product difference confidence intervals. Therefore Brockhoff et al. (2015) suggests a novel procedure to obtain appropriate product difference confidence intervals.

In this paper, we compare the hypothesis tests for overall product differences  
40 when using the MAM, the 2-way mixed model and the multiplicative mixed model, where the latter is fitted by the newly developed R-package *mumm* (Pødenphant & Brockhoff, 2016) by optimization of the likelihood function (section 2). Further, we investigate whether the suggested procedure in Brockhoff et al. (2015) actually does produce appropriate product difference confidence  
45 intervals. We will do this by using the procedure to estimate confidence intervals for simulated data sets and, thereafter, calculating coverage probabilities for the estimated confidence intervals. Additionally, we will use R-package *mumm* to find the profile likelihood based confidence intervals for the product differences for the multiplicative mixed model, and calculate the resulting coverage prob-  
50 abilities. Subsequently we will compare the performance of the full likelihood approach and the novel method suggested by Brockhoff et al. (2015) (section 3). In section 4, we propose a new test for product difference for the MAM. Finally, section 5 includes a summary, a discussion about computation time and some

final remarks.

## 55 2. Power to detect product differences

The hypothesis tests for overall product differences were conducted by F-tests for model (1) and (2) and by likelihood ratio tests for model (3). The product difference F-test based on the MAM differs from the F-test based on the 2-way mixed model by having  $MS_{Disagreement}$  instead of  $MS_{Interaction}$  as denominator in the F-statistic. This is the reason for the increased power to detect product differences (Brockhoff et al., 2015). The F-test based on the MAM was carried out by the use of R-package *SensMixed* (Kuznetsova et al., 2016b). The likelihood ratio test for product differences in (3) was performed by testing the reduced model with no product effect

$$Y_{ijk} = \mu + a_i + d_{ij} + \epsilon_{ijk} \quad (4)$$

$$a_i \sim N(0, \sigma_{PAN}^2), d_{ij} \sim N(0, \sigma_D^2), \epsilon_{ijk} \sim N(0, \sigma^2),$$

against the full model. The full model and its likelihood were found by using the R-package *mumm* (Pødenphant & Brockhoff, 2016), which is a wrapper of the Template Model Builder R-package (*TMB*) (Kristensen et al., 2016). *TMB* enables fast optimization of the Laplace approximation of the marginal log-likelihood function for the multiplicative mixed model. The "Laplace approximation" is a standard procedure within likelihood theory often used for nonlinear mixed models as a way to approximate the complicated likelihood function which otherwise is a complex multi-integral expression (Wolfinger, 1993; Vonesh, 1996). Fast optimization of the Laplace approximation is made possible with *TMB* through the use of Automatic Differentiation (AD). With this technique, *TMB* obtains the gradient, the Hessian, and the third order derivatives of the joint log-likelihood function with respect to the random effect coefficients, in a very time-efficient manner. These values are used to construct the Laplace approximation and its gradient, which are given as inputs to a standard minimizer in R to finally perform the maximization of the likelihood function.

The number of degrees of freedom in the likelihood ratio test is not obvious, due to the multiplicative term. When hypothesizing that the products are all equal, a  $J - 1$  degrees of freedom hypothesis, then the potential scaling differences also vanish from the model: When the products are all the same, there is no way that the assessors can range them differently. This is known in statistical likelihood theory as the problem of "nuisance parameters only present under the alternative". This kind of challenge has received attention over the last decades, see, e.g., Davies (1977); Ritz & Skovgaard (2005). The practical challenge is that one does not know the exact nor approximate distribution of the likelihood ratio test statistic. In sensory applications a similar situation and challenge is seen for the joint test of product difference in the corrected beta-binomial analysis of replicated difference testing data, cf. Brockhoff & Linander (2017). In other words, we do not know the best possible choice of number of degrees of freedom for the test.

However, we do know that, with a significance level at 5%, the false discovery rate should also be 5%. In the likelihood ratio test for product difference, we have therefore used the number of degrees of freedom that, in a simulation study under the null hypothesis, fulfills this criteria. In the simulation study we used 1000 data sets, simulated from the null model (4), where the parameters are set equal to the values we get from fitting (4) to the example data set in question.

### 2.1. Data example

We have analyzed the *TVbo* data set from the R-package *lmerTest* (Kuznetsova et al., 2016a, 2017). The data stems from sensory evaluations of Bang & Olufsen (BO) televisions, which were characterized by two design factors, *Picture* and *TVset*, with 4 and 3 levels, respectively. To compare the power of the methods in an illustrative way, a data set in which the product effect is on the boundary of significance is preferred. In this example, we have therefore allowed ourselves to analyze a subset of the data to lower the significance of the product effect. Thus, we have discarded all the observations for the two last levels of *Picture*,

such that this factor only has 2 levels. Afterwards, the two design factors were crossed, yielding 6 products in total. The products were evaluated by 8 assessors in 2 replications, yielding 16 observations per product, and the assessment was based on 15 attributes, of which we chose *Dim glass effect* as the response variable for this example. The *Dim glass effect* is big if parts of the picture seems dim/dull and the scale goes from none to a lot.

Table 1 shows the ANOVA table for this data set and Table 2 shows the result of the hypothesis tests for overall product differences. The approximate number of degrees of freedom in the likelihood ratio test is 9, which is found from the method described previously. In Table 1, we see that a significant assessor-by-product interaction is present, of which a big part is explained by the scaling effect, whereas the disagreement effect is non-significant. Observing the p-values for the product effect in Table 2, we see that significance at a 5% level is obtained for all the methods. However, we fail to find a significant product effect at a 1% level, when using the standard 2-way mixed model. To see if the results from this example illustrate a general tendency, we have conducted a simulation study, to be presented in the next section.

	SS	MS	DF	F	P-value
Assessor	92.52	13.22	7	5.38	0.0003
Product	32.66	6.53	5		
Interaction:	86.00	2.46	35	3.76	< .0001
<i>Scaling</i>	69.68	9.95	7	17.08	< .0001
<i>Disagreement</i>	16.32	0.58	28	0.89	0.6205
Error	31.37	0.65	48		

Table 1: ANOVA table for the *TVbo* dataset with *Dim glass effect* as the response variable.

## 2.2. Simulation study

For the comparison of the hypothesis tests for overall product difference, 1000 data sets have been simulated from the multiplicative mixed model (3). Each

Model	Test	DF <sub>1</sub>	DF <sub>2</sub>	F	$\chi^2$	P-value
2-way (1)	F-test	5	35	2.66		$3.87 \cdot 10^{-2}$
MAM (2)	F-test	5	28	11.21		$5.32 \cdot 10^{-6}$
mumm (3)	LRT	9			59.93	$1.38 \cdot 10^{-9}$

Table 2: Hypothesis tests for overall product differences with *Dim glass effect* as the response variable.

data set contains scores given by 8 assessors to 6 products in 2 replications. The product effect parameters and the variance components were set equal to the parameter estimates obtained from fitting model (3) by the *mumm* R-package to the *TVbo* data set with *Dim glass effect* as the response variable (without the two last levels of *Picture*). The parameter estimates are given in Table 3.

$\mu + \nu_1$	$\mu + \nu_2$	$\mu + \nu_3$	$\mu + \nu_4$	$\mu + \nu_5$	$\mu + \nu_6$
2.2324	2.9687	3.5215	2.0347	2.0388	2.0665
$\sigma$	$\sigma_{PAN}$	$\sigma_{SCALE}$	$\sigma_D$	$\rho$	
0.7299	0.9588	1.5193	$3.0291 \cdot 10^{-05}$	0.6924	

Table 3: The product effect parameters and the variance components used in the simulation of data for comparison of the hypothesis tests.

125

In this study, we found that in 962 cases out of 1000, the P-value for the MAM was smaller than for the 2-way mixed model. This number was a bit larger for the multiplicative mixed model, with 991 P-values being smaller than for the 2-way model. When looking at the overall product difference significance, with a significance level at 0.05, the MAM finds a significant product difference in 78.0% of the attributes, whereas the multiplicative mixed model finds a difference in 99.2% of the cases. The 2-way model finds the fewest significant differences, with a percentage of 55.8%. The difference between the methods becomes bigger if the significance level is lowered to 0.01. In this case,

130



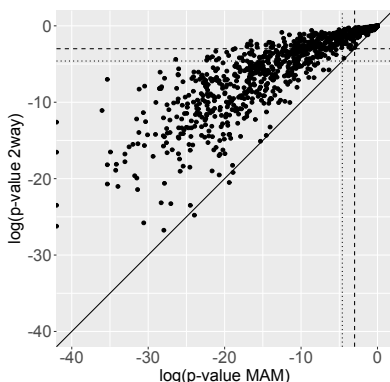


Figure 1: The logarithm of the p-values for the 2-way model plotted against the logarithm of the p-values for the MAM. The dotted and the dashed lines mark  $p\text{-value} = 0.01$  and  $p\text{-value} = 0.05$ , respectively. For the dots on the boundary of the axis the computed value of  $\log(p\text{-value})$  is  $-\infty$ .

135 a significant difference is found for the MAM, the multiplicative mixed model and the 2-way model in 72.2%, 98.4% and 43.8% of the cases, respectively. This reflects the increased power obtained from using the MAM instead of the common 2-way mixed model, as described in Brockhoff et al. (2015). It further shows that using the multiplicative mixed model we obtain the greatest power  
140 to detect product differences.

To get a better understanding of the behavior of the p-values, Figure 1, 2 and 3 show scatter plots of the logarithm of the p-values. The log-transformation was chosen to "spread out" the values close to zero, since the p-values on to the border of significance are of most interest. The p-values for the 2-way model is  
145 plotted against the p-values for the MAM in Figure 1, which clearly illustrates that the p-value for the MAM in general is lower than for the 2-way model. It is further observed that there are quite a lot of situations, where the p-value is large ( $> 0.05$ ) for the 2-way model but small for the MAM ( $< 0.01$ ). Figure 2 shows the p-values for the 2-way model plotted against the p-values for the  
150 multiplicative mixed model. Here we also see a lot of situations, where the

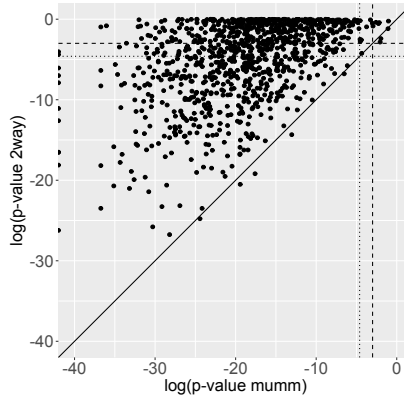


Figure 2: The logarithm of the p-values for the multiplicative mixed model plotted against the logarithm of the p-values for the 2-way model. The dotted and the dashed lines mark  $p\text{-value} = 0.01$  and  $p\text{-value} = 0.05$ , respectively. For the dots on the boundary of the axis the computed value of  $\log(p\text{-value})$  is  $-\infty$ .

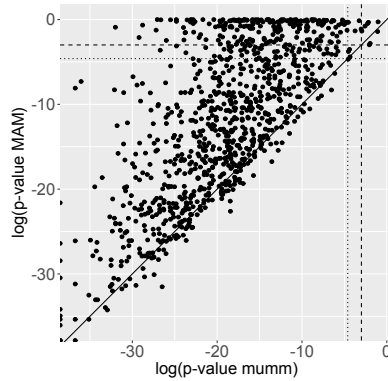


Figure 3: The logarithm of the p-values for the multiplicative mixed model plotted against the logarithm of the p-values for the MAM. The dotted and the dashed lines mark  $p\text{-value} = 0.01$  and  $p\text{-value} = 0.05$ , respectively. For the dots on the boundary of the axis the computed value of  $\log(p\text{-value})$  is  $-\infty$ .

p-value is above 0.05 for the 2-way model but below 0.01 for the multiplicative mixed model. In figure 3, which shows the p-values for the MAM plotted against the p-values for the multiplicative mixed model, we see that the p-values for the multiplicative mixed model more resembles the p-values for the MAM than for the 2-way model. However, a similar pattern as before is seen, with a lot of situations, where the p-value is above 0.05 for the MAM but below 0.01 for the multiplicative mixed model. This indicates that using the multiplicative mixed model, we are able to capture some information about product differences that the two other models fail to. We elaborate more on this in section 4.

### 2.3. Different data scenarios

In the previous data example, the scaling effect was present while the disagreement effect was non-significant. In this section, we want to investigate how generic our findings are, regarding the power to detect product differences, by considering two other types of data scenarios. In the first alternative example, we have analyzed a data set, where both the scaling effect and the disagreement effect are present (at a 5% level). The data also stems from the TVbo data set, but this time we choose the attribute *Colourbalance* as our response variable, and we have discarded all the observations for the first and the fourth level of *Picture*. In the last example, we have considered a data set, where the disagreement effect is present, while the scaling effect is non-significant. Yet again, the data stems from the TVbo data set, where we choose *Depth* as the response variable and discarded the observations for the second and third level of *Picture*.

Table 4 shows the p-values for the interaction effects and the overall product effect for the three data scenarios. When both the scaling effect and the disagreement effect are present, the p-values for an overall product effect are lower for the MAM and for the multiplicative mixed model than for the 2-way model, with the MAM resulting in the lowest p-value. On the other hand, when the scaling effect is non-significant the standard 2-way mixed model gives the lowest p-value, when testing for an overall product effect. We have conducted

Data	ANOVA (interaction)			Product effect		
	Inter.	Scaling	Disagr.	2-way	MAM	mumm
1	< .0001	< .0001	0.6205	$3.87 \cdot 10^{-2}$	$5.32 \cdot 10^{-6}$	$1.38 \cdot 10^{-9}$
2	0.0016	0.0145	0.0413	0.0537	0.0145	0.0420
3	0.0049	0.5665	0.0051	$6.96 \cdot 10^{-3}$	$1.04 \cdot 10^{-2}$	$4.11 \cdot 10^{-2}$

Table 4: p-values for the interaction effects and for the overall product effect in the three data sets. In data set 1, *Dim glass effect* is the response variable, in data set 2, *Colourbalance* is the response, and in data set 3, *Depth* is the response.

Data	Comparison		Power					
			$\alpha = 0.05$			$\alpha = 0.01$		
	1	2	2-way	MAM	mumm	2-way	MAM	mumm
1	96.2	99.1	55.8	78.0	99.2	43.8	72.2	98.4
2	82.5	63.9	66.7	75.2	77.8	47.5	59.3	60.1
3	26.9	7.2	93.4	92.6	83.2	79.1	75.5	61.2

Table 5: The results from the simulation studies. Comparison 1 states the percentage of times the p-value for the MAM is lower than the p-value for the 2-way model. Comparison 2 states the percentage of times the p-value for the mumm is lower than for the 2-way model. The rest of the results are the percentage of times the models find a significant product effect (the power).

a simulation study for the two alternative data scenarios, similar to the one conducted for the first data set. The results are shown in Table 5. When the interaction consists of a significant scaling effect and a significant disagreement effect, the MAM and the multiplicative mixed model have a very similar power to detect product differences, whereas the 2-way model finds a significant effect less often. However, the difference between the methods are smaller in this scenario, than in the scenario without a significant disagreement effect. When the scaling effect, on the other hand, is not significant, the 2-way mixed model has the largest power to detect product differences. This is anyhow not surprising, since the MAM and the mumm waste degrees of freedom on estimating a

non-existing scaling effect.

### 3. Confidence intervals for product differences

The procedure suggested by Brockhoff et al. (2015) for obtaining product difference confidence intervals based on (2), while taking (3) into consideration, is implemented in the R-package *SensMixed*. The exact command used to achieve the estimated confidence intervals is shown in Appendix A. This novel procedure gives non-symmetrical confidence intervals, because the scaling variance is taken into account. This contradicts the confidence intervals one would obtain from running the MAM in standard statistical software for linear models, i.e. computing the confidence intervals in the following way (Brockhoff et al., 2015):

$$\bar{y}_{.j} - \bar{y}_{.j'} \pm t_{0.975}(f) \sqrt{\frac{2MS_{Disagreement}}{IK}},$$

where  $t_{0.975}(f)$  is the 0.975 quantile of the Student's t-distribution with degrees of freedom  $f = DF_{Disagreement} = (I - 1)(J - 2)$ .

195 It is important to note that this method is not recommended, since the scaling variance is falsely ignored. This method of computing confidence intervals will from now on be referred to as method "MAM<sub>naive</sub>", whereas method "MAM" will denote the novel procedure suggested by Brockhoff et al. (2015), which takes the scaling variance into account.

200

The profile likelihood estimated confidence intervals based on (3) were found by R-package *mumm*. The *mumm* function calls, for obtaining the estimated confidence intervals, are shown in Appendix A. Let this method be referred to as method "mumm".

#### 205 3.1. Data example

We consider again the *TVbo* data set, with the attribute *Dim glass effect* as the response variable, but this time we keep all of the observations, meaning that we now analyze 12 products.

Table 6 shows the ANOVA table for this data set and Table 7 shows the  
 210 p-values for the product effect. We see that all the effects are highly significant,  
 except for the disagreement effect, which is non-significant.

	SS	MS	DF	F	P-value
Assessor	277.11	39.59	7	9.05	< .0001
Product	295.76	26.89	11		
Interaction:	337.00	4.38	77	1.96	0.0009
<i>Scaling</i>	215.09	30.73	7	17.64	< .0001
<i>Disagreement</i>	121.90	1.74	70	0.78	0.8656
Error	214.89	2.24	96		

Table 6: ANOVA table for the *TVbo* dataset with *Dim glass effect* as the response variable.

Model	Test	DF <sub>1</sub>	DF <sub>2</sub>	F	$\chi^2$	P-value
2-way (1)	F-test	11	77	6.14		$3.87 \cdot 10^{-07}$
MAM (2)	F-test	11	70	15.44		$1.02 \cdot 10^{-14}$
mumm (3)	LRT	17			128.16	$< 10^{-14}$

Table 7: Hypothesis tests for overall product differences with *Dim glass effect* as the response variable.

It is clear that the product effect is significant no matter which model is  
 used. We also see that the disagreement effect is non-significant, meaning that  
 the scaling effect alone explains the assessor-by-product interaction.

215 We have estimated the product contrasts and their corresponding 95%-  
 confidence intervals by method "MAM", method "MAM<sub>naive</sub>", method "mumm",  
 and by using the standard 2-way mixed model. To illustrate the difference be-  
 tween these four methods, the contrast estimates and the confidence intervals  
 for the three smallest and largest contrasts, ordered according to the estimates  
 220 obtained from fitting model (3), are plotted in Figure 4. It is clear that the  
 estimated confidence intervals from method "MAM" and method "mumm" are

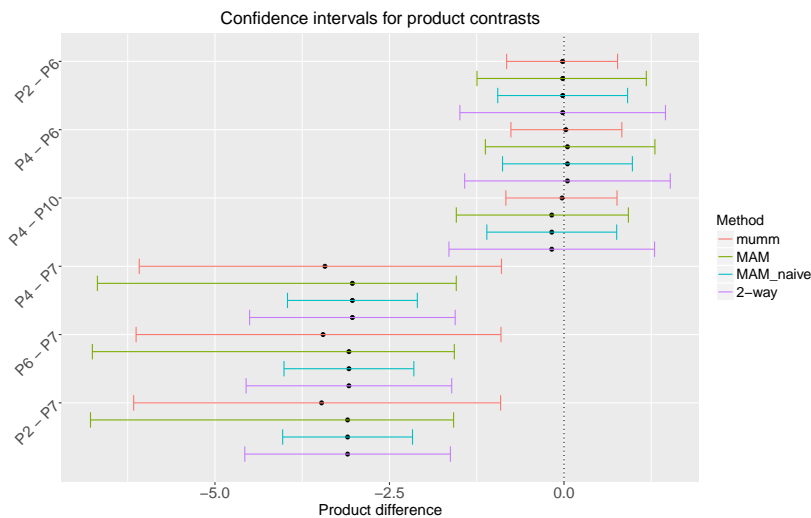


Figure 4: The estimated product contrasts (black dots) and the corresponding 95%-confidence intervals. Method "MAM" refers to the procedure suggested in Brockhoff et al. (2015), method "MAM<sub>naive</sub>" refers to using MAM directly, and method "mumm" refers to the use of R-package *mumm* for finding the profile likelihood based confidence intervals. Only the confidence intervals for the three smallest and the three largest contrasts are plotted.

asymmetric as expected, being wider "away from zero" than "towards zero" (Brockhoff et al., 2015). Moreover, the intervals obtained by method "MAM" are wider than the intervals obtained from method "mumm", for the small contrasts. The confidence intervals from method "MAM<sub>naive</sub>" and the 2-way model are symmetric and have a constant width for all of the contrasts, which makes them notably shorter than the intervals from the two other methods, for large contrasts, and wider for small contrasts. For small contrasts, the intervals for "MAM<sub>naive</sub>" look reasonable, whereas the intervals for the 2-way model appear to be too wide compared to the others. However, from this plot we cannot say much about which intervals are more correct. Therefore, we have conducted a simulation study to compute the coverage probabilities for the confidence intervals, which will be described in the following section.

### 3.2. Simulation study

235 To compute coverage probabilities for the estimated confidence intervals, 1000 datasets were simulated from the multiplicative mixed model (3). Each data set contains scores given by 8 assessors to 12 products in 2 replications. The product effect parameters and the variance components were set equal to the parameter estimates obtained from fitting model (3) to the *TVbo* data set with  
 240 *Dim glass effect* as the response variable. The estimated model parameters used in the simulation study are shown in Table 8. We have estimated the

$\mu + \nu_1$	$\mu + \nu_2$	$\mu + \nu_3$	$\mu + \nu_4$	$\mu + \nu_5$	$\mu + \nu_6$
2.2259	1.9794	3.8116	2.0273	3.1773	2.0004
$\mu + \nu_7$	$\mu + \nu_8$	$\mu + \nu_9$	$\mu + \nu_{10}$	$\mu + \nu_{11}$	$\mu + \nu_{12}$
5.4519	2.4355	3.7462	2.0544	5.1984	2.0916
$\sigma$	$\sigma_{PAN}$	$\sigma_{SCALE}$	$\sigma_D$	$\rho$	
1.3366	1.1700	0.9117	$6.62 \cdot 10^{-5}$	0.8025	

Table 8: The model parameters used in the simulation of data sets.

product difference confidence intervals for all of the 1000 datasets by the four methods. Hereafter, the coverage probabilities of the confidence intervals have been calculated from the "true" known product differences.

245 Table 9 shows the coverage probabilities of the estimated confidence intervals, found from the simulation study. As seen, the coverage probabilities of the confidence intervals produced by method "MAM" look very reasonable and are very similar to the coverage probabilities for method "mumm". On the other hand, the coverage probabilities for method "MAM<sub>naive</sub>" and for the 2-way  
 250 model are clearly too high for small contrasts and they are unreasonably below 0.95 for large contrasts.



	Contrast	mumm	MAM	MAM <sub>naive</sub>	2-way
P2-P6	-0.0210	0.9580	0.9670	0.9780	0.9940
P4-P6	0.0269	0.9600	0.9660	0.9830	0.9970
P4-P10	-0.0271	0.9430	0.9710	0.9830	0.9930
P10-P12	-0.0371	0.9550	0.9690	0.9790	0.9950
P2-P4	-0.0479	0.9530	0.9600	0.9760	0.9960
⋮	⋮	⋮	⋮		
P7-P12	3.3604	0.9350	0.9310	0.6940	0.7920
P7-P10	3.3975	0.9390	0.9270	0.7170	0.7990
P4-P7	-3.4246	0.9340	0.9310	0.6950	0.8000
P6-P7	-3.4515	0.9410	0.9310	0.6830	0.7780
P2-P7	-3.4725	0.9350	0.9380	0.6950	0.7850

Table 9: The coverage probabilities.

#### 4. New F-test for the MAM

In section 2 it was found that the MAM gives a higher power to detect product differences than the 2-way mixed model, when a scaling effect is present.

255 Using the multiplicative mixed model, however, we obtain an even higher power to detect product differences. In this section, we therefore propose a new and improved F-test for the MAM, for testing the significance of an overall product effect.

When using the 2-way mixed model, the denominator in the F-test, for  
 260 testing the significance of an overall product effect, is the mean square for the product-by-assessor interaction ( $MS_{Interaction}$ ). However, the interaction contains the scale effect, which means that the variation due to the assessors different use of scale range ends up in the denominator, i.e in the "error". When using the MAM, on the other hand, the denominator is  $MS_{Disagreement}$ , which  
 265 is the mean square for the interaction without the mean square for the scale effect. Hence, the scale effect has been removed from the "error" in the F-test,

giving an increased power to detect product differences.

In the likelihood ratio test, when using the multiplicative mixed model, the full model is tested against a reduced model, where both the product effect and the scaling effect is removed. This idea is similar to using an F-test for the MAM with the mean square for the combined effect of product and scaling ( $MS_{Product+Scaling}$ ) in the numerator instead of just  $MS_{Product}$ . This makes sense, since a large scaling effect is a sign of product difference. Therefore we propose a new F-test for testing the significance of an overall product effect, with the F-statistic:

$$F_{Product} = \frac{MS_{Product+Scaling}}{MS_{Disagreement}} = \frac{(SS_{Product} + SS_{Scaling})/(J + I - 2)}{MS_{Disagreement}}$$

When performing this F-test in the simulation study described in 2.2 (corresponding to Data 1 in Table 5), we find a significant product difference in  
 270 91.1% and 88.0% of the attributes with a 5% and 1% level, respectively. This is closer to the results we obtained when using the multiplicative mixed model. Figure 5 shows the p-values for the 2-way model plotted against the p-values from the new F-test, which looks rather similar to the Figure 2, showing the p-values for the 2-way model plotted against the p-values from the multiplicative  
 275 mixed model. Figure 6 shows the p-values from the new F-test plotted against the p-values from the multiplicative mixed model, which shows that the values seem to follow each other pretty nicely. Thus, using the MAM together with the newly proposed F-test we increase the power to detect product differences, in a similar fashion as when using the multiplicative mixed model together with  
 280 the likelihood ratio test.

## 5. Summary and discussion

In this paper we have compared the hypothesis tests for overall product difference when using the MAM, the 2-way mixed model and the multiplicative mixed model. It was found that the use of the multiplicative mixed model  
 285 results in the highest power to detect product differences, when a scaling effect

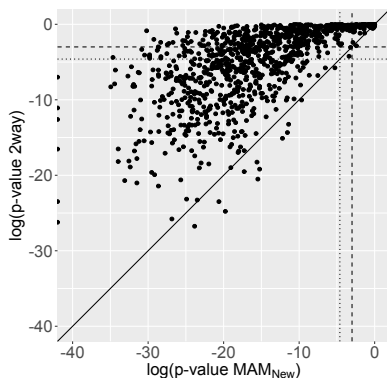


Figure 5: The logarithm of the p-values for the 2-way model plotted against the logarithm of the p-values for the MAM with the new F-test. The dotted and the dashed lines mark  $p\text{-value} = 0.01$  and  $p\text{-value} = 0.05$ , respectively. For the dots on the boundary of the axis the computed value of  $\log(p\text{-value})$  is  $-\infty$ .

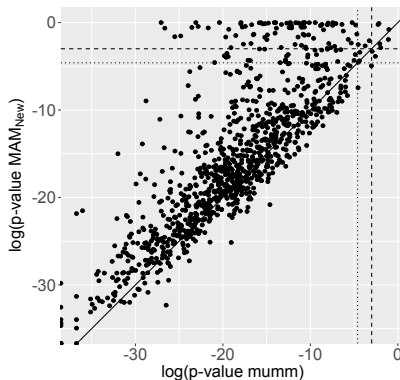


Figure 6: The logarithm of the p-values for the MAM with the new F-test plotted against the logarithm of the p-values for the multiplicative mixed model. The dotted and the dashed lines mark  $p\text{-value} = 0.01$  and  $p\text{-value} = 0.05$ , respectively. For the dots on the boundary of the axis the computed value of  $\log(p\text{-value})$  is  $-\infty$ .

is present. It was further found, in accordance with Brockhoff et al. (2015), that also the MAM gives an increased power to detect product differences compared to the 2-way mixed model.

Through simulation studies, this paper also investigated whether the suggested procedure in Brockhoff et al. (2015), based on the MAM, produce appropriate confidence intervals for product differences. We compared those intervals with the profile likelihood based confidence intervals, based on the multiplicative mixed model, one can obtain from using the R-package *mumm*. It was found that both methods result in very proper confidence intervals with reasonable coverage probabilities.

### 5.1. Computation time

The estimation of the parameters in the multiplicative mixed model (3), when fitting the model to the *Dim glass effect* attribute by *mumm*, takes on average 0.36 seconds (average of 100 runs). To achieve the corresponding confidence intervals for all of the 66 product contrasts, 410.54 seconds, i.e. almost 7 minutes, are needed on average<sup>1</sup>. In Pødenphant et al. (Unpublished results) the computation time for the model estimation when using *mumm* is compared to the computation time for fitting the model by the *NLMIXED* procedure in SAS (Littell et al., 2007), and the authors state that *mumm* is faster than *NLMIXED*. When fitted to the *TVbo* data set, *mumm* is more than 20% faster than *NLMIXED*, and when fitted to larger data sets, the difference can be substantial; In one of their examples, *mumm* is more than 40 times as fast. They further note that proper confidence intervals for the multiplicative mixed model cannot be estimated by *NLMIXED*.

The computational burden is, however, immensely reduced for the "MAM" method, with an average computation time of only 0.29 seconds for fitting the MAM and estimating the 66 confidence intervals. This gives a strong advan-

---

<sup>1</sup>The computer used has 16 GB RAM, an Intel Core i7-6500U processor and runs under the operating system Windows 7 Enterprise. The version of R is 3.4.2.

tage to MAM, especially regarding simulation studies and analysis of data from experiments with many products.

### 315 5.2. Concluding remarks

The formal modeling approach of fitting the multiplicative mixed model with likelihood methods is advantageous in terms of insight and understanding. Further, it makes it possible to get profile likelihood based confidence intervals. The formal modeling, however, comes with the cost of a relative high computation  
320 time. The Mixed Assessor Model is on the other hand very fast to estimate. When a scaling effect is present in the data, both models give an improved power to detect product differences compared to a 2-way mixed analysis of variance, with the multiplicative model resulting in the highest. We have, however, proposed a new F-test for the MAM, which results in a power that resembles the  
325 power of the formal modeling approach. Further, we found that the suggested procedure in Brockhoff et al. (2015), based on the MAM, produces appropriate confidence intervals for product differences. In the light of the reduced computation time, we therefore see the MAM as a good alternative to the formal multiplicative mixed model.

## 330 6. Compliance with Ethical Standards

**Funding:** The conduct of the research is fully funded by the Technical University of Denmark (DTU).

**Declaration of interest:** None

### Appendix A. R-code

Listing 1: R-code for estimating product contrast confidence intervals.

```
335 library(SensMixed)
library(mumm)

#Loading and preparing the data
```

```

340 data = TVbo
data$Product = factor(data$TVset:data$Picture)
data$y = data$Dimglasseffect

#Fitting the MAM
345 fit_MAM = sensmixed(c("y","y"), c("Product"),c("Assessor"),
                      data = data, product_structure = 1,
                      error_structure = "ONLY-ASS", MAM = TRUE,
                      control = sensmixedControl(calc_post_hoc = TRUE,
350                                             MAM_balanced = TRUE,
                                             MAM_adjusted = FALSE))

#Estimating the product contrasts and their confidence intervals
MAM_contrasts = fit_MAM[[5]][,1][,1]
fit_MAM_posth = fit_MAM[[8]][,1]
355 MAM_conf_lower = fit_MAM_posth[,1]
MAM_conf_upper = fit_MAM_posth[,2]

#Fitting the multiplicative mixed model
fit_mumm = mumm(y ~ -1 + Product + (1|Assessor) +
360                (1|Product:Assessor) +
                mp(Assessor,Product), data = data)

#Estimating the product contrasts and their confidence intervals
c1 = combn(12,2)[1,]
365 c2 = combn(12,2)[2,]
matrix_contrasts = matrix(0,length(c1),nlevels(data$Product))
matrix_contrasts[cbind(1:length(c1),c1)] = 1
matrix_contrasts[cbind(1:length(c2),c2)] = -1
matrix_contrasts_full = cbind(matrix_contrasts,matrix(0,66,5))
370
mumm_contrasts = matrix_contrasts%*%fit_mumm$par.fix
mumm_conf = confint(fit_mumm, parm = matrix_contrasts_full, level =
0.95)

```

375 **References**

- Brockhoff, P. B. (2003). Statistical testing of individual differences in sensory profiling. *Food Quality and Preference*, *14*, 425–434.
- Brockhoff, P. B., & Belmonte, F. (2018). Applied univariate statistics. In N. Zacharov (Ed.), *Sensory Evaluation of Sound* chapter 6. CRC Pr I Llc.
- 380 Brockhoff, P. B., & Linander, C. B. (2017). Analysis of the data using the r package sensr. In *Discrimination Testing in Sensory Science* (pp. 303–344). Elsevier.
- Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account by analyzing profile data with the mixed assessor  
385 model. *Food Quality and Preference*, *39*, 156–166.
- Brockhoff, P. M., & Skovgaard, I. M. (1994). Modelling individual differences between assessors in sensory evaluations. *Food Quality and Preference*, *5*, 215–224.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present  
390 only under the alternative. *Biometrika*, *64*, 247–254.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, *70*, 1–21. doi:10.18637/jss.v070.i05.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016a). *lmerTest:*  
395 *Tests in Linear Mixed Effects Models*. URL: <https://CRAN.R-project.org/package=lmerTest> r package version 2.0-32.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*.
- 400 Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016b). *SensMixed: Analysis of Sensory and Consumer Data in a Mixed Model*

- 
- Framework*. URL: <http://CRAN.R-project.org/package=SensMixed> r package version 2.0-9.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2007). *SAS for mixed models*. SAS institute.
- Peltier, C., Brockhoff, P. B., Visalli, M., & Schlich, P. (2014). The mam-cap table: A new tool for monitoring panel performances. *Food Quality and Preference*, *32*, 24–27.
- Pødenphant, S., & Brockhoff, P. B. (2016). *mumm: Multiplicative Mixed Models using Template Model Builder*. URL: <https://CRAN.R-project.org/package=mumm> r package version 0.2.0.
- Pødenphant, S., Kristensen, K., & Brockhoff, P. B. (Unpublished results). The multiplicative mixed model with the mumm r-package as a general and easy random interaction model tool. *The American Statistician*, *Submitted*.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: <https://www.R-project.org/>.
- Ritz, C., & Skovgaard, I. M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*, *92*, 507–517.
- Smith, A., Cullis, B., Brockhoff, P., & Thompson, R. (2003). Multiplicative mixed models for the analysis of sensory evaluation data. *Food Quality and Preference*, *14*, 387–395.
- Vonesh, E. F. (1996). A note on the use of laplace’s approximation for nonlinear mixed-effects models. *Biometrika*, *83*, 447–452.
- Wolfinger, R. (1993). Laplace’s approximation for nonlinear mixed models. *Biometrika*, *80*, 791–795.





## APPENDIX C

# Biadditive Models - a model overview and estimation by the Template Model Builder

---

**Pødenphant, S.**, Hadasch, S., Brockhoff, P. B., Piepho, H.-P. (2018). *Biadditive Models and the Template Model Builder - a model overview and estimation methods* (Draft intended for a statistics journal).

# Biadditive Models

## - a model overview and estimation by the Template Model Builder

Sofie Pødenphant<sup>a,\*</sup>, Steffen Hadasch<sup>b</sup>, Per B. Brockhoff<sup>a</sup>, Hans-Peter Piepho<sup>b</sup>

<sup>a</sup>*DTU Compute, Section for Statistics and Data Analysis, Technical University of  
Denmark, Richard Petersens Plads, DK-2800 Kongens Lyngby, Denmark*

<sup>b</sup>*Institute of Crop Science, Biostatistics Unit, University of Hohenheim, Fruwirthstr. 23,  
DE-70599 Stuttgart, Germany.*

---

### Abstract

---

### 1. Introduction

Biadditive models are defined by Denis & Gower (1994) and Denis & Gower (1996) as models of the form shown in (1), together with models that results from dropping one or both main effects from (1).

$$y_{ij} = \mu + \alpha_i + \gamma_j + \sum_{k=1}^K \beta_{ik} \delta_{jk} + e_{ij}, \quad (1)$$

$$i = 1, \dots, I; \quad j = 1, \dots, J; \quad e_{ij} \sim N(0, \sigma^2).$$

This model was first presented in Gollob (1968) and is frequently used in the area of agricultural science, where it is fitted to genotype-environment data. In that case,  $y_{ij}$  is the yield of genotype  $i$  in environment  $j$  or vice versa. However, biadditive models are also used in other scientific disciplines, e.g. sensory science and medicine. The full model is often denoted as Additive Main effects Multiplicative Interaction (AMMI) model (Gauch Jr, 1988), especially in agronomy. In this paper we will for the sake of simplicity limit our focus to models with only one multiplicative term, i.e. to models with  $K = 1$ , but extension to  $K > 1$  is straightforward.

---

\*Corresponding author  
Email address: sofp@dtu.dk (Sofie Pødenphant)

## 2. Fixed effects models

In a fixed-effect model perspective, model (1) with  $K = 1$  can be written as

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \gamma_j + \beta_i \delta_j + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \quad (2)$$

with the following identifiability constraints

$$\begin{aligned} \|\delta\| &= 1 \\ \sum_{i=1}^I \alpha_i &= \sum_{j=1}^J \gamma_j = \sum_{i=1}^I \beta_i = \sum_{j=1}^J \delta_j = 0, \end{aligned}$$

where  $\|\delta\|$  denotes the  $\ell^2$ -norm of  $\delta$ . In this model,  $\mu$  is the grand mean,  $\alpha_i$  is the row main effect,  $\gamma_j$  is the column main effect and  $e_{ij}$  is the error term. Depending on the context,  $\beta_i$  can be interpreted as a latent variable to which the different levels in the column factor respond differently through the regression coefficient  $\delta_j$ , or the other way around. This model is popular in agricultural and biological applications (Piepho, 1997), where it is known as AMMI1, with "1" referring to the single multiplicative term (Gauch Jr, 1992).

*Dropping the row main effect:*

Model (2) might be reduced by dropping either the row main effect or the column main effect. Dropping the row main effect leads to the following model

$$\begin{aligned} y_{ij} &= \mu + \gamma_j + \beta_i \delta_j + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \quad (3)$$

with the following identifiability constraints

$$\begin{aligned} \|\delta\| &= 1 \\ \sum_{j=1}^J \gamma_j &= \sum_{i=1}^I \beta_i = 0 \end{aligned}$$

This model is also often written as

$$\begin{aligned} y_{ij} &= \tilde{\gamma}_j + \tilde{\delta}_j \tilde{\beta}_i + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \quad (3b)$$

with the following identifiability constraints

$$\sum_{j=1}^J \tilde{\delta}_j = J, \quad \sum_{i=1}^I \tilde{\beta}_i = 0.$$

In this version,  $\tilde{\gamma}_j$  is the column mean for column  $j$ ,  $\tilde{\beta}_i$  is the row main effect, and the  $\tilde{\delta}_j$ s can be interpreted as regression coefficients. The relationship between the parameters in (3) and (3b) can be described as follows:

$$15 \quad \mu + \gamma_j = \tilde{\gamma}_j, \quad \delta_j = \tilde{\delta}_j / \|\tilde{\delta}\|, \quad \beta_i = \tilde{\beta}_i \|\tilde{\delta}\|.$$

This model is often used to analyze data from e.g. the field of agricultural science and the field of sensory science. In the first mentioned field the model (3b) is known as Finlay-Wilkinson regression if  $i$  is the environment index and  $j$  is the genotype index. In that case,  $\tilde{\gamma}_j$  is the expected performance of genotype  $j$  in an "average" environment,  $\tilde{\beta}_i$  is the main effect of environment  $i$ ,  $\tilde{\delta}_j$  is the sensitivity of genotype  $j$  to the environment, and  $e_{ij}$  is the error term. This model results in genotype-specific regression lines on the environmental quality, assessed by the main effect  $\tilde{\beta}_i$ , with  $\tilde{\gamma}_j$  as the intercept and  $\tilde{\delta}_j$  as the slope. The model was first proposed by Yates & Cochran (1938), but later popularized by Finlay & Wilkinson (1963), hence the name of the method. Another version of model (3), often applied in the field of agricultural science, is the GGE (genotype main effects and genotype-by-environment interaction) model with one multiplicative term. The GGE model is often written as

$$\begin{aligned} y_{ij} &= \tilde{\gamma}_j + \lambda u_i v_j + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \\ \|\mathbf{u}\| &= \|\mathbf{v}\| = 1 \\ \sum_{i=1}^I u_i &= 0 \end{aligned} \quad (3c)$$

Now  $i$  is the genotype index and  $j$  is the environment index, meaning that  $\tilde{\gamma}_j$  is the expected performance of an "average" genotype in environment  $j$ . This model is equal to model (3b), with  $v_j = \tilde{\delta}_j / \|\tilde{\delta}\|$ ,  $u_i = \tilde{\beta}_i / \|\tilde{\beta}\|$  and  $\lambda = \|\tilde{\delta}\| \|\tilde{\beta}\|$ , which means that this model results in environment-specific regression lines on the genotype effect, with  $\tilde{\gamma}_j$  as the intercept and  $\tilde{\delta}_j = v_j / \bar{v}$  as the slope.

Already in Eberhart & Russell (1966), model (3b) was expanded to include column (genotype) specific error variances to describe the performance of genotypes over a series of environments, where  $i$  is the environment index and  $j$  denotes the genotype. Also in the field of sensory science model (3b) has been expanded with column (assessor) specific error variances. This version of the model is called the basic assessor model and was presented in Brockhoff & Skovgaard (1994):

$$y_{ij} = \tilde{\gamma}_j + \tilde{\delta}_j \tilde{\beta}_i + e_{ij}, \quad (3d)$$

$$e_{ij} \sim N(0, \sigma_j^2),$$

where  $y_{ij}$  is the score of product  $i$  given by assessor  $j$ ,  $\tilde{\gamma}_j$  is the main effect of assessor  $j$ ,  $\tilde{\beta}_i$  is the main effect of product  $i$  and  $\tilde{\delta}_j$  is the "scaling" slope for assessor  $j$ , which is a measure of how much the assessor in question spreads the products on the subjective scoring scale. Brockhoff & Skovgaard (1994) impose the restrictions  $\bar{\tilde{\beta}} = 0$  and  $MS_{\tilde{\beta}}$ , where the latter is the average over the squared  $\tilde{\beta}_i$ s. We will, however, keep the restrictions from (3b).

*Dropping the row main effect and the column main effect:*

Model (2) might be reduced by dropping both the row main effect and the column main effect, giving us the following model:

$$y_{ij} = \mu + \beta_i \delta_j + e_{ij}, \quad (4)$$

$$e_{ij} \sim N(0, \sigma^2)$$

with the following identifiability constraints

$$\|\delta\| = 1$$

In this model,  $\mu$  is not the grand mean but a so-called "shift parameter", which gives rise to the name "Shifted Multiplicative Model" (Seyedsadr & Cornelius, 1992). Model (4) can be reparametrized into the following form

$$\begin{aligned} y_{ij} &= \phi + \alpha_i + \gamma_j + \lambda\alpha_i\gamma_j + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{4b}$$

with the following identifiability constraints

$$\|\alpha\| = \|\gamma\| = 1$$

The parameters in (4b) can be described by the parameters from (4) with  $\mu = \phi - 1/\lambda$ ,  $\beta_i = (\lambda\alpha_i + 1)/\|\gamma + 1/\lambda\|$ ,  $\delta_j = (\gamma_j + 1/\lambda)/\|\gamma + 1/\lambda\|$ .

*Dropping the row main effect, the column main effect and the grand mean/shift parameter:*

Model (2) might be reduced by dropping both the row main effect and the column main effect together with the grand mean, giving us the following model:

$$\begin{aligned} y_{ij} &= \beta_i\delta_j + e_{ij}, \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{5}$$

This model was introduced by Fisher & Mackenzie (1923).

### 3. Mixed effects models

In a mixed-effect model perspective, model (1) with  $K = 1$  can be written as

$$y_{ij} = \mu + \alpha_i + \gamma_j + \beta_i \delta_j + e_{ij}, \quad (6)$$

$$(\gamma_j, \delta_j) \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} \sigma_\gamma^2 & \rho \sigma_\gamma \\ \rho \sigma_\gamma & \sigma_\delta^2 \end{bmatrix} \right), \quad e_{ij} \sim N(0, \sigma^2)$$

with the following identifiability constraints

$$\sum_{i=1}^I \alpha_i = \sum_{i=1}^I \beta_i = 0, \quad \sigma_\delta = 1.$$

The interpretation of the model parameters are similar to the interpretation in model (2), except that the column factor is considered random. Due to the random column factor, the first two moments of  $\mathbf{y}_j$  is given by

$$E(\mathbf{y}_j) = \mu \mathbf{1}_I + \boldsymbol{\alpha}$$

$$\text{var}(\mathbf{y}_j) = \sigma_\gamma^2 \mathbf{1}_J \mathbf{1}_J^T + \sigma_\delta^2 \boldsymbol{\beta} \boldsymbol{\beta}^T + \rho \sigma_\gamma \sigma_\delta (\boldsymbol{\beta} \mathbf{1}_I^T + \mathbf{1}_I \boldsymbol{\beta}^T) + \sigma^2 \mathbf{I}_I,$$

This mixed version of the AMMI model was proposed by Piepho (1997) for the analysis of genotype-environment data, where  $y_{ij}$  is the yield of genotype  $i$  in environment  $j$ . Piepho (1997), however, imposes independence between  $\gamma_j$  and  $\delta_j$  and suggests to restrict the mean of  $\beta$  to 1 and lift the constrain on  $\sigma_\delta$ . In Piepho et al. (1998) and Piepho (1999) the model with independence between  $\gamma_j$  and  $\delta_j$  is also considered, but there  $\delta_j$  has unit variance while  $\beta$  is left unconstrained. These constraints are a good alternative to the constraints proposed in this paper, especially if the preferred optimization routine does not allow for a correlation between random effects. In that case, the variance of  $\mathbf{y}_j$  can be written as

$$\text{var}(\mathbf{y}_j) = \sigma_\gamma^2 \mathbf{1}_J \mathbf{1}_J^T + \boldsymbol{\beta} \boldsymbol{\beta}^T + \sigma^2 \mathbf{I}_I,$$

where it is clear that the model has a factor analytic covariance structure.

It should be noticed, however, that the interpretation of the parameters depends on the constraints; if the  $\beta_i$ s do not sum to zero,  $\gamma_j$  is no longer the



column main effect. This can be seen by writing up the expected column mean for column  $j$ :

$$E(\bar{y}_{.j}) = \frac{\sum_{i=1}^I (\mu + \alpha_i + \gamma_j + \beta_i \delta_j + e_{ij})}{I} = \mu + \gamma_j + \bar{\beta} \delta_j.$$

From this, it is clear that the average of  $\beta_i$ , and thereby also the sum, needs to  
 25 be zero, for  $\gamma_j$  to be the column main effect.

Smith et al. (2005) gives an overview of different mixed model approaches for the analysis of series of variety trials. Model (6) is among the described approaches, where the case with fixed environment and random genotype effects are considered as well as the converse. The first-mentioned version, however,  
 30 allows for environment dependent residual variances.

More general and extended versions of model (6) are suggested in Smith et al. (2001) and Smith et al. (2003) for the analysis of genotype-environment data and sensory evaluation data, respectively. These models includes replications, multiple multiplicative terms and additional random and fixed effects associated with the design of the experiment.

*Dropping the row main effect:*

Dropping the row main effect, from model (6) gives

$$y_{ij} = \mu + \gamma_j + \beta_i \delta_j + e_{ij} = \mu + \gamma_j + \beta_i + \theta_j \beta_i + e_{ij}, \quad (7)$$

$$(\gamma_j, \delta_j) \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \sigma_\gamma^2 & \rho \sigma_\gamma \sigma_\delta \\ \rho \sigma_\gamma \sigma_\delta & \sigma_\delta^2 \end{bmatrix} \right), \quad e_{ij} \sim N(0, \sigma^2),$$

with  $\theta_j = \delta_j - 1$  and  $\sum_{i=1}^I \beta_i = 0$ . Now  $\beta_i$  is the row main effect, meaning that the model can be interpreted as a regression on the row main effect with  $\delta_j$  as the slope and  $\mu + \gamma_j$  as the intercept. The mean and the variance of  $y_{ij}$  are

$$E(y_j) = \mu \mathbf{1}_I + \beta$$

$$\text{var}(\mathbf{y}_j) = \sigma_\gamma^2 \mathbf{1}_J \mathbf{1}_J^T + \sigma_\delta^2 \beta \beta^T + \rho \sigma_\gamma \sigma_\delta (\beta \mathbf{1}_I^T + \mathbf{1}_I \beta^T) + \sigma^2 \mathbf{I}_I,$$

where it is clear that the mean parameter vector,  $\beta$ , is included in the variance structure, which complicates the estimation of the model, see Section 4. This

model was introduced by Piepho (1999) for the analysis of genotype-environment data, where  $y_{ij}$  denotes the heading dates of grasses, measured in days after April 1, for genotype  $i$  in environment  $j$ . In Brockhoff et al. (2015), the model is proposed for the analysis of sensory profile data, where  $y_{ij}$  is the evaluation by the  $j$ th assessor of the  $i$ th product. Pødenphant et al. (2018) gives a thorough description of model (7) in regards to estimation, interpretation and application. The paper includes examples, where the model is applied to sensory profile data, genotype-environment data and data from method comparison studies in medicine. In the latter application,  $y_{ij}$  is the measured value obtained from using method  $j$  on item/patient  $i$ .

*Dropping the column main effect:*

Dropping the column main effect from (6) gives

$$y_{ij} = \mu + \alpha_i + \beta_i \delta_j + e_{ij} \quad (8)$$

$$\delta_j \sim N(0, \sigma_\delta^2) \quad e_{ij} \sim N(0, \sigma^2), \quad (9)$$

with  $\sum_{i=1}^I \alpha_i = 0$  and  $\sum_{i=1}^I \beta_i = I$ . Here  $\alpha_i$  is the row main effect and  $\delta_j$  the column main effect, meaning that the model can be interpreted as a regression on the column main effect with  $\beta_i$  as the slope and  $\mu + \alpha_i$  as the intercept. The first two moments of  $y_{ij}$  are

$$E(\mathbf{y}_j) = \mu \mathbf{1}_I + \boldsymbol{\alpha}$$

$$\text{var}(\mathbf{y}_j) = \sigma_\delta^2 \boldsymbol{\beta} \boldsymbol{\beta}^T + \sigma^2 \mathbf{I}_I,$$

where it is again clear that the covariance has a factor analytic structure. A version of this model, which includes replications and heteroscedastic residual variances, was introduced by Oman (1991). The exact version of (8) is considered in Piepho (1997) for the analysis of genotype-environment data, where  $y_{ij}$  is the mean yield of genotype  $i$  in environment  $j$ . In this case, the model can be interpreted as a mixed model version of the Finlay & Wilkinson (1963) model. In Smith et al. (2003) an extended version of the model, with multiple multiplicative terms and heteroscedastic residual variances, is used to analyze

data from sensory evaluation experiments. In that case,  $y_{ij}$  is the evaluation by  
 40 the  $j$ th assessor of the  $i$ th product. Model (8) is also considered in Piepho et al.  
 (2015) for meta-analysis in medicine, where  $y_{ij}$  is the value of the linear predic-  
 tor for the  $i$ th treatment in the  $j$ th trial. In Piepho et al. (2015), however, the  
 authors reparameterize the model by requiring  $\sigma_\delta^2 = 1$  instead of  $\sum_{i=1}^I \beta_i = I$ .  
 This changes the interpretation of the parameters slightly, since  $\delta_j$  becomes a  
 45 scaled version of the column main effect.

#### 4. Estimation

##### 4.1. Fixed effects models

The main effects in the AMMI1 model (2) are commonly estimated from the  
 raw means:  $\hat{\mu} = \bar{y}_{..}$ ,  $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$ ,  $\hat{\gamma}_j = \bar{y}_{.j} - \bar{y}_{..}$ . The parameters in the  
 50 multiplicative term can thereafter be estimated by performing a singular value  
 decomposition of  $\hat{\mathbf{Z}}$ , with  $\hat{z}_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\gamma}_j$ , where the  $\hat{\delta}_j$ s are elements  
 of the first column eigenvector and the  $\hat{\beta}_i$ s are the elements of the first row  
 eigenvector multiplied by the first singular value. This solution is equal to the  
 maximum likelihood solution.

55 The reduced model (3) can be estimated in the same way:  $\hat{\mu} = \bar{y}_{..}$ ,  $\hat{\gamma}_j = \bar{y}_{.j} - \bar{y}_{..}$ ,  
 and the multiplicative term is found by performing a singular value decomposi-  
 tion of  $\hat{\mathbf{Z}}$ , with  $\hat{z}_{ij} = y_{ij} - \hat{\mu} - \hat{\gamma}_j$ . Again, the  $\hat{\delta}_j$ s are elements of the first column  
 eigenvector and the  $\hat{\beta}_i$ s are the elements of the first row eigenvector multiplied  
 60 by the first singular value. When estimating the GGE in (3c), the only differ-  
 ence is that the overall mean and the row main effect is combined into  $\hat{\gamma}_j = \bar{y}_{.j}$ ,  
 and that the elements of the first row eigenvector and the first singular value  
 are not multiplied together, but kept apart, with  $\hat{\lambda}$  being the singular value and  
 the  $\hat{u}_i$ s being the elements of the first row eigenvector. The  $\hat{v}_j$ s corresponds to  
 65 the  $\hat{\delta}_j$ , being the elements of the first column eigenvector. The parameters in  
 the multiplicative term in model (3b) can be found from the parameters in (3),

by the following relationship:  $\tilde{\delta}_j = \frac{J\delta_j}{\sum_{j=1}^J \delta_j}$  and  $\tilde{\beta}_i = \frac{\beta_i \sum_{j=1}^J \delta_j}{J}$ . These solutions are all equal to the maximum likelihood solution.

Model (3d) is not as straightforward to estimate, because of the column specific error variances. In Eberhart & Russell (1966) they first estimate the row main effects,  $\tilde{\beta}_i$ , from the raw means, and secondly estimate the rest of the parameters by a usual regression of  $\mathbf{y}$  on  $\boldsymbol{\beta}$ . In Brockhoff & Skovgaard (1994), the model is estimated by optimization of the log-likelihood function in an iterative manner, alternating between the estimation of the  $\beta_i$ s while assuming the rest of the parameters are known, and the estimation of the  $\tilde{\gamma}_j$ s, the  $\tilde{\delta}_j$ s and the  $\sigma_j$ s while assuming the  $\beta_i$ s are known.

The Shifted Multiplicative Model in (4) is more complicated to estimate, since the maximum likelihood fit or the least squares fit of the model is not equivalent to first fitting  $\mu$  from the raw overall mean and then performing a singular value decomposition to find the multiplicative term (Gabriel, 1978); in Gabriel (1978) it is denoted as "an unsolved case". Seyedsadr & Cornelius (1992) propose a Newton-Raphson iterative method to minimize the residual sum of squares, which, they argue, work extremely well if the starting value for  $\hat{\mu}$  is close to the solution. In case the Newton-Raphson method does not converge, they propose the use of a generalized EM algorithm, which, however, moves much more slowly. One main reason for non-convergence of the New-Raphson method, suggested in Seyedsadr & Cornelius (1992), is that  $\hat{\mu}$  is moving towards  $\pm\infty$ . This problem is due to the residual sum of squares, as a function of  $\hat{\mu}$ , approaches an asymptote as  $\hat{\mu}$  tends to  $-\infty$  or  $+\infty$ . To handle this particular issue Piepho (1998) proposes an alternative algorithm, which is more robust but slower. As soon as the shift parameter is estimated, the parameters in the multiplicative term can be found by singular value decomposition of  $\mathbf{Y} - \hat{\mu}$ .

Model (5) can simply be estimated by performing a singular value decomposition of  $\hat{\mathbf{Z}}$ , with  $\hat{z}_{ij} = y_{ij}$ .

#### 4.2. Mixed effects models

Model (6) and (8) can be estimated by utilizing the factor-analytic covariance structure of  $\mathbf{y}$ . Hence, the ML or REML estimated parameters can be found  
100 in SAS by the MIXED procedure with the factor-analytic covariance structure option (Piepho, 1997). Model (7) is harder to estimate, due to the confounding between the mean and the variance. So far, the model parameters can therefore only be estimated by maximum likelihood methods, in e.g. SAS by the NLMIXED procedure or in R by the `mumm` package Pødenphant & Brockhoff  
105 (2018).

#### 4.3. The Template Model Builder R-package TMB

All of the models described in this paper can be fitted by the R-package TMB (Kristensen et al., 2016), which finds the maximum likelihood estimated model parameters. In short, TMB optimizes the Laplace approximation of the  
110 marginal log-likelihood function of nonlinear mixed models. The optimization process is very efficient due to the use of automatic differentiation, which is employed to find the gradient, the Hessian and the third order derivatives of the joint log-likelihood function with respect to the random effect parameters. These values are used to build the Laplace approximation itself and its gradient,  
115 which can be given as input to a standard minimizer in R, e.g. `nlmminb`, which returns the maximum likelihood estimated parameters. However, the TMB is not particularly user-friendly, since the user needs to define the negative joint log-likelihood function in a C++-template. As soon as the C++-template is correctly implemented, the optimization is straightforward, though. Appendix  
120 A includes examples of C++-templates, which define the joint log-likelihood function of some selected biadditive models.

The TMB is employed in R-package `mumm`, which can fit multiplicative models of the type in (7). In other words, `mumm` is basically a wrapper around TMB, such that the user can fit the model in R with a syntax very similar to  
125 the well-known `lmer` function from the `lme4` R-package, and thereby avoids the coding of C++-templates.

The TMB is also interesting for the estimation of the Shifted Multiplicative Model in (4), since it, as described, is complicated to estimate. C++-templates for model (4) and for the Shifted Multiplicative Mixed Model with two multiplicative terms are shown in Appendix B. Extensions to three or more multiplicative terms are straightforward. In Piepho (1998) a data set from the Bundessortenamt (Hannover, Germany) with observations from a fodder beet variety trial is analyzed, where the shift parameter is found by minimization of the residual sum of squares (RSS). This is, however, a bit troublesome for the model with two multiplicative terms, since the RSS as a function of the shift parameter has both a global and a local minimum. For some starting guesses, the algorithm therefore finds the local minimum instead of the global minimum. However, using the TMB, the shift parameter can be found by optimization of the likelihood-function in a very quick (less than 50 ms) and stable manner, without a good starting guess.

## Appendix A. C++-templates

Listing 1: The C++-template for the mixed AMMI model (6).

```

#include <TMB.hpp>

145 template<class Type>
Type objective_function<Type>::operator() ()
{
  /* Data Section*/
  DATA_FACTOR(fixed);
  150 DATA_FACTOR(random);
  DATA_VECTOR(y);

  /* Parameter Section */
  PARAMETER(mu);
  155 PARAMETER_VECTOR(alpha);
  PARAMETER_VECTOR(gamma);
  PARAMETER_VECTOR(beta);
  PARAMETER_VECTOR(delta);
  PARAMETER(log_sigma);
  160 PARAMETER(log_sigma_g); //log(sigma_gamma)
  PARAMETER(transf_rho); //unrestricted version of rho

  Type sigma = exp(log_sigma);
  Type sigma_g = exp(log_sigma_g);
  165 Type rho = transf_rho / sqrt(1. + transf_rho*transf_rho); //-1<rho<1
  vector<Type> alphab(alpha.size()+1);
  vector<Type> betab(beta.size()+1);

  /*alpha and beta are restricted to sum to zero
  * The resulting vectors are named alphab and betab
  */
  170 for(int i=0; i<alpha.size(); i++){
    alphab[i] = alpha[i];
  }
  175 alphab[alpha.size()] = -alpha.sum(); //To insure mean = 0

  for(int i=0; i<beta.size(); i++){
    betab[i] = beta[i];
  }
  180 betab[beta.size()] = -beta.sum(); //To insure mean = 0

  /* Including the altered parameters in the sdreport */
  ADREPORT(sigma);
  ADREPORT(sigma_g);
  185 ADREPORT(alphab);
  ADREPORT(betab);
  ADREPORT(rho);

  /* Joint log-likelihood function */
  190 Type nll = 0;
  int nobs = y.size();
  for(int i=0; i<nobs; i++){
    nll -= dnorm(y[i],
                 mu + alphab[fixed[i]] + gamma[random[i]] +

```

```

195         betab[fixed[i]]*delta[random[i]],
           sigma, true);
    }

    /* The covarinace matrix of gamma and delta */
200    matrix<Type> Sigma(2,2);
    Sigma(0,0) = sigma_g*sigma_g;
    Sigma(1,1) = 1;
    Sigma(0,1) = rho*sigma_g;
    Sigma(1,0) = Sigma(0,1);

205    /* The "random effect part" of the joint log-likelihood */
    for(int k=0; k<gamma.size(); k++) {
        vector<Type> gd(2); gd(0)=gamma(k); gd(1)=delta(k);
        nll += density::MVNORM(Sigma)(gd);
210    }
    return nll;
}

```

## Appendix B. C++-template for the Shifted Multiplicative Model

Listing 2: The C++-template for the shifted multiplicative model with 2 multiplicative terms.

```

215 #include <TMB.hpp>

    /*Function to force a vector's length to 1*/
    template<class Type>
220 vector<Type> vec_length_one (vector<Type> x) {
        int n = x.size();
        vector<Type> xt(n+1);
        xt.setZero();
        xt(0) = 1;
225     for(int i=0; i<n; i++) {
            xt = xt * cos(x(i));
            xt(i+1) = sin(x(i));
        }
        return xt;
230 }

    template<class Type>
    Type objective_function<Type>::operator() ()
    {
235     /* Data Section*/
        DATA_FACTOR(fixed1);
        DATA_FACTOR(fixed2);
        DATA_VECTOR(y);

240     /* Parameter Section */
        PARAMETER(mu);           //shift parameter
        PARAMETER_VECTOR(beta);
        PARAMETER_VECTOR(beta2); //for the 2nd mult term
        PARAMETER_VECTOR(delta);
245     PARAMETER_VECTOR(delta2); //for the 2nd mult term
        PARAMETER(log_sigma);
    }

```



```
250 Type sigma = exp(log_sigma);
vector<Type> deltab(delta.size()+1);
vector<Type> deltab2(delta2.size()+1);

/*The delta vectors are forced to length 1.
 * The resulting vectors are named deltab
 * and deltab2
255 */
deltab = vec_length_one(delta);
deltab2 = vec_length_one(delta2);

/* Including the altered parameters in the sdreport */
260 ADREPORT(sigma);
ADREPORT(deltab);
ADREPORT(deltab2);

/* Joint log-likelihood function */
265 Type nll = 0;
int nobs = y.size();
for(int i=0; i<nobs; i++){
    nll -= dnorm(y[i],
270                mu + beta[fixed1[i]]*deltab[fixed2[i]]+
                beta2[fixed1[i]]*deltab2[fixed2[i]],
                sigma, true);
}
return nll;
275 }
```

## References

- Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account by analyzing profile data with the mixed assessor model. *Food Quality and Preference*, *39*, 156–166.
- 280 Brockhoff, P. M., & Skovgaard, I. M. (1994). Modelling individual differences between assessors in sensory evaluations. *Food Quality and Preference*, *5*, 215–224.
- Denis, J.-B., & Gower, J. C. (1994). Biadditive models. *Biometrics*, *50*, 310–311.
- 285 Denis, J.-B., & Gower, J. C. (1996). Asymptotic confidence regions for biadditive models: interpreting genotype-environment interactions. *Applied Statistics*, (pp. 479–493).

- 
- Eberhart, S. t., & Russell, W. (1966). Stability parameters for comparing varieties 1. *Crop science*, *6*, 36–40.
- 290 Finlay, K., & Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*, 742–754.
- Fisher, R. A., & Mackenzie, W. A. (1923). Studies in crop variation. ii. the manurial response of different potato varieties. *The Journal of Agricultural Science*, *13*, 311–320.
- 295 Gabriel, K. R. (1978). Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 186–196).
- Gauch Jr, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, (pp. 705–715).
- 300 Gauch Jr, H. G. (1992). Statistical analysis of regional yield trials: Ammi analysis of factorial designs. Elsevier.
- Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, *33*, 73–115.
- 305 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, *70*, 1–21. doi:10.18637/jss.v070.i05.
- Oman, S. D. (1991). Multiplicative effects in mixed model analysis of variance. *Biometrika*, *78*, 729–739.
- 310 Piepho, H.-P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics*, *53*, 761–766.
- Piepho, H.-p. (1998). An algorithm for fitting the shifted multiplicative model. *Journal of Statistical Computation and Simulation*, *62*, 29–43.

- Piepho, H.-P. (1999). Fitting a regression model for genotype-by-environment  
315 data on heading dates in grasses by methods for nonlinear mixed models.  
*Biometrics*, *55*, 1120–1128.
- Piepho, H.-P., Denis, J., & van Eeuwijk, F. (1998). Mixed biadditive models. In  
*Proceedings of the 19th International Biometric Conference, Capetown* (pp.  
79–89).
- 320 Piepho, H.-P., Madden, L. V., & Williams, E. R. (2015). Multiplicative inter-  
action in network meta-analysis. *Statistics in Medicine*, *34*, 582–594.
- Pødenphant, S., & Brockhoff, P. B. (2018). *mumm: Multiplicative Mixed Mod-  
els using the Template Model Builder*. URL: [https://CRAN.R-project.  
org/package=mumm](https://CRAN.R-project.org/package=mumm) r package version 0.2.1.
- 325 Pødenphant, S., Kristensen, K., & Brockhoff, P. B. (2018). The multiplica-  
tive mixed model with the mumm r-package as a general and easy random  
interaction model tool. *The American Statistician*, *Submitted*.
- Seyedsadr, M., & Cornelius, P. L. (1992). Shifted multiplicative models for  
nonadditive two-way tables. *Communications in Statistics-Simulation and  
330 Computation*, *21*, 807–832.
- Smith, A., Cullis, B., Brockhoff, P., & Thompson, R. (2003). Multiplicative  
mixed models for the analysis of sensory evaluation data. *Food Quality and  
Preference*, *14*, 387–395.
- Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environ-  
335 ment data using multiplicative mixed models and adjustments for spatial field  
trend. *Biometrics*, *57*, 1138–1147.
- Smith, A., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cul-  
tivar breeding and evaluation trials: an overview of current mixed model  
approaches. *The Journal of Agricultural Science*, *143*, 449–462.
- 340 Yates, F., & Cochran, W. (1938). The analysis of groups of experiments. *The  
Journal of Agricultural Science*, *28*, 556–580.

APPENDIX D

# Reference Manual

---

**Sofie Pødenphant** and Per Bruun Brockhoff (2018). *mumm: Multiplicative Mixed Models using the Template Model Builder*. R package version 0.2.1.  
<https://CRAN.R-project.org/package=mumm>

## Package ‘mumm’

August 15, 2018

**Type** Package

**Title** Multiplicative Mixed Models using the Template Model Builder

**Version** 0.2.1

**Date** 2018-08-14

**Maintainer** Sofie Poedenphant <sofp@dtu.dk>

**Description** Fit multiplicative mixed models using maximum likelihood estimation via the Template Model Builder (TMB), Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM (2016) <doi:10.18637/jss.v070.i05>. One version of the multiplicative mixed model is applied in Piepho (1999) <doi:10.1111/j.0006-341X.1999.01120.x>. The package provides functions for calculating confidence intervals for the model parameters and for performing likelihood ratio tests.

**License** GPL (>= 2)

**Imports** TMB, Rcpp, Matrix, stringr, methods

**Depends** lme4

**LinkingTo** TMB, RcppEigen, Rcpp

**LazyData** TRUE

**RoxygenNote** 6.1.0

**URL** <http://github.com/sofpj/mumm>

**BugReports** <http://github.com/sofpj/mumm/issues>

**Encoding** UTF-8

**Suggests** knitr

**Repository** CRAN

**NeedsCompilation** yes

**Author** Sofie Poedenphant [aut, cre],  
Per Bruun Brockhoff [aut]

**Date/Publication** 2018-08-15 11:00:03 UTC

**R topics documented:**

confint.mumm . . . . .	2
lrt . . . . .	3
mumm . . . . .	4
ranef.mumm . . . . .	6

<b>Index</b>	<b>8</b>
--------------	----------

---

confint.mumm	<i>Confidence Intervals for Model Parameters</i>
--------------	--

---

**Description**

Computes confidence intervals for the fixed effect parameters and the variance components for an object of class `mumm`.

**Usage**

```
## S3 method for class 'mumm'
confint(object, parm = "all", level = 0.95, ...)
```

**Arguments**

<code>object</code>	an object of class <code>mumm</code> .
<code>parm</code>	a vector of parameter names or a matrix, where the rows specify linear combinations of the model parameters. If missing, confidence intervals will be computed for all of the fixed effect parameters and all of the variance components.
<code>level</code>	the confidence level.
<code>...</code>	Currently not used.

**Details**

The confidence intervals are computed by the profile likelihood method.

**Value**

A matrix with the first column showing the lower confidence limit and the second column showing the upper limit for each parameter or linear combination of parameters.

**Examples**

```
set.seed(100)
sigma_e <- 1.5
sigma_a <- 0.8
sigma_b <- 0.5
sigma_d <- 0.7
nu <- c(8.2, 6.2, 2.3, 10.4, 7.5, 1.9)
```

```

nA <- 15
nP <- 6
nR <- 5

a <- rnorm(nA, mean = 0, sd = sigma_a)
b <- rnorm(nA, mean = 0, sd = sigma_b)
d <- rnorm(nA*nP, mean = 0, sd = sigma_d)
e <- rnorm(nA*nP*nR, mean = 0, sd = sigma_e)

Assessor <- factor(rep(seq(1,nA),each = (nP*nR)))
Product <- factor(rep(rep(seq(1,nP),each = nR), nA))
AssessorProduct <- (Assessor:Product)

y <- nu[Product] + a[Assessor] + b[Assessor]*(nu[Product]-mean(nu)) + d[AssessorProduct] + e

sim_data <- data.frame(y, Assessor, Product)

fit <- mumm(y ~ 1 + Product + (1|Assessor) + (1|Assessor:Product) +
            mp(Assessor,Product) ,data = sim_data)

confint(fit, parm = c('Product3', 'mp Assessor:Product'), level = 0.90)

```

---

lrt

*Likelihood Ratio Test*


---

### Description

A function to perform a likelihood ratio test for testing two nested models against each other.

### Usage

```
lrt(fit1, fit2)
```

### Arguments

fit1	a fitted model object of class mumm.
fit2	a fitted model object of class mumm, lm or merMod.

### Details

Performs the likelihood ratio test for testing two nested models against each other. The model in fit2 should be nested within the model in fit1.

### Value

A matrix with the likelihood ratio test statistic and the corresponding p-value.

**Examples**

```

set.seed(100)
sigma_e <- 1.5
sigma_a <- 0.8
sigma_b <- 0.5
sigma_d <- 0.7
nu <- c(8.2, 6.2, 2.3, 10.4, 7.5, 1.9)

nA <- 15
nP <- 6
nR <- 5

a <- rnorm(nA, mean = 0, sd = sigma_a)
b <- rnorm(nA, mean = 0, sd = sigma_b)
d <- rnorm(nA*nP, mean = 0, sd = sigma_d)
e <- rnorm(nA*nP*nR, mean = 0, sd = sigma_e)

Assessor <- factor(rep(seq(1,nA),each = (nP*nR)))
Product <- factor(rep(rep(seq(1,nP),each = nR), nA))
AssessorProduct <- (Assessor:Product)

y <- nu[Product] + a[Assessor] + b[Assessor]*(nu[Product]-mean(nu)) + d[AssessorProduct] + e

sim_data <- data.frame(y, Assessor, Product)

fit <- mumm(y ~ 1 + Product + (1|Assessor) + (1|Assessor:Product) +
           mp(Assessor,Product) ,data = sim_data)

fit2 <- mumm(y ~ 1 + Product + (1|Assessor) + mp(Assessor,Product) ,data = sim_data)
lrt(fit,fit2)

```

mumm

*Fit Multiplicative Mixed Models with TMB***Description**

Fit a multiplicative mixed-effects model to data with use of the Template Model Builder.

**Usage**

```
mumm(formula, data, cor = TRUE, start = c(), control = list())
```

**Arguments**

**formula** a two-sided formula object describing the linear fixed-effects and random-effects part together with the multiplicative part. The response is on the left of a  $\sim$  operator and the terms which are separated by  $+$  operators are on the right. The random-effect terms are recognized by vertical bars " $|$ ", separating an expression



*mumm*

5

	for a model matrix and a grouping factor. The syntax for the multiplicative term is 'mp("random effect","fixed effect")'.
data	a data frame containing the variables in the formula.
cor	logical. If FALSE the random effect in the multiplicative term is assumed to be independent of the corresponding random main effect.
start	a numeric vector of starting values for the parameters in the model.
control	a list of control parameters passed on to the nlmInb function used for the optimization.

**Details**

Fit a multiplicative mixed model via maximum likelihood with use of the Template Model Builder. A multiplicative mixed model is here considered as a model with a linear mixed model part and one multiplicative term. A multiplicative term is here defined as a product of a random effect and a fixed effect, i.e. a term that models a part of the interaction as a random coefficient model based on linear regression on a fixed main effect.

**Value**

An object of class *mumm*.

**Examples**

```

set.seed(100)
sigma_e <- 1.5
sigma_a <- 0.8
sigma_b <- 0.5
sigma_d <- 0.7
nu <- c(8.2, 6.2, 2.3, 10.4, 7.5, 1.9)
nA <- 15
nP <- 6
nR <- 5
a <- rnorm(nA, mean = 0, sd = sigma_a)
b <- rnorm(nA, mean = 0, sd = sigma_b)
d <- rnorm(nA*nP, mean = 0, sd = sigma_d)
e <- rnorm(nA*nP*nR, mean = 0, sd = sigma_e)
Assessor <- factor(rep(seq(1,nA),each = (nP*nR)))
Product <- factor(rep(rep(seq(1,nP),each = nR), nA))
AssessorProduct <- (Assessor:Product)
y <- nu[Product] + a[Assessor] + b[Assessor]*(nu[Product]-mean(nu)) + d[AssessorProduct] + e
sim_data <- data.frame(y, Assessor, Product)
fit <- mumm(y ~ 1 + Product + (1|Assessor) + (1|Assessor:Product) +
  mp(Assessor,Product) ,data = sim_data)

```

---

ranef.mumm	<i>Extract Random Effects</i>
------------	-------------------------------

---

**Description**

A function to extract the estimated random effects from a model object of class `mumm`.

**Usage**

```
## S3 method for class 'mumm'
ranef(object, ...)
```

**Arguments**

<code>object</code>	an object of class "mumm"
<code>...</code>	Currently not used

**Value**

A named list with the estimated random effects, where each element in the list is a numeric vector consisting of the estimated random effect coefficients for a random factor in the model.

**Examples**

```
set.seed(100)
sigma_e <- 1.5
sigma_a <- 0.8
sigma_b <- 0.5
sigma_d <- 0.7
nu <- c(8.2, 6.2, 2.3, 10.4, 7.5, 1.9)

nA <- 15
nP <- 6
nR <- 5

a <- rnorm(nA, mean = 0, sd = sigma_a)
b <- rnorm(nA, mean = 0, sd = sigma_b)
d <- rnorm(nA*nP, mean = 0, sd = sigma_d)
e <- rnorm(nA*nP*nR, mean = 0, sd = sigma_e)

Assessor <- factor(rep(seq(1,nA),each = (nP*nR)))
Product <- factor(rep(rep(seq(1,nP),each = nR), nA))
AssessorProduct <- (Assessor:Product)

y <- nu[Product] + a[Assessor] + b[Assessor]*(nu[Product]-mean(nu)) + d[AssessorProduct] + e

sim_data <- data.frame(y, Assessor, Product)

fit <- mumm(y ~ 1 + Product + (1|Assessor) + (1|Assessor:Product) +
```

---

*ranef.mumm*

7

```
mp(Assessor,Product) ,data = sim_data)  
  
ranef(fit)
```

# Index

`confint.mumm`, 2

`lrt`, 3

`mumm`, 4

`ranef.mumm`, 6