Made available by Hasselt University Library in https://documentserver.uhasselt.be

Statistical models for analyzing repeated quality measurements of horticultural products. Model evaluation and practical example.

Peer-reviewed author version

De Ketelaere, Bart; Lammertyn, Jeroen; MOLENBERGHS, Geert; Nicolai, Bart M. & De Baerdemaker, Josse (2003) Statistical models for analyzing repeated quality measurements of horticultural products. Model evaluation and practical example.. In: Mathematical Biosciences, 185(2). p. 169-189.

DOI: 10.1016/S0025-5564(03)00092-0
Handle: http://hdl.handle.net/1942/432

Statistical models for analyzing repeated quality measurements of horticultural products. Model evaluations and practical example.

Bart De Ketelaere[1], Jeroen Lammertyn[2], Geert Molenberghs[3], Bart Nicolaï[2] and Josse De Baerdemaeker[1]

[1] K.U.Leuven, Department of Agro-Engineering and -Economics, Laboratory for Agricultural Machinery and –Processing; Kasteelpark Arenberg 30, 3001 Leuven, Belgium;

[2] K.U.Leuven, Department of Agro-Engineering and –Economics, Laboratory of Postharvest Technology; de Croylaan 42, 3001 Heverlee, Belgium;

[3] Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, 3590 Diepenbeek, Belgium

corresponding author: Bart De Ketelaere; K.U.Leuven, Department of Agro-Engineering and –Economics, Laboratory for Agricultural Machinery and –Processing; Kasteelpark Arenberg 30, 3001 Leuven, Belgium;

e–mail: bart.deketelaere@agr.kuleuven.ac.be;

tel: +32(0)16 32 85 93

fax: +32(0)16 32 85 90

Abstract

Four different types of statistical models used to analyze repeated measures are discussed and compared. Repeated measures analysis is gaining importance during recent years and several software packages offer a broad class of routines. In the field of postharvest quality assessment of horticultural products, research on the development of non–destructive quality sensors, replacing destructive and often time consuming sensors, has spurred in the last decennium offering the possibility of taking repeated quality measures on the same product. A dataset dealing with the postharvest quality evolution of different tomato cultivars serves as practical example for model comparisons. Starting from an analysis at each time point and an ordinary least squares regression model as standard and widely used methods, this contribution aims at comparing these two methods to a repeated measures analysis and a longitudinal mixed model. It is shown that the flexibility of such a mixed model, both towards the repeated measures design of the experiments as towards the large product variability inherent to these horticultural products, is an important advantage over classical techniques. This research shows that different conclusions could be drawn depending on which technique is used due to the basic assumptions of each model and which are not always fulfilled. The results further demonstrate the flexibility of the mixed model concept. Using a mixed model for repeated measures, the different sources of variability, being inter–tomato variability, intra–tomato variability and measurement error were characterized being of great benefit to the researcher.

Keywords: statistical models, repeated measures, product variability, mixed models, tomato quality.

1. Introduction

In the field of applied sciences, one is often confronted with correlated data. The term correlated data embraces a multitude of data structures, such as multivariate observations, clustered data, repeated measurements, longitudinal data and spatially correlated data [1]. Although multivariate analysis techniques have received most attention in literature, repeated measures analysis has gained much attention during recent years. The term repeated measures points to data structures where multiple measurements are obtained from a single experimental unit. This experimental unit can be, for instance, a family and a certain parameter is measured for all its members. As another example, repetitions can be made over a certain period of time for each subject. In this case, the term longitudinal data is often used. When repeated measures of each subject are taken on different locations the term spatial data applies.

In this contribution, we will focus on longitudinal data as a subclass of repeated data. In order to make the different models and their comparisons more interpretable, a practical dataset in the field of postharvest crop monitoring will be used. In this sector, quality inspection and classification of products are of great need in the modern market, where large quantities are sold within seconds, sometimes without access of the buyers to see the product. The conventional quality inspection often involves destructive and / or time consuming measurements and may be applied only to small samples of large shipments. High quality standards and the necessity for shelf–life determination have increased the need for simple and quick evaluation of the internal properties of each product sold, preferably making use of non–destructive devices that ‘sense’ the product’s quality attributes such as firmness or flavor

[2]. One of the most important advantages of these non–destructive measurement techniques, besides their objective nature, is the possibility they offer towards monitoring individual products during the experimental period, which on its turn allows for modeling the quality change. The modeling of the quality evolution of horticultural products during storage has been described in literature by several authors [3–9]. In these contributions, two different approaches for modeling the repeated quality measures over time can be distinguished.

A *first* approach makes use of the analysis of the data at each time point separately [3, 7, 8]. For instance, at each measuring day the average quality attribute is calculated, and these means are compared. This approach allows a simple interpretation of the data and allows easily communication to non–statisticians, but it does not consider overall differences since only one time point is analyzed at a time. Consequently, the method does not allow studying the evolution of the quality during storage, which is, however, of prime interest in many experiments. A *second* approach makes use of an ordinary least squares (OLS) regression model to study the quality evolution. The advantage of the OLS regression approach is that it is easily implemented in standard software and that it allows a prediction of the time at which a batch of products reaches a pre–set lower bound of the quality parameter of interest. The latter was not possible in case of the analysis at each time point. The disadvantage of such analysis, but also of the analysis at each time point, is that it does not take into account the repeated nature of the data – it naively treats observations across time as independent – affecting significance levels of estimated parameters. In the case of biological specimen, such as fruits, this is reinforced by the fact that biological material exhibits a large natural variation in quality and this subject specific variability is not accounted for in such models. For instance, Thai et al. [4] remarked that the fit of their model decreased considerably when modeling a batch of tomatoes, compared to the modeling of the individual tomato profiles. As such, the amount of *unexplained* variability in their data increased due to this batch heterogeneity inherent to biological produce. The presence of a large inter–subject variance combined with the negation of the covariance structure of the repeated measures could lead to wrong conclusions. Moreover, these models inherently assume that the variance of the data remains constant over time (homoscedasticity). From research of several authors it can be questioned whether this homoscedasticity assumption is valid [7, 10, 11]. When repeats are available for the quality measure at each time point for each product, which is often the case, yet another point is the question whether it is advisable to use only the average quality measure of a single product in modeling its behavior during storage or to use all available measurements, which allows not only the estimation of the variance of the measurements on a single subject during storage, but also takes into account this variation – and its possible dependence on the quality measure – when estimating a model's parameters. This could be an important factor since the reliability of a quality sensor could depend the quality measure. The repeated measurements nature of such data, their heteroscedasticity combined with the large natural variation in quality of biological products raise the question whether the proposed analysis methods in literature describe the data adequately.

Laird and Ware (1982) proposed a statistical model that allows for a subject–specific effect above a population–specific effect [12]. These subject–specific regression parameters reflect the natural heterogeneity in the population and can also be interpreted as the deviation of the evolution of a specific subject from the overall population. For this reason they are usually assumed to follow a Gaussian distribution. Their mean then reflects the average evolution in the population, and is therefore called the vector of fixed effects. The assumption of a Gaussian distribution is not only intuitive, but is also mathematically convenient [1, 13]. This type of models are called mixed–effects models and are appropriate for data that exhibit a large inter–subject variability, as is expected for measurements on biological produce. Furthermore, the incorporation of, for instance, a subject–specific time trend allows for heteroscedasticity of the data. In the same context, this general framework was further broadened to allow for repeated measurements [1, 13]. The availability of the MIXED procedure in the SAS software [14] provides a broad class of linear mixed–effects models readily available for routine use, and such models allow to compensate for the shortcomings of analyses at each time point and ordinary least squares regression models.

These different types of models were used and compared to analyze tomato firmness during a two–week storage experiment. Such data are characterized by two main specific characteristics, being (1) the natural variability caused by the biological products and (2) the repeated measures design. This work shows that the specific data nature of such studies requires a specific data analysis that goes beyond classical techniques such as an analysis at each time point and an ordinary least squares regression model.

The objective of this paper is to provide an overview and comparison of methods with a clear description of the assumptions that are inherent to these methods.

2. Materials and Methods

2.1 Tomato firmness data

Tomatoes of 13 different varieties were harvested and their firmness was followed during 2 weeks of storage. Tomatoes came from two different research stations, namely ‘Proefbedrijf der Noorderkempen’ (Experimental farm of the Noorderkempen region) at Meerle and

‘Proefstation voor de groenteteelt’ (Vegetable research station) at Sint–Katelijne Waver, both situated in Belgium. Three varieties are commercial (Quest, Mariachi and Tradiro), with Tradiro tomatoes coming both from Meerle (coded as TradiroM) and Sint–Katelijne Waver (coded as TradiroSKW). In subsequent results, TradiroSKW and TradiroM tomatoes were analyzed separately as two different varieties, and were compared. All varieties came from Meerle except Tradiro tomatoes, which came from both stations. The data were measured at the Flanders Centre for Postharvest Technology (Leuven, Belgium).

For each variety, 20 tomatoes were analyzed for two harvest periods, August and October. Tomatoes of both harvest periods came from the same plants. Tomatoes were harvested twice a week, with tomatoes used in this study originating all from the same harvest day. For each harvest period, measurements were taken at harvest (day 0), day 3, 5, 7, 10, 12 and 14 of storage. Tomatoes were stored at controlled atmosphere conditions (18 °C and 80 % RH) to accelerate the ripening process.

Tomato firmness was assessed using a commercial acoustic firmness tester (AWETA, Nootdorp, The Netherlands). The device produces a stiffness index S as indicator for fruit firmness. Stiffness was measured three times at the south pole of the tomatoes for each measurement day. Both the average stiffness as the individual measurements were used throughout further analyses. In the remainder of the text, the term stiffness will be used when indicating values produced by the acoustic tester and which are an estimate of the firmness. An overview of the data is provided by figures 1 and 2.

The starting point for the analyses where time is treated as a continuous variable is the first order degradation model most widely found in literature [15–16]. The solution of the first–order degradation model is given by

$$S(t) = S_0 e^{-\alpha t} \quad (1)$$

where $S(t)$ denotes the stiffness factor at time t, S_0 the initial stiffness ($\times\ 10^6 Hz^2 g^{2/3}$) and α the exponential decay factor (day^{-1}). In all analyses that follow, the natural logarithm of the stiffness was used in order to linearize the data as follows

$$s(t) \equiv ln(S(t)) = s_0 - \alpha t \quad (2)$$

where s_0 is defined as the natural logarithm of the initial stiffness S_0. This first order degradation model will be tested throughout the analysis against more complex models that consider also a quadratic time trend of $s(t)$.

2.2 Statistical methods

The SAS software (SAS version 8.2, The SAS Institute Inc., Cary, NC, USA) was used throughout all analyses. The data were modeled using 4 types of models of which the first two, an analysis at each time point and the ordinary least squares model are widely spread in literature (see introduction section). The third model presents a repeated measures analysis, while the last model includes random effects that are subject–specific.

2.2.1 Analysis at each time point

The first type of model consists of an analysis at each time point where a separate mean is fitted for each experimental setting. Inherently, time is considered as being a categorical variable. This is the type of analysis that is often found in literature concerning the postharvest treatment of horticultural produce and is given in the case of three main effects (for instance storage time δ, tomato cultivar τ and harvest λ)

$$Y_{ijkl} = \mu + \delta_i + \tau_j + \lambda_k + (\delta\,\tau\,\lambda)_{ijk} + \varepsilon_{ijkl} \quad (3)$$

where Y_{ijkl} refers to the response of subject l at storage time δ_i, belonging to cultivar τ_j and harvested at λ_k; μ is the overall mean and δ , τ and λ are three main effects with their interaction $(\delta\ \tau\ \lambda)$ and ε_{ijkl} the error term. Inferences about different behavior of different tomato cultivars are limited to each time point separately and are accomplished using a Tukey multiple comparison test.

2.2.2 Ordinary least squares regression model

A second type of model is an ordinary least squares (OLS) regression model given in its general form by

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}\,, \quad (4)$$

with $\boldsymbol{Y_i}$ the n_i–dimensional vector of all repeated measurements for the i–th subject (the repeated stiffness measures for a single tomato), X_i the appropriate ($n_i \times p$) matrix of known covariates (for instance cultivar and / or storage time); $\boldsymbol{\beta}$ a ($p \times 1$) vector of fixed effects and $\boldsymbol{\varepsilon_i}$ the vector of residual components ε_{ij}, $j = 1,\ldots,n_i$. It is stressed at this point that n_i refers to the the number of repeated measures for a subject i. In this setting, the error terms ε_{ij} are assumed to be independently and identically distributed with mean zero and variance σ^2. More precise, the error vector $\boldsymbol{\varepsilon_i}$ is assumed to be normally distributed with a zero mean vector and variance–covariance matrix equal to $\sigma^2 I$ with I the identity matrix of size ($n_i \times n_i$). As such, the two main assumptions of this model are (1) independence of all measurements and (2) homoscedasticity.

2.2.3 Repeated measures analysis

A third type of model allows the error terms of the repeated measures of a given subject to be dependent on each other and is of the same general form as equation 4. In contrast to that model, the components ε_{ij} are not independently and identically distributed with variance σ^2 but, in general, we can write that the $\boldsymbol{\varepsilon_i}$'s are normally distributed with a zero mean vector and variance covariance matrix Σ. Writing the distribution for $\boldsymbol{Y_i}$ as $N(X_i\boldsymbol{\beta}, V)$, the V matrix in its most general form is a $(n_i \times n_i)$ unstructured matrix. However, the number of parameters that have to be estimated for such general structure increases rapidly in function of the number of time points for each subject n_i and a simplification of the variance–covariance structure is often assumed. These simplified structures are borrowed from time series analysis that proposes a broad range of possibilities. A widely used structure for the V matrix is a first–order autoregressive structure. For a (3×3) variance–covariance matrix – denoting that three repeats were taken of each subject – this structure is given by

$$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix} \qquad (5)$$

where the σ parameters are used to denote variances and covariances, whereas the ρ parameters are used for correlations. This structure assumes that the correlation between measurements depends on the difference in time between them. When the variances are allowed to change as a function of time, this structure becomes

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \qquad (6)$$

and is called a heterogeneous first–order autoregressive structure. In this case σ^2_1 denotes the variance at the first measurement day, σ^2_2 at the second day, and so on. The selection of an appropriate covariance structure for the data can be made by minimizing Akaike's information criterion (AIC). The AIC is a function of the log–likelihood with a penalty for the number of parameters that were estimated [17].

2.2.4 Linear mixed model for longitudinal data

The fourth approach uses a longitudinal linear mixed model, defined as [11]:

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}\,, \qquad (7)$$

with $\boldsymbol{Y_i}$ the n_i–dimensional vector of all repeated measurements for the i–th subject (tomato); X_i the $(n_i \times p)$ design matrix of known covariates; $\boldsymbol{\beta}$ a $(p \times 1)$ vector of fixed effects; Z_i a $(n_i \times q)$ matrix of known covariates (for instance storage time) modeling how the response evolves over time for the i–th subject; $\boldsymbol{b_i}$ a $(q \times 1)$ vector of subject specific effects for which is assumed that $E(\boldsymbol{b_i}) = 0$ and $\boldsymbol{\varepsilon_i}$ the vector of residual components ε_{ij}, $j = 1,\ldots,n_i$. The random effects structure implies a covariance structure of a very specific form

$$Var(\boldsymbol{Y_i}) = V_i = \boldsymbol{Z_i}D\boldsymbol{Z_i}^T + \Sigma_i \qquad (8)$$

where D refers to the variance–covariance matrix of the random effects. It can be seen that the total covariance structure is partly determined by the $\boldsymbol{Z_i}$ vector of random effects and partly by the error variance–covariance matrix Σ_i. Since $\boldsymbol{Z_i}D\boldsymbol{Z_i}^{\boldsymbol{T}}$ mostly accounts for a large part of the variation in V_i, the structure of Σ_i is often assumed to be of the form $\sigma^2 I$ with I the identity matrix of size $(n_i \times n_i)$. Remark that this structure allows for heteroscedasticity as function of time, for instance when one assumes random intercepts and slopes so that $\boldsymbol{Z_i}$

equals [1 t]. The variance encountered here is referred to as *inter–subject* variance, indicating that the variance in response among subjects (tomatoes within a cultivar for instance) could be a function of time.

For models with only few random effects, choosing a simple ($\sigma^2 I$) variance–covariance structure Σ_i may prove to be an over–simplification. Where there is no evidence for the presence of additional random effects, or when random effects have no substantive meaning, the covariance assumption can be relaxed by allowing an appropriate, more general residual covariance structure Σ_i for the vector ε_i of subject–specific error components. The ε_i's are decomposed as $\varepsilon_{(1)i} + \varepsilon_{(2)i}$ with $\varepsilon_{(1)i}$ denoting the component of measurement error ($\sim N(\mathbf{0}, \sigma^2 I)$) and $\varepsilon_{(2)i}$ ($\sim N(\mathbf{0}, \sigma^2 H_i)$) the component of serial correlation [13]. Two examples of these functions are the Gaussian and exponential serial correlation functions determining the serial correlation matrix H_i. For a selection of the residual covariance structure Σ_i of the error components ε_i , Akaike's Information Criterion (AIC) can be used. Inclusion of such serial correlation should only be considered when using models having only random intercepts since the effect of such serial correlation is very often dominated by the combination of random effects and measurement error [18].

As an extension to this mixed model approach in the case of repeats for each subject at a given time point, all available measurements instead of the average response of each subject at each time point could be considered. In doing so, the model allows to estimate the variance of the measurements within a single subject, which was impossible when only the mean response at each time point was considered. This extra variance component will be referred to as *intra–subject* variance. The inclusion of such intra–subject variance is assured by adding a diagonal matrix to the error variance matrix Σ_i that is allowed to change as a function of storage time. These local effects have the form σ^2_o *diag[exp($U_i\delta$)]* where U_i is the full–rank design matrix corresponding to effects entered in the log–linear variance model for subject *i*, δ is the vector of parameters that are estimated and σ^2_0 is an estimate of the initial intra–subject variance. If one includes a time effect into the U_i matrix, the repeatability of measurements during the experimental period can be modeled. The exponential function ensures non–negative components for the variance. In case of a simple residual error variance–covariance matrix, the total measurement error variance–covariance matrix Σ_i can be written as

$$\Sigma_i = \sigma^2 I + \sigma^2_0 \; diag[exp(U_i\delta)] = diag[\sigma^2 + \sigma^2_0 \, exp(U_i\delta)] \qquad (9)$$

The model building process for mixed models is more complicated in this case than it is in ordinary regression since the model copes with a mean structure, a covariance structure and a random effects structure, all of which are not independent of each other. The scheme presented in figure 1 can be used in order to find the final model (after [13]).

For *fixed* effects model building purposes, maximum likelihood estimation (ML) should be used rather than the default restricted maximum likelihood estimation (REML), which is the standard setting in the MIXED procedure of SAS. This allows nested models to be compared with a likelihood ratio test defined as

$$G^2 = -2\ln\left[\frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})}\right], \qquad (10)$$

where L_{ML} denotes the ML likelihood function and $\hat{\theta}_{ML,0}$ the parameters estimated under ML for a model 0 and being a subset of the parameters $\hat{\theta}_{ML}$; G^2 then follows, asymptotically, under H_0 a chi–squared (χ^2) distribution with degrees of freedom equal to the difference between the dimensions $v - u$ of $\hat{\theta}_{ML,0}$ and $\hat{\theta}_{ML}$. In the results section, the log likelihood is denoted by the symbol ℓ.

To test whether *random* effects are needed in the model, the likelihood ratio test defined above was used but follows asymptotically a null distribution that is a mixture of chi–squared distributions, rather than the classical single chi–squared distribution that was used to test fixed effects [13]. For the case of testing no random effect versus one random effect, the null distribution is a mixture of χ_1^2 and χ_0^2 with equal weights 0.5, denoted by $\chi_{0:1}^2$. In case of testing one versus two random effects, the null distribution is a mixture of χ_2^2 and χ_1^2 distributions with equal weights 0.5, denoted by $\chi_{1:2}^2$.

For comparing different *variance–covariance structures*, the likelihood ratio test G^2 defined in (10) can be used to compare nested models. For unnested models, one does not get a formal testing procedure anymore and hence the AIC should be used. Once the final model is obtained, parameter estimates and standard errors should be computed under restricted maximum likelihood estimation in order to have trustworthy estimates of the variance components.

3. Model comparisons

The different assumptions that are inherent to the four models described above are summarized in table 1.

The analysis at each time point provides a model that allows easy interpretation and visualization of its parameters. Its most important drawback is given by the fact that it completely ignores the repeated measures design of the experiment, and that inferences are only restricted to those storage times at which measurements (and thus means) are at hand. This further implies that evolution differences cannot be treated.

The OLS regression model treats the time effect a continuous variable, being more realistic than the categorical assumption in the analysis at each time point. As such, it allows estimating the quality *evolution* under the strict assumptions that any two measurements are independent of each other, and that data variability remains constant over storage time. Both assumptions are highly questionable and put the use of this type of model open to discussion. From literature, it appears that the homoscedasticity assumption of the data is often not valid in storage experiments [6, 9, 10] who all encounter time dependent data variability. Even more questionable is the assumption of independence of error terms since two measurements taken on the same subject will be related. This (wrong) assumption of the OLS model has no important consequence on the estimated parameters since they are asymptotically consistent [19], but has a very important consequence on their standard errors, and hence on parameter significance. Due to the negation of the dependency among measurements of a given subject the information contained in the data set is overestimated leading to smaller but inconsistent standard errors. Model fits are often evaluated by looking at parameter variances, with a preference for low parameter variances, but this is only valid under justifiable model assumptions. A possible way around the dependence of variance estimates on these model assumptions is a using robust sandwich variance estimator for the OLS model [20]. However, this option does not present the flexibility of other methods discussed below.

The repeated measures analysis of the data relaxes the assumption of independent error terms and, for some covariance structures such as a heterogeneous covariance structure, even the homoscedasticity assumption. Comparison of the model fit of a repeated measures analysis to that of an ordinary least squares regression model with the same fixed effects structure can easily be obtained using the likelihood ratio test, since both models are nested. A heterogeneous covariance structure has its limiting use in case of a large number of repeats for each subject since it models a different variance parameter for each time point in the analysis, as was shown for instance in (6). As for the OLS model, the source of the variance cannot be split into the different sub–contributions due to, for instance product variability and measurement errors which, again, is a severe drawback of these models since interest of researchers could lay in quantification of homogeneity of the batch (inter–subject variance), and researchers and companies manufacturing quality devices could put emphasis on the repeatability of their equipment (intra–subject variance). These drawbacks are perhaps the most important stimuli to prefer the mixed model approach above higher–mentioned approaches in case of horticultural products.

The linear mixed model for longitudinal data provides a very flexible tool for analysing repeated measures on horticultural products. It offers the possibility to account for the repeated measures nature of the data, the product variability inherent to those horticultural products and offers a broad spectrum of variance–covariance structures by the inclusion of subject specific parameters. Furthermore, the data variance can be split into the desired components that were named inter– and intra–subject variances.

4. Results

4.1 Analysis at each time point (model 0)

At each time point of the storage experiment a separate mean is fit to the data for each of the 28 (2 × 14) combinations for harvest and cultivar. A graphical view on this model was already given in figures 2 and 3. The model has a -2ℓ of –4567.1 using 196 (14 × 2 × 7) parameters.

This kind of analysis does not allow comparing the stiffness evolution of different varieties, but only allows comparing results for each time point separately. For instance, table 2 gives the tomato variety grouping for 0 and 14 days of storage for the August data. Varieties sharing the same letter for a given day are not different at a 5 % significance level.

The table shows that at harvest (day 0), *RZ7457* tomatoes had a significant higher stiffness than all other varieties. For the other varieties, the differences are much smaller resulting in no clear separate groupings with most of the varieties showing no significant difference. For instance at harvest, 8 varieties ranging from *TradiroM* to *Mariachi* show no significant difference. At the end of storage (day 14), the difference between varieties *RZ7457* and *DRW6391* is not significant anymore. Again, for the other varieties differences are much smaller.

For some varieties, such as *Quest* and *TradiroM* and which are both commercial, no significant difference was found for neither of the storage times for the August data (data for 3, 5, 7, 10 and 12 days of storage not shown). However, inspecting figure 2 it appears that both varieties behave different during storage: while *TradiroM* has a higher initial stiffness than *Quest*, it shows a larger decrease in stiffness during storage, resulting in a lower stiffness than *Quest* tomatoes at the end of study. This last remark stresses the weak point of such analysis at each time point: while two varieties may not show any statistical difference at any point, it still might be possible that their overall stiffness evolution over time is different. This will be further analysed in the subsequent analyses.

4.2 Ordinary least squares regression model (model 1)

A model treating storage time as a continuous variable replaces the unstructured mean of the previous model 0. A separate intercept, linear as well as quadratic time trend for each variety × harvest combination are taken as starting point for the analysis. This model has $3 \times 2 \times 14 = 84$ parameters and has a -2ℓ of –4483.2 leading to a value for the LRT test G^2 of 83.9 on 112 degrees of freedom, clearly favoring this simpler model above the over–elaborated model 0 ($P = 0.9781$). The quadratic time trends for each harvest × variety combination were not overall significant ($P = 0.3000$) and were removed from the model. Similarly, the quadratic term for each harvest, for each variety and the overall quadratic term were removed, leading to model 1 with only an intercept and linear time trends for each variety × harvest combination and a -2ℓ of –4451.5 using 56 parameters, which is preferable above the starting model in this paragraph ($G^2 = 31.7$ on 28 DF, $P = 0.2869$). Common slopes for each harvest or tomato variety could not be used ($P < 0.0001$). Table 3 lists the parameter estimates and the standard errors obtained under restricted maximum likelihood estimation. For simplicity, parameter estimates are only given for one tomato variety (*DRW5730*, August); standard errors hold for all tomato varieties. Standard errors of this model will be compared to later models.

Figure 4 presents the variance estimate provided by a spline–smooth curve based on the squared residuals (residuals not shown), together with the modeled variance function, which is assumed to be constant over storage time and equals 0.01835. This plot already indicates that the assumptions made in this model are highly questionable and need further investigation. This will be the topic of next two paragraphs where the two assumptions made by the OLS model will be relaxed.

A contrast statement was constructed in order to compare *Quest* tomatoes to *TradiroM* tomatoes for the August data, as was done in the analysis at each time point. The contrast statement simultaneously tests for equal intercept and slope for both varieties and rejects the null hypothesis of equal behavior of both varieties (P = 0.0135). This simple example already indicates the advantage of treating the storage time as a continuous variable since intuitively, by inspecting figure 2, one would indeed be tempted to conclude a different behavior for both varieties.

4.3 Repeated measures analysis (model 2)

Instead of treating all individual measurements as independent, measures taken on 1 tomato are now allowed to depend on each other which is a much more realistic approach. Several covariance structures were tested, with the first order heterogeneous autoregressive structure ARH(1) found to be the most plausible solution indicating that different variances for each time point occur in the data, with long storage inducing larger variances.

The results of this model are given in table 3 under models 2a and 2b. From this table it may be noted that the inclusion of quadratic time trends for each variety–harvest combination (model 2b) is preferable over including only linear time trends given by model 2a ($G^2 = 89.1$, DF = 28, $P < 0.0001$).

Figure 5 gives the spline–smooth of the variance function based on the OLS residuals (full line) together with the modeled variances σ^2_i ($i = 1, \ldots, 7$) using model 2b. It may be stressed that the full line is *not* a fit of the σ^2_i 's (given by triangles in the figure) by this repeated measures analysis. Comparing figures 4 and 5 leads to the conclusion that allowing for heteroscedasticity provides a more plausible variance modeling.

Comparing model 2a to the OLS model 1, it is seen that the model fit improves drastically with a -2ℓ of –7615.5 for model 2a versus –4451.5 for model 1. The reader may, however, not be mislead by the increase in standard error for the repeated measures analysis. Indeed, in model 1 the assumption was made that each point in the analysis was independent from all others, while model 2a proved that not all variability is independent leading to the larger but correctly estimated standard errors. This incorrect independence assumption of the OLS model results in estimators that are not consistent and thus a comparison of standard errors is not justified. On the other hand, when constructing difference estimates, these differences could be much more pronounced, although this was not the case in comparing *TradiroM* tomatoes to *Quest* tomatoes in August ($P = 0.1047$) leading to the conclusion that both profiles are not significantly different, in contrast to the conclusion that was drawn from the OLS model. This last fact again clearly stresses the importance of model assumptions on the significance of parameters and hence on the conclusions drawn from the study.

4.4 Linear mixed model for longitudinal data (models 3 and 4)

The residuals of the OLS model $\boldsymbol{r_i}^{\boldsymbol{OLS}}$, describing the remaining variability of the data that is not explained by the model are, under the assumptions of the OLS model, constant over time. However, inspecting for instance figure 4 presenting the OLS residuals, it can be concluded that this variability is not constant over time, but exhibits a quadratic pattern. The linear mixed model discussed here allows for heteroscedasticity in time, as the residuals of the OLS model can be written as

$$\boldsymbol{r_i}^{\boldsymbol{OLS}} \approx Z_i \boldsymbol{b_i} + \varepsilon_i \tag{11}$$

Indeed, the covariance between any two points t_1 and t_2 can be written as follows for the case of random intercept and slope:

$$\begin{aligned} Cov(\boldsymbol{Y}_i(t_1), \boldsymbol{Y}_i(t_2)) &= \begin{pmatrix} 1 & t_1 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \end{pmatrix} + \sigma^2 \\ &= d_{22} t_1 t_2 + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \end{aligned} \tag{12}$$

with D the variance–covariance matrix of the random effects. It implies that the variance function of the response behaves quadratically over time with positive curvature d_{22} so that figure 4 points to the possible inclusion of a random slope into the model.

Since the covariance structure models all variability not explained by the fixed effects, all systematic trends need to be removed first. For this purpose, the parameters used in model 1 were taken as a preliminary mean structure.

Random effects are now added to the model, which can be interpreted as subject–specific corrections to the overall mean structure. For the inclusion of random effects, it is favoured to include too many random effects instead of too few to ensure that the remaining variability is not due to missing random effects [12]. Since the model assumes random effects $\boldsymbol{b_i}$ to have zero mean, we consider only covariates Z_i that have already appeared in the fixed part X_i. For this reason, random intercepts and random slopes were used as starting point together with the preliminary mean structure of model 1. The random–effects variance matrix D was assumed to be unstructured.

The -2ℓ of this model 3c is –7643.4, which is an improvement of model 1. Deleting the random slopes from previous model gives a -2ℓ of –7275.0, clearly favouring the inclusion of random slopes ($G^2 = 368.4$, DF = 1:2[1], $P < 0.0001$).

For this preliminary mean structure – the mean structure of the OLS model that included an intercept and slope for each harvest × variety combination – the possible inclusion of a serial correlation was investigated in case of only a random intercept. A Gaussian and exponential serial correlation were tested, replacing the simple covariance structure. Both serial correlations make use of two parameters, whereas the simple structure uses only 1 parameter. The exponential serial correlation was the most suitable to describe the data, leading to an AIC of –7470.2 which is inferior to the model that includes a random intercept and slope together with a simple covariance structure.

With the extended covariance structure that was modelled by the inclusion of random intercepts and slopes (model 3c) and that captured a large amount of the variation in the data, it was investigated whether the preliminary mean structure still holds. In a first step a quadratic time effect for each variety × harvest combination was included into the mean structure and resulted in a -2ℓ of –7780.7 (model 3f), which clearly is better than model 3c ($G^2 = 137.3$, DF = 28, $P < 0.0001$). Common quadratic time effects for harvest or variety did not improve the model ($P < 0.0001$). While quadratic profiles as function of time were found to be not significant in the OLS model without random effects (P = 0.2854), they turn up in this mixed model (where product variability and repeated measures are included in the model) to be highly significant (P < 0.0001).

The extended mean structure of model 3f – including an intercept, slope and quadratic effect for each harvest × variety combination – was used on its turn to investigate whether the random structure needs further adjustment. Since quadratic profiles are included in the mean structure, it was investigated whether a random quadratic effect further improves the model fit (model 3g). Again, the random–effects variance matrix D was assumed to have an unstructured form. The -2ℓ of this model is –7884.4 and hence is to be preferred above model 3f (G^2 = 103.7, DF = 2:3, P < 0.0001). Since the mean structure still holds with this random structure, model 3g was taken as final model. Figure 6 shows the modelled variance and the spline–smoothed variance obtained using OLS residuals. Both variance functions show a rather similar pattern indicating that this final model is capable of tracking the data variance.

Constructing a contrast statement to test for a significant stiffness profile for *TradiroM* versus *Quest* tomatoes in August reveals a significant difference between both varieties using model 3g (P = 0.0098).

A next step in the analysis is the inclusion of the three measurements taken on each tomato at each time point, instead of its mean value used throughout models 3. The same model building steps were followed as described above, leading to a model that includes an intercept, slope and quadratic trend for each harvest × variety combination, together with a random intercept, slope and quadratic trend. The stepwise selection of a fixed, random and covariance of this model is not discussed in this text but is analogous to that applied to obtain model 3g. The final model using three measurements on each tomato at each time (model 4g) has a -2ℓ of –20153.8. Comparing models 3g and 4g shows (table 3) that the parameter estimates are more precise if three measurements are taken. This was to be expected since the number of measurements on which parameter estimates are based are triple–fold, but the increase in data variability is only moderate since the three measurements taken on each tomato are correlated.

Model 4g allows the inter–tomato variability to increase as function of storage time, but assumes the intra–tomato variability to be constant over time. This last assumption is now relaxed in model 4h where a diagonal matrix of the form as $diag[\sigma^2_o exp(U_i\delta)]$ is added to the variance covariance matrix Σ_i. A preliminary choice for the design matrix U_i was made using the squared residuals of a model with an unstructured mean for each tomato at each time point. Such model has no practical sense and was only used for this data–exploration purpose because it allows estimating the intra–tomato variance we are interested in. A spline–smooth estimate of the variance (not shown) followed a positive quadratic pattern as function of storage time. The variance remains more or less constant for the first week of storage but then increases rapidly in the second week of storage. A model allowing a quadratic increase in intra–tomato variance will be used as starting point. As mean structure, the factors used in model 4g are taken. This model 4h has a -2ℓ of –21437.2 (AIC = –21249.2), which is a clear improvement over model 4g, using only 3 extra parameters (σ^2_o, δ_1 *and* δ_2). Restricting the intra–tomato variance to behave linear increases the -2ℓ to –21145.9 (AIC = –20959.9) and did not improve the model fit. At harvest, the variance of the three measurements is 0.0017 units on logarithmic stiffness scale and increases almost four–fold to 0.0066 after two weeks of storage. For instance, for a tomato with stiffness $8 \times 10^6 Hz^2 g^{2/3}$ at harvest, the 95 %

confidence limits for its stiffness are 7.68 to 8.32 × $10^6 Hz^2 g^{2/3}$. For a tomato with a stiffness of 5 × $10^6 Hz^2 g^{2/3}$ after two weeks of storage, its 95 % confidence limits are 4.40 and 5.60 × $10^6 Hz^2 g^{2/3}$, a substantially broader interval than at harvest. These 95% confidence limits are only based on the intra–tomato variance $\sigma^2{}_o \, diag[exp(U_i \delta)]$; as such, they provide an estimate of the repeatability of the acoustic firmness tester. The course of this repeatability as function of storage time is given in figure 7, together with the intra–tomato variance $\boldsymbol{Z_i} D \boldsymbol{Z_i}^T$ and the residual error variance σ^2. At harvest, the repeatability of the measurements is good and the intra–tomato variance only accounts for a minor part in the error variance $diag(\Sigma_i)$. In this example the random effects account for most of the data variance. In other words, most of the variability in the data is due to the different behaviour of the different tomatoes belonging to a given variety since the random effects describe the dispersion of the profiles around their fixed effects that were modelled as separate quadratic trends for each variety.

5. Discussion

The analysis at each time point provides a quick view on the differences that occurred between tomato varieties at each time point. However, an overall comparison of tomato varieties could not readily be obtained, rendering the interpretation of firmness evolution more cumbersome.

The OLS regression model treats the storage time as a continuous variable, being more realistic than the categorical assumption in the analysis at each time point. As such, it allows estimating the stiffness decay factor and hence shelf–life of the tomatoes under the strict assumptions that any two measurements are independent of each other, and that data variability remains constant over storage time. Both assumptions are highly questionable and put the use of this type of model open to discussion. From literature, it appears that the homoscedasticity assumption of the data is often not valid [7, 10, 11, 21] who all encounter time dependent data variability. This was also found in the results presented here where figure 4 indicated that the squared OLS residuals show increased amplitude towards the end of the experiment. This increase can be due to two different causes. *First*, because tomatoes are harvested when they attain a certain maturity stage, their variability in stiffness, which is correlated to maturity [22], is lower than during storage where different tomatoes react in a different way to climate conditions increasing data variance. *Second*, as proven in literature [10, 11, 23], the repeatability of the acoustic firmness tester worsens with decreasing product stiffness. Both components are likely to have their influence but the OLS setting does not allow separating their contributions to the data variability.

Even more questionable is the assumption of independence of error terms since two measurements taken on the same tomato will be related. This (wrong) assumption of the OLS model has no important consequence on the estimated parameters since they are asymptotically consistent [19], but has a very important consequence on their standard errors, and hence on parameter significance. Due to the negation of the dependency among measurements of a given tomato, the information contained in the data set is overestimated leading to smaller but inconsistent standard errors. Model fits are often evaluated by looking at parameter variances, but this is only valid under justifiable model assumptions. This was shown in the practical example of comparing the stiffness evolution of two tomato varieties where the significance of their difference is highly dependent on the model assumptions that were made, making it very difficult to correctly interpret results in literature where the model assumptions are not checked thoroughly. For instance, using the repeated measures analysis, the conclusion would be that *TradiroM* and *Quest* tomatoes in August show a similar stiffness evolution (P = 0.1047) while the final mixed model that used the same data (model 3g) concludes the opposite (P = 0.0098).

The repeated measures analysis of the data relaxes the assumption of independent error terms and even of homoscedasticity since a heterogeneous covariance structure was used. Although the model fit improved drastically over the OLS model in terms of the -2ℓ value, the standard errors of the estimates increased remarkably due to reasons that were mentioned above. The heterogeneous autoregressive covariance structure tracks the data variance adequately, but probably is a too complex structure (it assumes a separate variance estimate for each storage time) to describe the variance trend when considering figure 5. As was the case in the OLS model, the source of the variance cannot be split into the different sub–contributions due to tomato variability and measurement error which, again, is a severe drawback of these models since interest of growers could lay in quantification of homogeneity of the batch or variety (inter–tomato variance), and researchers and companies manufacturing acoustic firmness devices put emphasis on the repeatability of their equipment (intra–tomato variance). These drawbacks are perhaps the most important stimuli to prefer the mixed model approach above higher–mentioned approaches.

The linear mixed model for longitudinal data, considering three repetitions for each tomato at each measurement day, was capable to divide the data variance into the wanted components that were named inter– and intra–tomato variances. This total variance decomposition demonstrates one of the important advantages of longitudinal model 4h over, for instance, the analysis at each time point and the OLS model. Using the OLS model, it was possible to quantify the total data variance by squaring the residuals and fitting a cubic spline through it. Although the so–obtained variance function showed an increase during storage, this increase could not be accounted for in the OLS model. That analysis only gives an indication of the average, constant variance being an oversimplification of the true underlying data behaviour. Using model 4h, it is possible to allow for heteroscedasticity and to quantify the different components of the data variance, which is of utmost importance. During the whole experiment, the inter–tomato variance proves to dominate the intra–tomato variance (figure 7) indicating that deviations of individual patterns from the variety mean are in the first place due to the different behaviour of the different tomatoes within that variety. Much smaller is the deviation that is due to the imperfect repeatability of the acoustic firmness technique. When comparing the intra–and inter–tomato variance with the residual error variance σ^2 it can be noted that σ^2 is larger than the intra–tomato variance until day 10, but that at the last day of the experiment the intra–tomato variance becomes larger (figure 7); the combined contribution of σ^2 and intra–tomato variance never accounts for more than one third of the total variance. As such, the unexplained variance of the model σ^2 remains very low in contrast with the OLS and repeated measures analysis where all variance could be regarded as ‘unexplained by the model’.

6. Conclusions

Four methods for analysing repeated measures data on horticultural products – inherently exhibiting a large inter–subject variability – were discussed and compared. Although many research still make use of classical techniques such as an analysis at each time point or an ordinary least squares regression model, other techniques are available in statistical software packages and that are much more flexible that higher–mentioned techniques. Perhaps the most flexible of those techniques is the concept of mixed models for repeated measures as it is able to describe several contributions of variance (such as intra– and inter–subject variance) and allow for complex variance–covariance structures. The findings that were postulated were applied to a practical example where the tomato firmness of different cultivars is followed during postharvest storage. This research shows the caveats of interpreting results when the basic assumptions of a given model are not fully fulfilled, leading to controversial results. The mixed model approach presented in the text proves to be the most flexible in order to be able to describe the different variance contributions (within tomatoes and between tomatoes), together with the correct treatment of the repeated measures design of the experiment. By including random effects in the model, it was shown that most of the random variation was due to the different behaviour of tomatoes within a variety. Much smaller was the variation that was due to the measurement error when taking repeated measures on one single tomato. This intra–tomato variation was shown to increase during storage, meaning that the repeatability of the firmness device was lower for soft tomatoes.

Footnotes

[1] The notation 1:2 refers to the mixture of chi–squared distributions.

[2] Only if a heterogeneous covariance structure is assumed.

[3] Only if random time trends are included, or a heterogeneous covariance structure is assumed.

[4] Analysis at each time point using a separate mean for each harvest – variety combination

[5] OLS model, $slope^2$ was not significant and hence not added to the model

[6] Repeated measures analysis, ARH(1) covariance structure

[7] Models 1 to 3 are based on the average stiffness of each tomato.

[8] Models 4 represent 3 repetitions for each tomato. –2Log L and AIC's hence are not comparable to models 1, 2 and 3.

[9] Model 4h has the same structure as model 4g except that it includes the local effects σ^2_o $diag[exp(U_i\delta)]$.

References

[1] G. Verbeke, G Molenberghs, Linear mixed models for longitudinal data, Springer–Verlag, New York, 2000.

[2] N. Galili, J. De Baerdemaeker, Performance of acoustic test methods for quality evolution of agricultural products, ISMA Conference, Leuven, Belgium (1996).

[3] R.L. Jackman, A.G. Marangoni, D.W. Stanley, Measurement of tomato firmness, Hortsc. 25(7) (1990) 781–783.

[4] C.N. Thai, R.L. Shewfelt, Peach quality changes at different constant storage temperatures: empirical models, Trans. ASAE 33(1) (1990) 227–233.

[5] H. Chen, J. De Baerdemaeker, Resonance frequency and firmness of tomatoes during ripening, Proceedings of the 22^{nd} international conference on agricultural mechanisation, Zaragoza, Spain 1 (1990) 61–68.

[6] P. Chen, Z. Sun, L. Huarng, Factors affecting acoustic responses of apples, Trans. ASAE 35(6) (1992) 1915–1920.

[7] H. Chen, Analysis on the acoustic impulse resonance of apples for non–destructive estimation of fruit quality, Ph.D. thesis, KULeuven, Belgium, 1993.

[8] D. Rosenfeld, I. Shmulevich, N. Galili, Measuring firmness through mechanical acoustic excitation for quality control of tomatoes, Food automation congress, Orlando, USA (1994).

[9] N. De Belie, S. Schotte, P. Coucke, J. De Baerdemaeker, Development of an automated monitoring device to quantify changes in firmness of apples during storage, Postharvest Biol. Tec. 18 (2000) 1–8.

[10] B. De Ketelaere, J. De Baerdemaeker, Advances in Spectral Analysis for Non–Destructive Tomato Stiffness Estimation Using Vibration Measurements, J. Agr. Eng. Res. 78(2) (2001 a) 177–185.

[11] B. De Ketelaere, J. De Baerdemaeker, Nonparametric Smoothing: Theory and Application to spectral analysis, Math. Comput. Simulat. 56(4–5) (2001 b) 385–394.

[12] N.M. Laird, J.H. Ware, Random effects models for longitudinal data, Biometrics 38 (1982) 963–974.

[13] G. Verbeke, G Molenberghs, Linear mixed models in practice: a SAS oriented approach. Lecture notes in statistics 126, Springer–Verlag, New York, 1997.

[14] R.C. Littell, G.A. Milliken, W.W. Stroup, R.D. Wolfinger, SAS system for mixed models, SAS Institute Inc., Cary, NC, 1996.

[15] J. Felföldi, A. Fekete, M. Gilinger Pankotai, Firmness–based assessment of tomato shelf–life, Proceedings of AgEng, Madrid, Spain (1996).

[16] S. Schotte, N. De Belie, J. De Baerdemaeker, Acoustic impulse response technique for evaluation and modelling of the firmness evolution of tomatoes, Postharvest Biol. Tec. 17(2) (1999) 105–115.

[17] H. Akaike, A new look at the statistical model identification, IEEE T. Automat Contr. 19(6) (1974) 716–723.

[18] P.J. Diggle, K.Y. Liang, S.L. Zeger, Analysis of longitudinal data, Oxford Science Publications, Oxford, 1994.

[19] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[20] P.J. Huber, The behavior of maximum likelihood estimates under nonstandard

conditions, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA 1(967) 221–223.

[21] J.A. Abbott, L.A. Liljedahl, Relationship of sonic resonant frequency to compression tests and Magness–Taylor firmness of apples during refrigerated storage, Trans. ASAE 37(4) (1994) 1211–1215.

[22] S. Schotte, J. De Baerdemaeker, Ontwikkeling en uittesten van mechanische en electrische impedantiemetingen in relatie met vruchthardheids– en kwaliteitsbeoordeling. Final project report (in Dutch), 1998.

[23] F. Duprat, M. Grotte, E. Pietri, D. Loonis, The Acoustic Impulse Response Method for Measuring the Overall Firmness of Fruit, J. Agr. Eng. Res. 66 (1997) 251–259.

Tables

Table 1: Overview of flexibilities of models used to model repeated quality measures.

Analysis type	Allow for repeated measures?	Allow to model product variability?	Allow for heteroscedasticity?
Analysis at each time point	no	no	yes
OLS regression model	no	no	no
Repeated measures analysis	yes	no	yes[2]
Linear mixed model for longitudinal data	yes	yes	yes[3]

Table 2: Tukey multiple comparisons at 0 and 14 days of storage for the August data. Cultivars with the same letter for a given day are not different at a 5 % significance level.

Day	Tukey grouping					Mean stiffness ($\times 10^6 Hz^2 g^{2/3}$)	cultivar
0	A					8.26	*RZ7457*
		B				7.12	*DRW5736*
		B				7.08	*DRW6391*
		B	C			6.97	*S&G18161*
		B	C	D		6.77	*S&G49107*
		B	C	D		6.62	*RZ72503*
		B	C	D	E	6.46	*TradiroM*
		B	C	D	E	6.40	*DRW6492*
		B	C	D	E	6.26	*BS9445*
			C	D	E	6.18	*DRW6340*
				D	E	6.06	*TradiroSKW*
				D	E	5.95	*Quest*
					E	5.95	*E2031152*
					E	5.66	*Mariachi*
14	A					6.80	*RZ7457*
	A	B				6.10	*DRW6391*
		B	C			5.38	*DRW5736*
		B	C	D		5.21	*S&G49107*
			C	D		5.09	*S&G18161*
			C	D		4.82	*RZ72503*

C	D	4.80	*Quest*
C	D	4.79	*E2031152*
C	D	4.66	*DRW6492*
C	D	4.58	*TradiroM*
C	D	4.57	*Mariachi*
	D	4.45	*TradiroSKW*
	D	4.34	*DRW6340*

Table 3: Tomato segmentation data. Overview of the model building process with parameter estimates and standard errors using REML; model fit statistics obtained under ML. Parameter estimates hold for DRW5736 tomatoes in August. Models 1–3 based on average stiffness; model 4 includes three repetitions for each tomato. Values for random effects denote variances.

Model Nr	Intc	s.e	Slope	s.e	Slope²	s.e	Random intercept	Random slope	Random slope²	–2Log L	AIC
0[4]	–	–	–	–	–	–	–	–	–	–4567.1	–4185.1
1[5]	1.9370	0.0213	–0.0225	0.002461	–	–	–	–	–	–4451.5	–4337.5
2a[6]	1.9532	0.0272	–0.0223	0.002816	–	–	–	–	–	–7615.5	–7487.5
2b	1.9686	0.0284	–0.0323	0.005910	0.00080	0.00041	–	–	–	–7704.6	–7520.6
3a[7]	1.9370	0.0211	–0.0225	0.002443	–	–	–	–	–	–4451.5	–4337.5
3b	1.9370	0.0275	–0.0225	0.001374	–	–	0.0131	–	–	–7275.0	–7159.0
3c	1.9370	0.0277	–0.0225	0.002123	–	–	0.0142	6.7×10^{-5}	–	–7643.4	–7523.4

Table 3 continued.

Model Nr	Intc	s.e	Slope	s.e.	Slope²	s.e	Random intercept	Random slope	Random slope²	–2Log L	AIC
3d	1.9704	0.0273	–0.0389	0.008917	0.001150	0.00060	–	–	–	–4483.2	–4313.2
3e	1.9704	0.0289	–0.0389	0.004579	0.001150	0.00029	0.0131	–	–	–7374.9	–7202.9
3f	1.9704	0.0281	–0.0389	0.004586	0.001150	0.00034	0.0143	6.9×10^{-5}	–	–7780.7	–7604.7
3g	1.9704	0.0289	–0.0389	0.004712	0.001150	0.00035	0.0137	1.2×10^{-4}	9.9×10^{-7}	–7884.4	–7702.4
4a[8]	1.9365	0.0136	–0.0227	0.001569	–	–	–	–	–	–10939.6	–10825.6
4b	1.9365	0.0267	–0.0227	0.001032	–	–	0.0134	–	–	–18616.1	–18500.1
4c	1.9365	0.0279	–0.0227	0.002191	–	–	0.0152	8.4×10^{-5}	–	–19708.2	–19588.2
4d	1.9696	0.0176	–0.0389	0.005732	0.001139	0.00039	–	–	–	–11013.6	–10843.6
4e	1.9696	0.0277	–0.0389	0.003752	0.001139	0.00025	0.0134	–	–	–18785.5	–18613.5
4f	1.9696	0.0287	–0.0389	0.003951	0.001139	0.00023	0.0152	8.4×10^{-5}	–	–19912.8	–19736.8

4g	1.9696	0.0283	–0.0389	0.004635	0.001139	0.00034	0.0148	2.3×10^{-4}	1.4×10^{-6}	–20153.8	–19971.8
4h[9]	1.9673	0.0290	–0.0367	0.004687	0.000945	0.00034	0.0141	1.3×10^{-4}	7.1×10^{-7}	–21437.2	–21249.2

Figure legends

Figure 1: Graphical representation of the mixed model building process used throughout this chapter (after [13]).

Figure 2: Variety–specific tomato stiffness profiles as function of storage time, modelled using an unstructured mean. August harvest. Δ: BS9445; ∇: DRW5736; +: DRW6340; –: DRW6391; o: DRW6492; ×: E2031152; ∗: Mariachi; □: Quest; ◊: RZ72503; ◁: RZ7457; ▷: S&G18161; ❶: S&G49107; ①: TradiroM (full line) and TradiroSKW (dashed line).

Figure 3: Variety–specific tomato stiffness profiles as function of storage time, modelled using an unstructured mean. October harvest. Δ: BS9445; ∇: DRW5736; +: DRW6340; –: DRW6391; o: DRW6492; ×: E2031152; ∗: Mariachi; □: Quest; ◊: RZ72503; ◁: RZ7457; ▷: S&G18161; ❶: S&G49107; ①: TradiroM (full line) and TradiroSKW (dashed line).

Figure 4: Spline–smoothed average trend of the squared OLS residuals of model 1 (–) and modelled variance function using the same model 1 (– –).

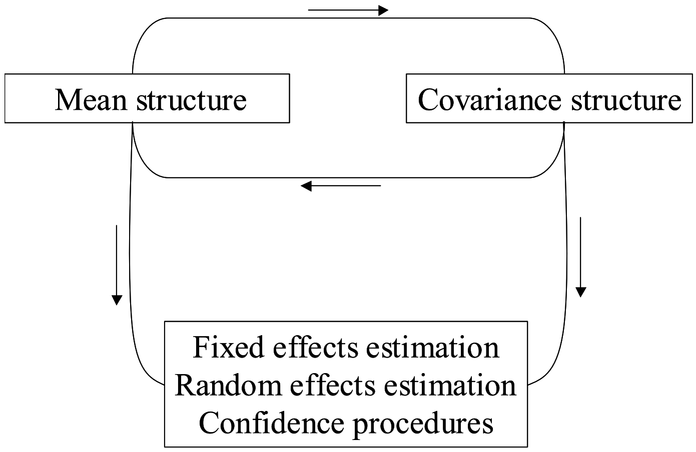
Figure 5: Spline–smoothed average trend of the squared OLS residuals of model 1 (–) and modelled variances σ^2_i *using model 2b with a heterogeneous autoregressive structure (Δ).*

Figure 6: Spline–smoothed average trend of the squared OLS residuals of model 1 (–) and modelled variance function using model3g (– –).

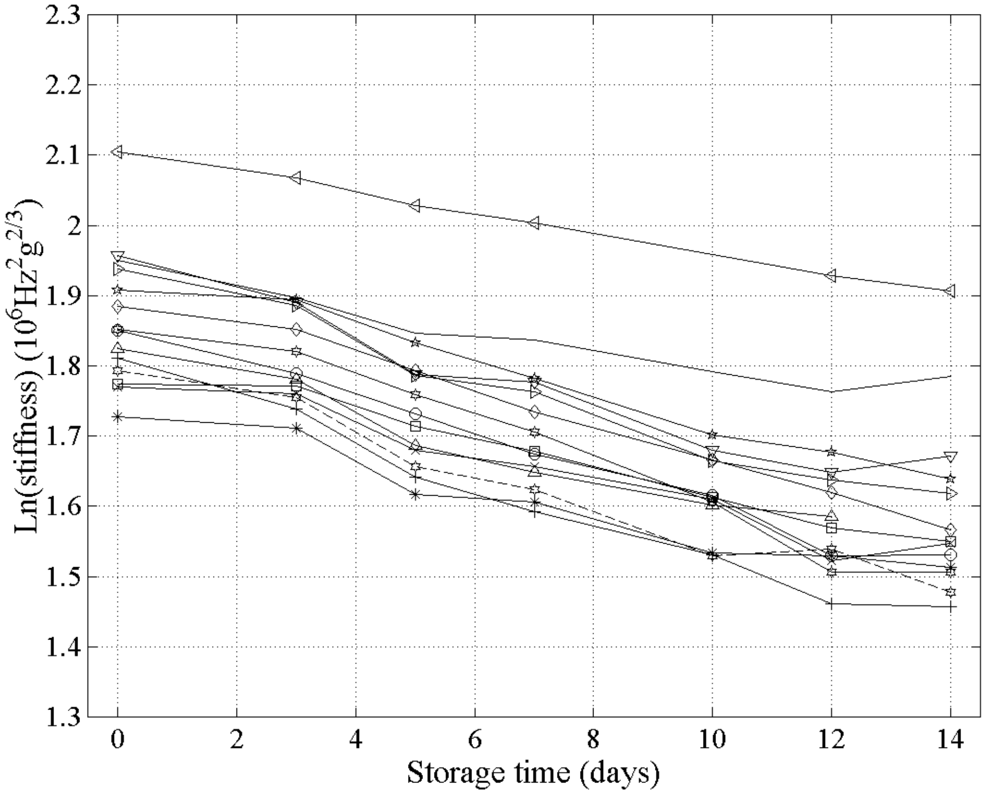
Figure 7: Decomposition of the total variance as function of storage time into its three components (model 4h). The intra–tomato variance is given by σ^2_o *diag[exp(*$U_i\delta$*)]; the*

inter–tomato variance by $Z_iDZ_i^T$ *and the residual variance by* σ^2.

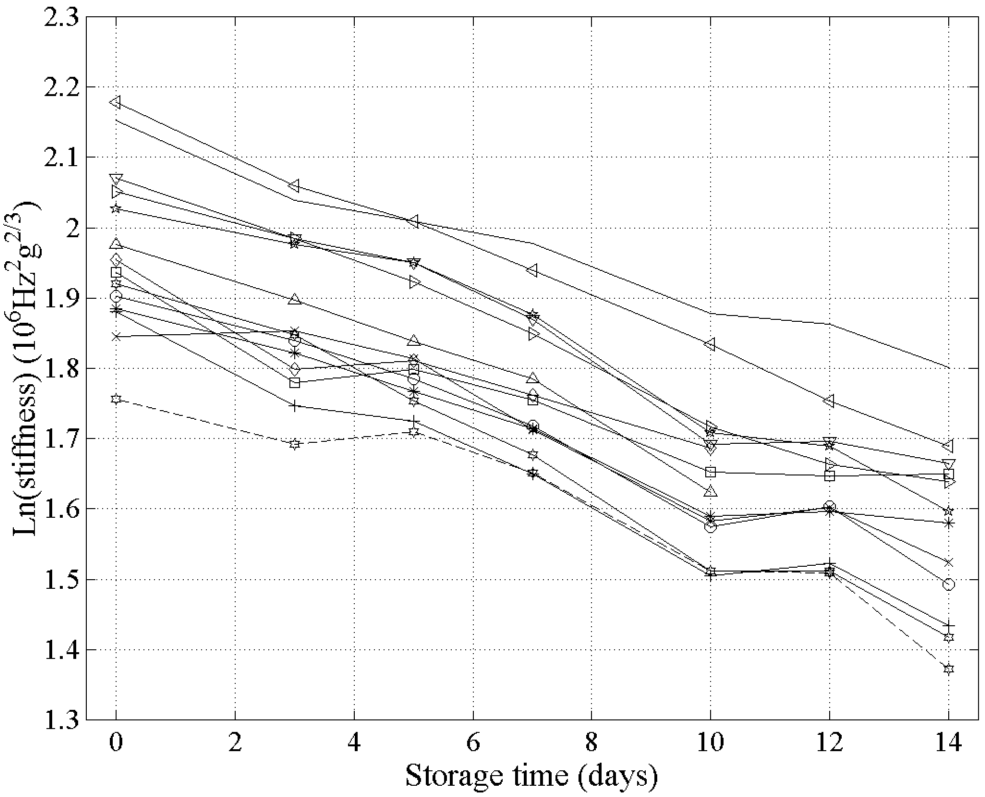
Figures

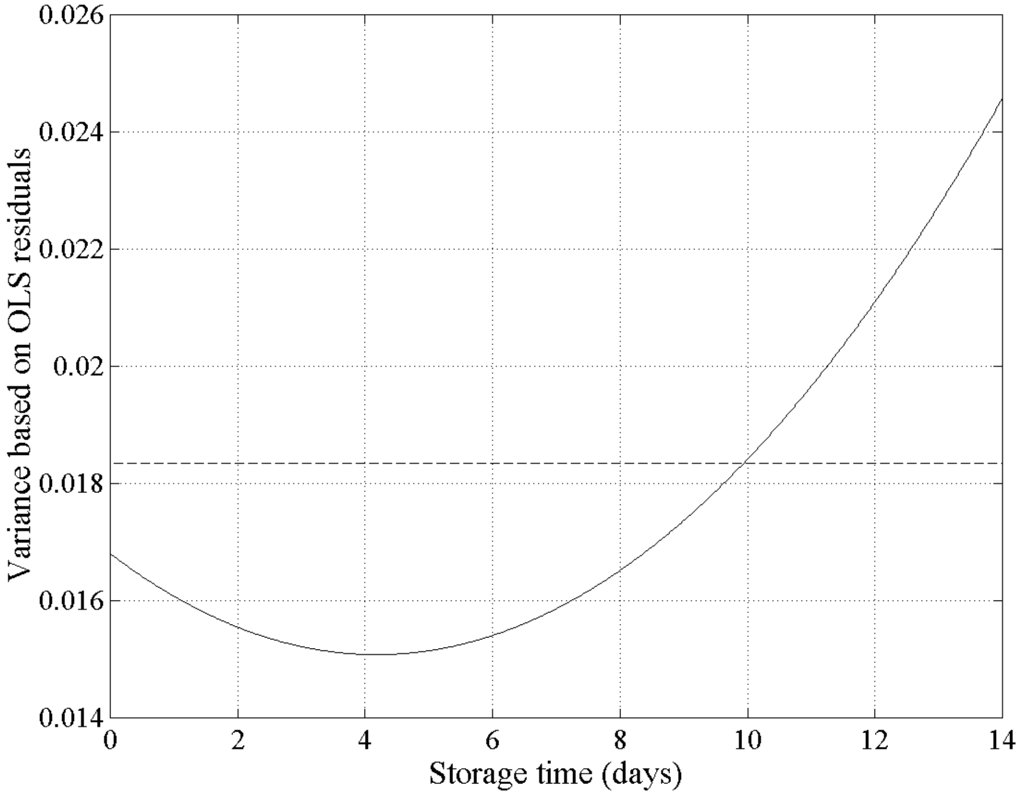

Mean structure
Covariance structure
Fixed effects estimation
Random effects estimation
Confidence procedures

641

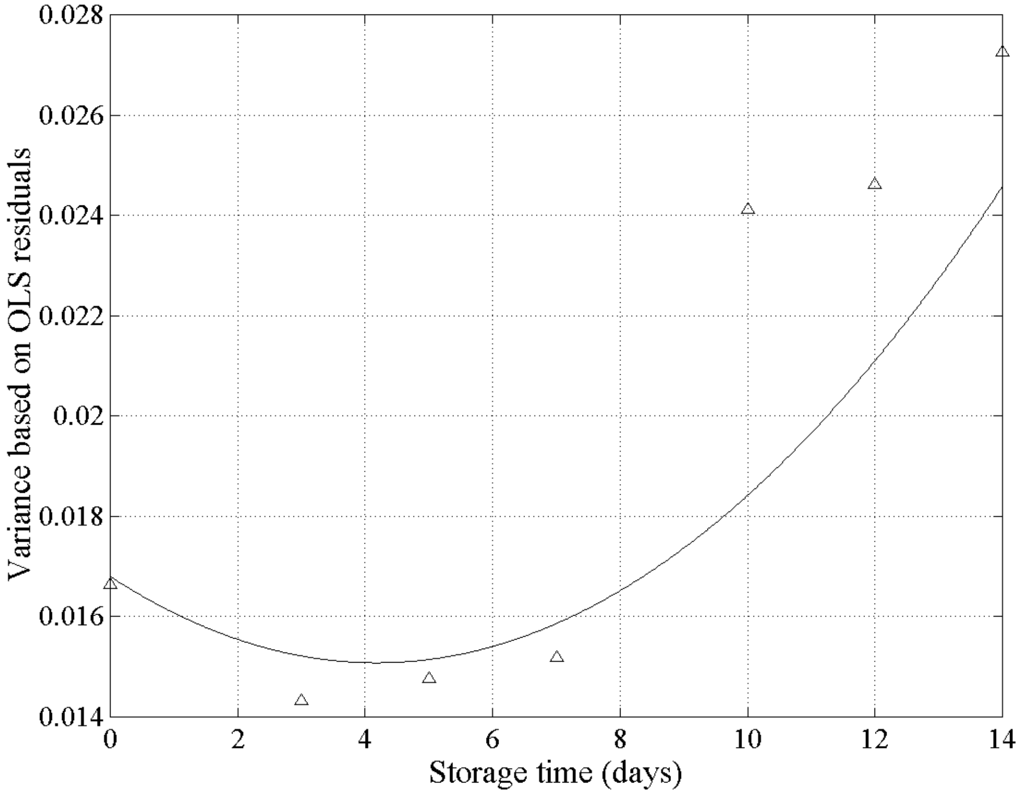


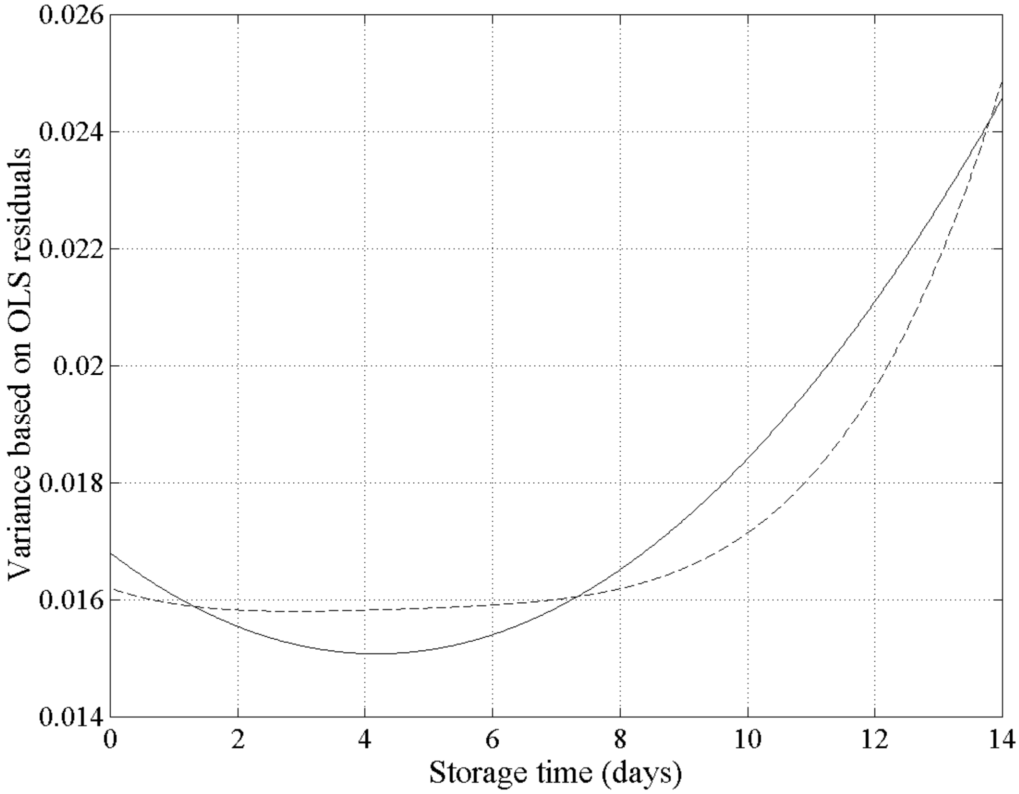
Ln(stiffness) (10^6Hz^2g^2/3)
Storage time (days)
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
0
2
4
6
8
10
12
14

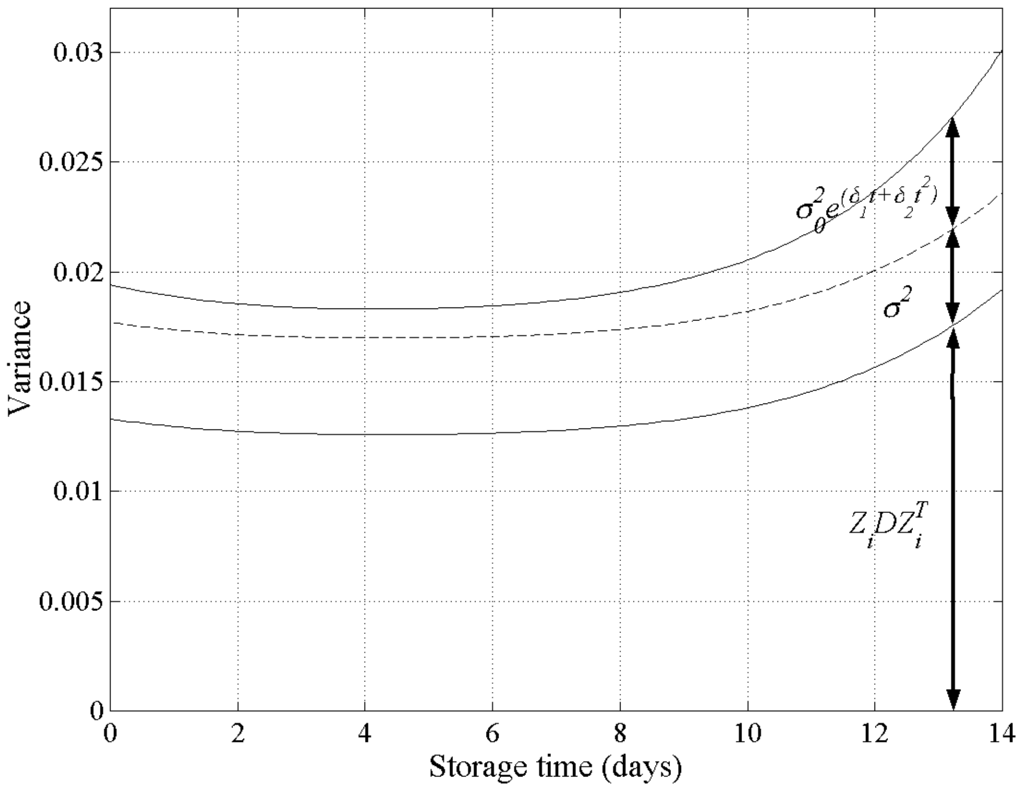

2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
0
2
4
6
8
10
12
14
Storage time (days)
Ln(stiffness) ($10^6 Hz^2 g^{2/3}$)



0.026
0.024
0.022
0.02
0.018
0.016
0.014
0
2
4
6
8
10
12
14
Variance based on OLS residuals
Storage time (days)


0.028
0.026
0.024
0.022
0.02
0.018
0.016
0.014
0
2
4
6
8
10
12
14
Variance based on OLS residuals
Storage time (days)


0.026
0.024
0.022
0.02
0.018
0.016
0.014
0
2
4
6
8
10
12
14
Variance based on OLS residuals
Storage time (days)



0.03
0.025
0.02
0.015
0.01
0.005
0
Variance
0
2
4
6
8
10
12
14
Storage time (days)
$\sigma_0^2 e^{(\delta_1 t+\delta_2 t^2)}$
σ^2
$Z_i D Z_i^T$