Multilevel modeling in food science: A case study on heat-induced ascorbic acid degradation kinetics

M.A.J.S. van Boekel, Stéphanie Roux

Multilevel modeling in food science: a case study on heat-induced ascorbic acid degradation kinetics

M.A.J.S. van Boekel[1] & S. Roux[2]

[1] Food Quality & Design Group, Wageningen University & Research, Wageningen, the Netherlands

[2] Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 91300, Massy, France

Food Research International 158 (2022) 11565
https://doi.org/10.1016/j.foodres.2022.111565

Author note

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. Declarations of competing interest: none.

Correspondence concerning this article should be addressed to M.A.J.S. van Boekel, Food Quality & Design Group, Wageningen University & Research, P.O. Box 17, 6700 AA Wageningen, the Netherlands. E-mail: tiny.vanboekel@wur.nl

Abstract

Characterization of variation of experimental results is achieved by repeating experiments. Frequently, results are averaged before data are analysed but that may not be the best practice because valuable information is then lost. Three other ways to analyze repetitions are: 1) each experiment is analyzed on its own (no pooling of data), 2) all experiments are analyzed together in one go (complete pooling), 3) data are analyzed together while allowing for similarities as well as differences in the result (partial pooling). Multilevel modeling uses partial pooling by partitioning variance over more than one level. Level 1 consists of the measurements themselves, higher levels consist of groups or clusters of measurements (repetitions, experiments at various temperatures, at various pH values, etc.) and parameters are analyzed both at the population and at the group/cluster level.

The approach is applied to a case study in which heat-induced isothermal degradation of ascorbic acid was studied with 15 repetitions in an aqueous solution, making it a two-level study. The data were analyzed using averaging and complete pooling, complete pooling without averaging, no-pooling at all, and partial pooling. The kinetic model was established by letting the data decide about the order of the reaction, while this was compared to a model where the order was fixed at 1 (first-order model). Results show that both averaging with complete pooling, as well as complete pooling without averaging, strongly underestimate variation. The no-pooling technique overestimates variation, while partial pooling partitions variation over the levels and thus gives a better impression of the variation involved. The kinetics of ascorbic acid appear to be subject to strong variation when each experiment is considered separately because it is a compound that is very sensitive to all kinds of experimental conditions. With multilevel modeling it appeared to be possible to characterize the uncertainties involved much better than with single level modeling. A Bayesian analysis was performed, in which parameters are allowed to be variable, which is useful because multilevel modeling leads to characterization of variation of parameters. The Bayesian method allows to visualize the posterior distribution of parameters, thereby giving more insight in their behaviour. Also, a Bayesian analysis focuses more strongly on predictive accuracy of models, including multilevel models. The predictive accuracy of 4 models describing the same ascorbic acid data was compared and the multilevel model with reaction order estimated from the data performed by far the best in this regard. The pros and cons of multilevel modeling are discussed and it is concluded that multilevel modeling is to be preferred whenever the data allow to perform such an analysis.

Keywords: Bayesian regression, kinetics, reaction order, heating, multilevel modeling, ascorbic acid, prediction, data pooling

1 Introduction

Variability is an important aspect to deal with in science in general, and in food science in particular because of all kinds of variability to which foods are subject. There is, in fact, variability at various levels. Foods are usually not homogeneous, their composition depends on growing conditions, genetic aspects play a role, the way sampling is done, the analysis method used can have an effect, and on top of that there is variation for which the reason is unknown. Obviously, it is important to be able to quantify variation as much as possible but it is also important to determine the level at which such variation occurs: is it at the level of the actual measurements or at the level of treatments. This is where the topic of multilevel modeling comes in. It is perhaps useful to consider some concepts a little closer. Accuracy refers to the ability to approach the "true" value of a quantity of interest as close as possible, but the problem is that it is usually unknown what the true value is (otherwise a measurement would not be needed). Accuracy is something that can only be achieved by proper calibration of methods and equipment: systematic bias should be absent. There is no statistical method that can correct for inaccuracy and it is the responsibility of the researcher that measurements are as accurate as possible. This is different for the concept of precision: this refers to how close measurements are together (note that a measurement can be very precise but inaccurate, while accurate measurements can be very imprecise). Precision can be estimated using statistics by doing repeated measurements. It is known both from experience and statistical theory that chemical and physical measurements lead to normally distributed results; this follows from the central limit theorem stating that many small errors lead to normally distributed errors. The beauty of the normal distribution is that it only needs two parameters, the mean and the standard devation (or equivalently the variance) to fully characterize the distribution. In the words of McElreath (2020) it is the most simple distribution that yields maximum entropy (in terms of information theory). Other measurements such as counts or yes/no scores require other distributions such as the Poisson and binomial distribution. Here the discussion is limited to normally distributed measurements. There are also variability concepts such as repeatability (variation occurring under the same conditions, same lab, same equipment, same operator) and reproducibility (variation occurring for the same experimental design in different labs, different operators, etc.). Generally,

variation will be lower (better precision) for repeatability studies than for reproducibility studies but the concept of precision remains the same: how are measurements spread between each other as characterized by variance/standard deviation.

In publications, food scientists may report that replicate measurements were done, which should mean that experiments were repeated in an independent way following exactly the same procedure. However, this is not always clearly stated and the danger of 'pseudo-replication' is lurking, meaning that measurements are thought to be independent when they are actually not (Lazic, Clarke-Williams, & Munafò, 2018; Lazic, Mellor, Ashby, & Munafo, 2020).

It is rather common to average obtained results to be able to deal with variation. This is, unfortunately, not a good habit as will be discussed later on, mainly because useful information is discarded upon averaging. The goal of this article is to show that there are better ways to deal with variation. Variation seems to be considered by many as a nuisance, but it can also be seen as a source of useful information that teaches a lot about the system under investigation. That will be the approach used here and the method to do that is called multilevel modeling. It is a topic that is gaining importance in many branches in science but not so much yet in food science, which is the motivation for the current article. Since the concept is not well known yet, a brief introduction on the topic follows. The concept will then be illustrated with a case study on heat-induced degradation of ascorbic acid. The data set consists of a substantial number of repetitions of experiments with exactly the same experimental design (published partly already by Gómez Ruiz, Roux, Courtois, and Bonazzi (2018) but supplemented here with additional measurements); moreover, the published data were not yet analyzed in a multilevel way. Since the experiments were done by four different researchers over a period of five years (2013-2017), using the same equipment and with freshly made solutions for each run, this could be classified as a study on reproducibility.

A very brief introduction to the concept of the multilevel approach is given, first, by explaining the differences between no pooling, complete pooling and partial pooling of data. Then, a very brief recapitulation about the Bayesian approach and its connection to multilevel modeling follows, before the results are discussed.

1.1 A brief account of multilevel modeling

Experimental designs are obviously important for what can be done in subsequent analysis and modeling. In that context, levels are to be understood as follows. Repetitions of similar experiments ('runs') can be grouped; the measurements within each run form then the lowest level 1, while the

runs themselves can be grouped into the next level 2; they can be considered as a subsample of the population of all possible runs. Simply said, whenever experiments can be subdivided in groups, multilevel modeling is possible. For example, experiments done at various temperatures, or at various pH values, or at various water activities can be grouped. Some model parameters are then allowed to vary per group around a central value. Response variables at level 1 are always the measurements or observations. Response variables at higher levels are regression coefficients from the level below that. Examples in food microbiology can be found in Garre, Zwietering, and Den Besten (2020) and Van Boekel (2021b), chemical measurements in Hickman, Ignatowich, Caracotsios, Sheehan, and D'Ottaviano (2018), Van Boekel (2022) and Van Boekel (2021a). For instance, suppose that there are n measurements $y_i(i = 1..n)$ in a particular run (level 1), that there are r runs$_j(j = 1..r)$ performed at the same condition (level 2) while runs done at the same temperature, for instance, could be grouped at level 3 indexed by the number of m different temperatures $T_k(k = 1..m)$ (a graphical illustration is shown in Figure S1 in the Supplement as an example of a so-called nested experimental design). In the current article, the analysis is limited to two levels: level 1 consists of measurements within each run, while the cluster of all runs together form level 2, all performed isothermally at one temperature $T = 70$ °C.

What can be done in further analysis of the obtained data depends on how experiments are designed. A *first* approach could be to pool the results from all repetitions and analyse them as if they are all generated without any variation between the repetitions, thus ignoring group structures; the remaining variation (i.e., not explained by the model) then piles up in residual variance. This is properly called **complete pooling**. Such an approach may lead to underfitting (Gelman & Hill, 2007), meaning that not all information in the data is used. A *second* approach is to average over all repetitions at each experimental setting, which results in **pooling and averaging**. The data are then compressed which may lead again to underfitting. A *third* approach is to analyse each repetition on its own, the **no pooling** approach. This leads to as many modeling outcomes (parameter estimates) as there were repetitions. The consequence is that the outcome of one repetition is in no way connected to the outcome of another one. When no pooling is applied, group means are estimated independently as if the variation between groups is infinitely large; it tends to make the groups more different than they actually are and tends to lead to overfitting (Gelman & Hill, 2007), i.e., putting too much trust in the data. The *fourth* approach connects the repetitions with each other, characterized as **partial pooling**. Partial pooling is achieved with multilevel modeling; group means are considered a random sample from an overarching distribution called

the population. Partial pooling is obviously in between no-pooling and complete pooling; one could consider it a compromise between under- and overfitting. One could also consider complete pooling and no pooling approaches as subsets of partial pooling. If the standard deviation between repetitions is characterized as σ_g, then $\sigma_g \to \infty$ for complete pooling, whereas $\sigma_g \to 0$ for no pooling. With partial pooling, $0 < \sigma_g < \infty$, while the actual value of σ_g can be estimated in the multilevel modeling approach. Single level classical regression is in that perspective a special case of multilevel regression.

Another important aspect to consider when analyzing repetitions is that measurements within each repetition may not be completely independent from a statistical point of view. Measurements may be correlated if samples are taken from within one run. Classical regression of such data is then not allowed because of this correlation; it leads to underestimation of variation and biased parameter estimates and their uncertainties. A major advantage of multilevel modeling is that it does take such correlation into account, leading to unbiased parameter estimates. It counteracts the danger of "pseudoreplication" (Lazic et al., 2018, 2020).

The terminology used in multilevel modeling can be confusing. Parameters describing effects at the population level are sometimes named 'fixed' while those describing variation at the cluster or group level are called 'random.' The term 'mixed effect models' refers then to a mixture of random and fixed effects, a term sometimes used instead of multilevel modeling. The strength of the multilevel method is that the levels inform each other; variation between experiments is accounted for but it is also taken into account that there are similarities between experiments. Information from one experiment is used in the analysis of another. Since one could also see a hierarchy in an experimental design, another term used is hierarchical modeling.

Readers interested in more background of multilevel modeling are referred to Garre et al. (2020), Gelman and Hill (2007), Gelman, Carlin, Stern, Vehtari, and Rubin (2013), Kruschke (2015), Lambert (2018), McElreath (2020), Van Boekel (2021a), Van Boekel (2021b), Van Boekel (2022) and numerous tutorials on the internet. This very short introduction serves as background for the approach followed in this paper. As reported elsewhere, the Bayesian approach is well suited for multilevel modeling (Van Boekel, 2021a, 2021b, 2022), as opposed to the more traditional and common frequentist approach.

1.2 A brief account of the Bayesian approach

Because the principles of the Bayesian approach are explained in the references mentioned above, only a very brief account will be given here. The Bayesian approach is the alternative for the frequentist approach that is, until now, used in most food science publications. The major difference is that in the Bayesian approach the attention is directed towards the plausibility of hypotheses and models given the available data, whereas in the frequentist approach the attention is on the plausibility of data given an assumed model or hypothesis. As a consequence, parameters are considered variable in the Bayesian concept and are allowed to have a probability distribution, whereas in the frequentist approach parameters are considered unknown but fixed and therefore cannot have a probability distribution. Frequentist confidence intervals, therefore, are not statements about probability of parameters, they refer to how often a parameter will be found in or out an interval upon (imaginary) frequent repetitions. The interpretation of such a 95% confidence interval is that upon 100 estimations a parameter will be in a specified interval in 95 cases and will not be in that interval in 5 cases. Also, frequentist estimation (usually via least-squares regression, which is a special case of maximum likelihood estimation) leads to point estimates of parameters (an estimate of the unknown but fixed value), Bayesian estimation leads to complete probability distributions of parameters (so-called posterior distributions). To distinguish from the frequentist concept of confidence intervals, the Bayesian equivalent is called a credible interval: it indicates the **probability** that the parameter is in a specified interval, not **how often** a parameter will be found in an interval as in the frequentist framework. In Bayesian estimation, researchers are forced to make their assumptions about parameters and models explicit by having to define so called prior distributions for parameters (reflecting expert knowledge on what is already known about parameters before data analysis) and a likelihood function for the data (reflecting the assumption about how the data are generated). In the frequentist world, this should actually also be done but that is rarely the case; it is usually tacitly assumed that assumptions for least-squares regression are fulfilled. Briefly, the assumptions are: correct model specification, additive errors, no errors in the independent predictor variable, normally distributed errors, mean of errors is zero, and the magnitude of errors are the same. These assumptions should be checked but this is, unfortunately, not always done, and consequently parameter estimates may be biased.

Since multilevel modeling considers random variation of parameters, it comes naturally to connect multilevel modeling to the Bayesian way of working, though it should be remarked that multilevel modeling is also done in the frequentist way, even though the assumption of fixed parameters is

then strictly speaking no longer valid, see Bates et al. (2015), Pinheiro et al. (2017), Matheson (2020). In the current publication, only the Bayesian approach is applied. For all but very simple problems, posterior distributions cannot be obtained analytically and Markov Chain Monte Carlo simulations are needed. This is by now a well established method (Betancourt, 2017) and various software packages are available.

1.3 Ascorbic acid kinetics

Ascorbic acid is, obviously, an important food constituent as a vitamin but also as an anti- and pro-oxidant. There is a vast amount of literature on kinetics of ascorbic acid during processing, notably heating. The literature was extensively reviewed by Gómez Ruiz et al. (2018) and more recently by Giannakourou and Taoukis (2021), so that will not be repeated here. Two additional references are from Al Fata, Georgé, André, and Renard (2017) and Shen et al. (2021). According to Al Fata et al. (2017), there are two degradation pathways, an aerobic one that goes via dehydroascorbic acid (a reversible step) leading to formation of 2,3-diketogulonic acid (an irreversible step). The other pathway is an anaerobic one in which the lactone ring is hydrolytically cleaved. They summarized global ascorbic acid (AA) degradation as depicted in equation (1):

$$\frac{\mathrm{d[AA]}}{\mathrm{d}t} = -k_{ox} \cdot [\mathrm{AA}]^{\alpha}[\mathrm{O_2}]^{\beta} - k_h \cdot [\mathrm{AA}]^{\gamma} \qquad (1)$$

In this equation, k_{ox} is the rate constant for aerobic and k_h for anaerobic (hydrolytic) degradation and α, β, γ the partial reaction orders. It shows clearly the kinetic complexity of the reaction. It is therefore not a surprise that there is no consensus in literature about kinetic characterization, as there are quite some conflicting results. The complex reaction mechanism indicates that ascorbic acid is quite sensitive to oxidation and degradation and so, when experimental conditions differ only slightly, different results will be obtained. In addition, its stability depends on many factors such as pH, metal traces (acting as catalysts), oxygen content, temperature, presence of other antioxidants. In other words, there is uncontrolled, and perhaps even uncontrollable experimental variation. A recent and comprehensive overview of the many possible kinetic pathways of degradation of ascorbic and dehydroascorbic acid can be found in Shen et al. (2021).

The fact that degradation of ascorbic acid leads to substantial experimental variation, and therefore also to variation in the subsequent kinetic analyses, makes it a suitable case to apply a multimodeling approach with the goal to characterize and better understand that variation. The data used in the current study come from a model system using an aqueous solution, studied under very

strict reaction conditions that may not be directly applicable to foods. However, since the reaction conditions were strictly controlled, it makes the data very suitable for the goal of the present study: to show how variability can be captured and quantified.

2 Materials and Methods

2.1 Experimental data

Heat-induced degradation of ascorbic acid solutions was studied in a reactor as described by Gómez Ruiz et al. (2018). There were 15 repetitions done by four researchers over a period of five years (2013-2017) using the same experimental settings (initial concentration around 5.6 mM of ascorbic acid (1 g/L), a gas mixture of 21% oxygen/79% nitrogen was continuously bubbled through the solution, heating temperature 70 °C, in a malate buffer with pH 3.8). Researcher 1 did runs 1-9, researcher 2 runs 10-11, researcher 3 runs 12-13, researcher 4 runs 14-15. Solutions with initial concentrations varying around 5.6 mM were prepared fresh for each experiment (small variations in initial concentration existed because the amount of ascorbic acid powder added was not precisely the same). Experiments were done in a batch reactor, and samples were removed at specific time points from the reactor over the course of one run. Fifteen repetitions (runs) were done (9 were reported before, 6 additional experiments are included here). The full experimental details can be found in Gómez Ruiz et al. (2018). These 15 experiments/repetitions are analyzed here. Compared to the original paper (Gómez Ruiz et al., 2018), data were recalculated as follows: time in seconds was recalculated to hours and concentration in M to mM. This was done to bring the numerical values in the same order of magnitude to avoid numerical difficulties in the MCMC (Markov Chain Monte Carlo) procedure to approximate the posterior distributions (see below). This has no effect on the modelling, only on the units of the resulting parameter estimates.

2.2 Software

The calculations, plots and writing were done in RStudio (version 1.4.1103) using the R package papaja. The R package brms (bayesian regression models using stan, version 2.16.1) was used for regression (Bürkner, 2017, 2018). `brms` uses the probabilistic programming language `Stan` in the background for the MCMC calculations, see the website from the Stan development team

https://mc-stan.org/users/documentation/, see also Carpenter et al. (2017). Graphs were produced using the R package ggplot.

MCMC sampling requires thorough checks to be sure that the algorithm converged. In all cases described here, four chains were run simultaneously with at least 4000 simulations of which half was discarded as warm-up. The standard checks for convergence are: trace plots showing graphically whether or not the chains have converged, $\hat{R}$ (Rhat) that should be very close to 1 (showing whether the chains have mixed well), the number of effective simulations showing how many of the simulations were effective (there is no clear-cut number for this parameter but if it is at least half of the total number of simulations that should be fine). These checks were performed for every analysis done and results are only reported if all the checks were OK; examples of such checks can be found in previous papers (Van Boekel, 2020, 2021a, 2021b, 2022). The R code used as well as the data sets can be found at the GitHub page of the first author: https://github.com/TinyvanBoekel/FRI.

2.3 Modelling approaches

2.3.1 Single level modeling with a normalized first-order reaction.

A first-order reaction using normalized concentrations (c_t/c_0) is displayed in equation (2) with parameter k_r as the rate constant:

$$\frac{c_t}{c_0} = \exp(-k_r \cdot t) \qquad (2)$$

Regression in the Bayesian way requires a likelihood function for the data (reflecting the assumption of how the data are statistically distributed) and prior distributions for the parameters (Van Boekel, 2020). For the likelihood, if a researcher proposes a first-order reaction as shown in equation (2), another assumption must be made how the data are generated in the statistical sense. When using least-squares regression in the frequentist framework, it is tacitly assumed that data are normally distributed. In the Bayesian framework, this must be explicitly stated and this is done by stating that the data are assumed to be generated from a normal distribution (symbol $\mathcal{N}$) according to a first-order reaction (more precisely: that the residuals are normally distributed). Since the data are normalized, there are only two parameters left to estimate, the rate constant k_r and the experimental standard deviation σ_e. For the prior of the rate constant, a normal distribution is assumed with a mean of 0.5 h^{-1} and a lower bound at zero (lb=0). The reason to choose a normal distribution is that it, according to McElreath (2020), "reflects the most natural expression of the

state of ignorance". Since it is physically impossible that a rate constant is negative, a lower bound of zero is added. Other possibilities are to opt for a log-normal distribution or an exponential distribution, which also allow only positive values. But since natural processes follow frequently a normal distribution this is chosen here. As for numerical values, a very rough rule of thumb in kinetic analysis is that $k_r \cdot t_{end} \approx 1 - 10$, with t_{end} the length of time for which the experiments were run. In this case, $t_{end} = 6.5$ h, hence a value of 0.5 is proposed, but a relatively large standard deviation of 0.3 h^{-1} is given to allow the software to search for quite different values: it expresses the uncertainty on the part of the researcher about the value of this parameter. To give such a large standard deviation is the mathematical way of stating this uncertainty. In most cases, the data (expressed in the likelihood function) will overrule the choices made for the prior. See Gelman et al. (2017) for details on the choice for priors; van Boekel (2020) also paid some attention to prior choice for kinetic models. For the experimental standard deviation, a half-cauchy distribution was used, as is frequently done for standard deviations because it has a large "fat" tail so that unlikely high values are still possible should the data suggest that (Van Boekel, 2020). Note that these prior distributions are not reflecting the 'true' parameter distributions but the uncertainty of the researchers. These assumptions will be combined with the information in the data and will result in posterior distributions. Displayed in statistical language, these assumptions lead to equation (3):

$$\begin{aligned} \frac{c_t}{c_0} &\sim \mathcal{N}(\mu_i, \sigma_e) \\ \mu_i &= \exp(-k_r \cdot t_i) \\ k_r &\sim \mathcal{N}(0.5,0.3) \quad \text{(lb=0)} \\ \sigma_e &\sim \text{half-cauchy}(10) \end{aligned} \qquad (3)$$

The symbol '$\sim$' should be read as: 'is distributed as,' this applies to the likelihood and the priors. The model itself is a deterministic equation as indicated by the symbol '=,' indicating that μ_i does not need to be estimated but is calculated from the parameters estimated.

2.3.2 Single-level modeling with the n^{th}-order model.

The equation to estimate the order of a reaction with respect to time is displayed in equation (4) (Van Boekel, 2021a):

$$c_{\mathrm{t}} = \left(c_0^{1-n_t} + (n_t - 1)k_r t\right)^{\frac{1}{1-n_t}} \qquad (4)$$

Next to the rate constant k_r and experimental standard deviation σ_e, there are two more parameters to estimate in this equation: the initial concentration c_0, and the order of the reaction n_t. The

proposed likelihood function to capture the data generating process is thus given by equation (4), which combined with the proposed priors for the parameters results in equation (5):

$$\begin{aligned} c_t &\sim \mathcal{N}(\mu_i, \sigma_e) \\ \mu_i &= \left(c_0^{(1-n_t)} + (n_t - 1) \cdot k_r \cdot t_i\right)^{\frac{1}{1-n_t}} \\ c_0 &\sim \mathcal{N}(6,1) \\ k_r &\sim \mathcal{N}(0.5,0.3) \quad (\text{lb}=0) \\ n_t &\sim \mathcal{N}(1,0.5) \\ \sigma_e &\sim \text{half-cauchy}(10) \end{aligned} \qquad (5)$$

The numerical values for the prior distributions are again based on expert knowledge, namely that the initial concentration should be in the neighborhood of 5.6 mM, and that the reaction order will be somewhere between 1 and 2. The standard deviations given are quite wide so that the priors allow for different values should the data suggest that; these can be considered as weakly regularizing priors, meaning that they prevent the software from searching for impossible values while they do allow to search for extreme values if the data suggest that. Unless expert knowledge suggests it, priors should in general not be dominating, meaning that the posterior distribution should not be determined mainly by the choice of prior but by the data. It is preferred to use weakly informative or regularizing priors (McElreath, 2020), where the function of the prior is to prevent impossible values (such as negative rate constants) but to allow for unlikely but extreme values should the data suggest that. Prior predictive checks are a way to find out the effect of priors on estimation as well as to test the implications of a proposed model (McElreath, 2020; Van Boekel, 2021a). These checks were done (results not shown) to be sure that the proposed priors were indeed weakly regularizing. Basically, these checks consist of generating plots that show possible outcomes of the dependent variable (in this case ascorbic acid concentrations) as a function of time. If such plots would show physically impossible values, prior choice should be adapted; it should show, however, unlikely values that are extreme but not physically impossible.

2.3.3 Multilevel modeling with the n^{th} order model.

The likelihood function for the data is again given by equation (4). Prior distributions for the parameters requires some extra work for multilevel modeling. Random effects are introduced to accommodate the between-experiment variation. This is done by acknowledging that the parameters are related because they are estimating the same effect. That is to say, the three parameters c_0, n_t and k_r are allowed to vary from experiment to experiment, by characterizing that variation with the introduction of three new parameters u, v, w. Variation in c_0 is then $c_0 \pm u$,

variation in n_t is $n_t \pm v$ and variation in k_r is $k_r \pm w$. The trick is that they are also connected by assuming that these three random effects are statistically distributed following a multivariate normal distribution MVN as displayed in equation (6):

$$u, v, w \sim \text{MVN}(0, \Sigma)$$
$$\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv}\rho_{uv} & \sigma_{uw}\rho_{uw} \\ \sigma_{vu}\rho_{vu} & \sigma_v^2 & \sigma_{vw}\rho_{vw} \\ \sigma_{wu}\rho_{wu} & \sigma_{wv}\rho_{wv} & \sigma_w^2 \end{pmatrix} \quad (6)$$

The variance-covariance matrix Σ holds the variances σ_u^2, σ_v^2, σ_w^2 and the covariances $\sigma_u\sigma_v\rho_{uv}$, $\sigma_u\sigma_w\rho_{uw}$, $\sigma_v\sigma_w\rho_{vw}$. It is a symmetric matrix, i.e., $\sigma_u\sigma_v\rho_{uv} = \sigma_v\sigma_u\rho_{vu}$, etc. The symbol ρ represents the correlation coefficient. For this particular case study with three random effects, it is thus a 3x3 random effects matrix. **Σ** can be rewritten as in equation (7):

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u & 0 & 0 \\ 0 & \sigma_v & 0 \\ 0 & 0 & \sigma_w \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_u & 0 & 0 \\ 0 & \sigma_v & 0 \\ 0 & 0 & \sigma_w \end{pmatrix} \quad (7)$$

R holds the symmetric correlation matrix with the parameter correlation coefficient ρ:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{uv} & \rho_{uw} \\ \rho_{vu} & 1 & \rho_{vw} \\ \rho_{wu} & \rho_{wv} & 1 \end{pmatrix} \quad (8)$$

Note from equation (6) that it is assumed that **on average** the random effects are zero. In other words, these random effects are considered as deviations from the "grand mean" obtained for all data. Noteworthy is also that not u, v, w will be estimated but rather $\sigma_u, \sigma_v, \sigma_w$, though u, v, w can be calculated afterwards. Since these are extra parameters, additional priors are needed for them. A normal distribution is again proposed for the likelihood function, as well as for the initial concentration c_0 and the rate constant k_r with a lower bound at zero, a half cauchy distribution is used for the random effects and experimental standard deviation, and the so-called LKJ prior is used for the correlation coefficients between the random parameters (LKJ stands for Lewandowski-Kurowicka-Joe, the people who proposed this prior, see McElreath (2020) for more information on this prior). All this translates to the statistical notation in equation (9):

$$
\begin{aligned}
c_t &\sim \mathcal{N}(\mu_i, \sigma_e) \\
\mu_i &= \left(c_0^{(1-n_t)} + (n_t - 1) \cdot k_r \cdot t_i\right)^{\frac{1}{1-n_t}} \\
c_0 &\sim \mathcal{N}(6,1) \\
k_r &\sim \mathcal{N}(0.5,0.3) \quad \text{(lb=0)} \\
n_t &\sim \mathcal{N}(1,0.5) \\
\sigma_u &\sim \text{half-cauchy}(10) \\
\sigma_v &\sim \text{half-cauchy}(10) \\
\sigma_w &\sim \text{half-cauchy}(10) \\
\sigma_e &\sim \text{half-cauchy}(10) \\
\rho &\sim \text{LKJ}(2)
\end{aligned}
\tag{9}
$$

3. Results and Discussion

Figure 1 displays all data points in one graph (Figure S2 in the Supplement shows the data per experiment). Two things become obvious from the plot in Figure 1 and Figure S2. First, that there is considerable variability between each experiment, which is typical for food-related experiments (note, though, that these experiments were done in homogeneous aqueous solutions so that there is no biological variation in this case). Second, that the degradation follows more or less the same pattern, so apart from the variability there is also similarity. These data will now be analyzed in various ways.

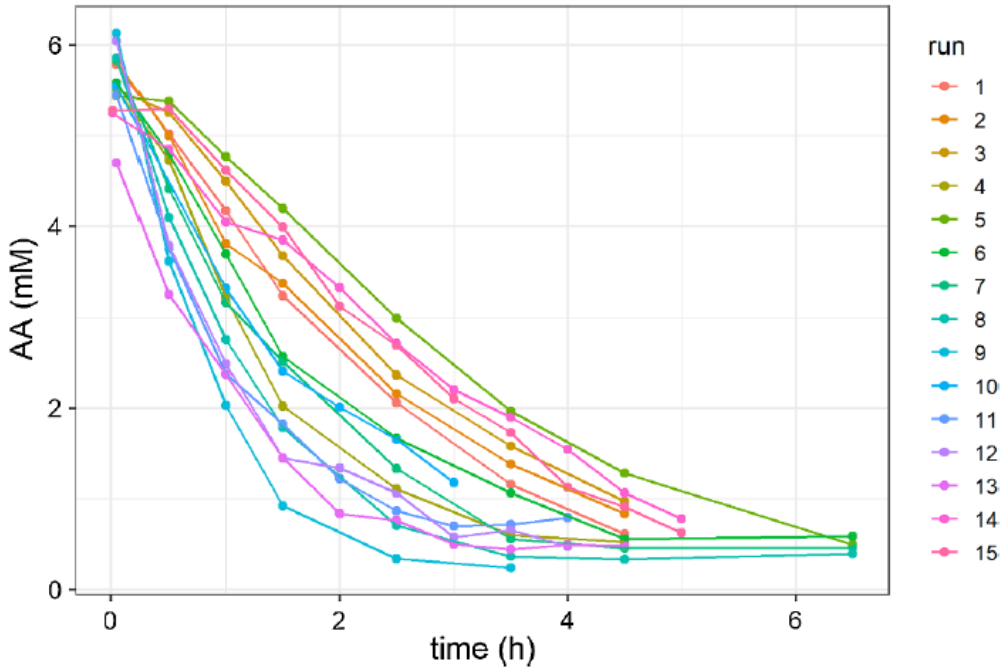

Figure 1. Overview of the data showing degradation of ascorbic acid (AA) in aqueous solution at 70 °C as a function of time: every run (15 in total) is given a different colour. Source: Gómez Ruiz et al. (2018)

3.1 Analysis of the kinetics of the normalized, averaged data using the first-order model

Gómez Ruiz et al. (2018) handled the observed variation in the data in two ways: i) by analyzing each dataset separately and studying the variation in obtained estimated rate constants (to be found in the Supplement of their article), and ii) by eliminating the variation *in initial concentration* by dividing the momentary concentration by the initial concentration for each experiment, c_t/c_0. The initial concentration used for normalization was estimated from the individual regressions. Next, they averaged the results obtained per sampled heating time from each run and performed nonlinear least-squares regression (i.e., in the frequentist way) of these normalized, averaged concentrations as a function of heating time according to a first-order model (equation (2)). This will be repeated here but then in the Bayesian way, for comparison with subsequent additional analyses. The results are shown as a numerical summary in Table 1 and graphically displayed in Figure 2.

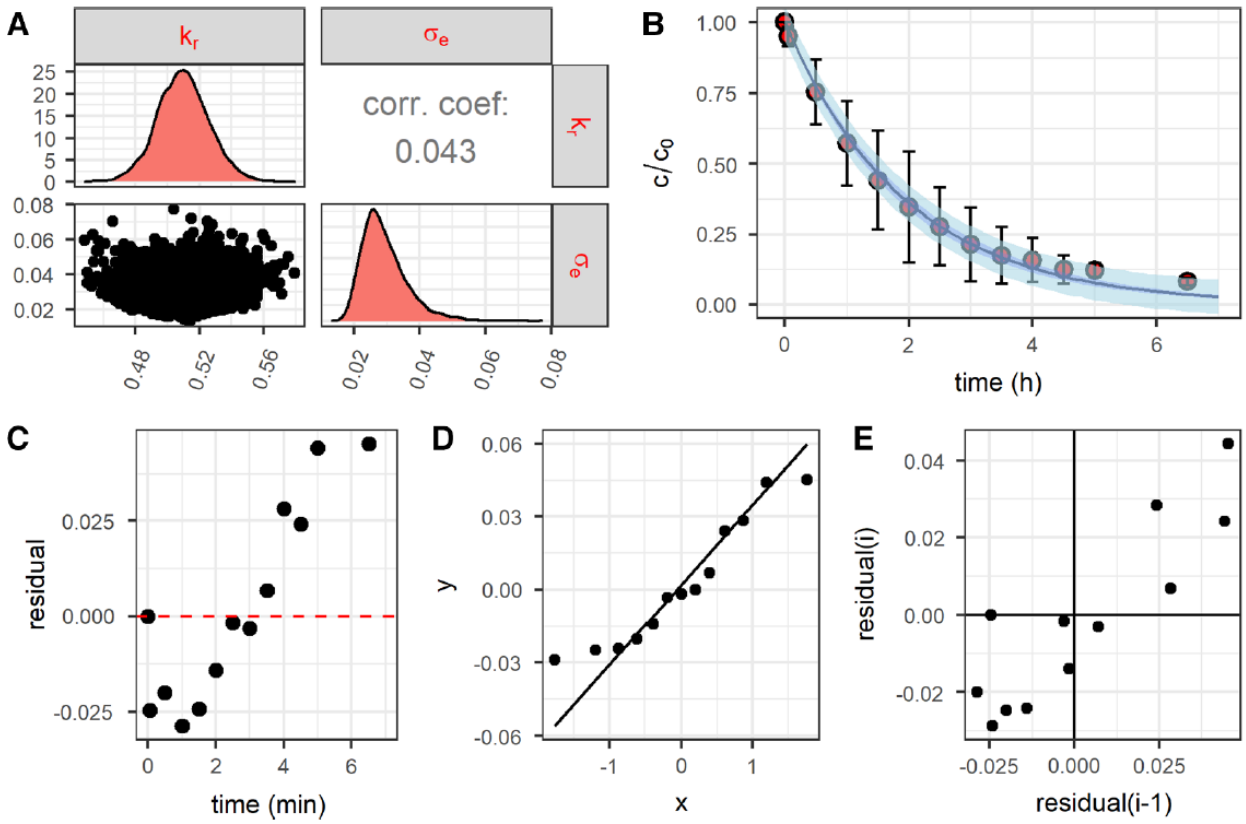

Figure 2. Bayesian nonlinear regression results of the normalized, averaged and pooled ascorbic acid data heated at 70 °C. A: pairs and posterior parameter density plots and correlation coefficient; B: regression line with 95% credible interval for the mean (dark blue ribbon, the credible interval is very narrow and almost coincides with the regression line) and 95% prediction interval for new values (light blue ribbon), the error bars represent the experimental standard deviations calculated for each of the 15 runs; C: residuals; D: normal Quantile-Quantile (QQ) plot; E: lag plot.

Table 1: Numerical summary of the posterior distribution resulting from Bayesian nonlinear regression for the normalized and averaged ascorbic acid data using the first-order model. SE= standard error, the lower and upper bound represent the 95% credible interval

Parameter	**mean**	**SE**	**lower bound**	**upper bound**
$\boldsymbol{k_r}$ $(\mathrm{h}^{\mathbf{-1}})$	0.510	0.016	0.478	0.543
$\boldsymbol{\sigma_e}$ (mM)	0.029	0.007	0.019	0.046

For comparison, Gómez Ruiz et al. (2018) reported a rate constant $k_r = 0.526 \pm 0.018$ h^{-1} obtained by least-squares regression, which is reasonably close to what is found here (the results reported in the paper were based on 9 experiments, meanwhile 6 more experiments were available included in the current analysis, which may cause slightly different results). The pairs and posterior parameter density plots and the correlation coefficient are in Figure 2A, while the resulting fit is in Figure 2B. McElreath (2020) coined the term 'retrodiction' as an alternative for the term 'fitting,' indicating how the model compares to the data in retrospect, emphasizing that it is not a prediction because it is based upon the same data that were used for establishing the model. The 95% credible and prediction intervals for these pooled, normalized and averaged data are also indicated in Figure 2B (the credible interval is very narrow and almost coincides with the regression line). A credible interval is the Bayesian equivalent of a confidence interval and indicates where the mean is expected to be found with 95% probability (note that this is different from the frequentist interpretation of a 95% confidence interval: this indicates where the mean will be expected in 95% of the cases if the experiment will be repeated). The 95% prediction interval indicates where future (not yet observed) values can be expected with 95% probability. This prediction interval does not cover the experimentally observed error bars, obviously because the model is not informed about these error bars.

The overall fit appears to be reasonable, as are the credible and prediction intervals. However, the large experimental standard deviations as shown in the error bars are not taken into account in this way: regression was only done on the average values. It illustrates that averaging removes information. The 95% credible interval is very tight around the regression line because it is only based upon the averages and not on the variation. Also the parameter estimate σ_e is only based on the averages and the deviations from the assumed first-order model, not on the experimental standard deviations. A remedy could be to do weighted regression where the standard deviations form the weights. This was not done in the original article and a different route will be taken here to account for the variation.

Regression results require model criticism, i.e., by investigating whether the model can stand some statistical tests. This can be checked by studying the residuals, normal Quantile-Quantile (QQ) plots to check the normality of the residuals and lag plots to check for serial correlation (Van Boekel, 2021a); these are displayed in Figure 2C, D, E, respectively. The residuals are clearly not randomly distributed; while the QQ plot is not too bad with some deviation from normality in the lower tail, but the lag plot suggests serial correlation. Overall, there are signs that the model is not performing well. The individual data sets were also analyzed per run using Bayesian nonlinear regression assuming first-order kinetics (analysis per run was also done by Gómez Ruiz et al. (2018) but then using nonlinear least-squares regression); the results of nonlinear Bayesian regression using the first-order model are shown in the Supplement. Reassuringly, the Bayesian regression results were numerically comparable to the least-squares regression results (shown in the Supplement of Gómez Ruiz et al. (2018)). Figure S3A in the Supplement shows a pairs and posterior parameter density plot for run 1 by way of example (other runs showed similar results), these plots do not give reason for concern. The individual fits obtained for first-order regression do not look bad at first sight (Figure S3B) but the residual plots in Figure S3C definitely show trends, similarly as was found for the normalized first-order model displayed in Figure 2C. All these results show that the residuals are not randomly distributed when using a first-order model. This is an indication that the first-order model is, perhaps, not the most optimal one. So, all in all, the results suggest that it could be interesting to take a closer look at the order of the reaction, while there also may be a problem due to normalizing and averaging. The next section investigates what happens if the data are not normalized and averaged, while also allowing the data to suggest the order of reaction, rather than fixing it at a value of 1.

3.2. Analysis of the kinetics of completely pooled data using the nth-order model

Relaxing the assumption of first-order kinetics prompts to take a closer look at the kinetic model. As equation (1) shows, the kinetics of ascorbic acid degradation is complex. For the experimental results analysed here, the oxygen concentration was kept constant at 21%. It may therefore be assumed that the anaerobic rate constant k_h is virtually zero while the constant oxygen concentration combines with the 'true' rate constant k_{ox} into a pseudo rate constant k_r. What remains to be established then is the value of the order parameter α in equation (1), which will be labeled here as n_t: see equation (4). This equation was also used by Al Fata et al. (2017) but they used a linearized form of it and a graphical analysis to estimate the order, so not via regression;

they reported an order of $n_t = 0.83$ but it should be noted that their experimental conditions were different. Here, nonlinear regression in the Bayesian way will be used to estimate all parameters, thereby avoiding having to interpret a graphical plot. As a reminder, the assumption is that the data are generated from a normal distribution with a mean μ_i and a (constant) standard deviation σ_e according to the n^{th}-order model: see equation (5). Prior information was, again, that the initial concentration is expected to be near 6 mM, while the numerical value for the rate constant was again expected to be around 0.5 with a large prior standard deviation as before. Expert knowledge suggests that it is reasonable to expect that the order n_t is not too far from 1, but also here a relatively high standard deviation is given. These are, again, weakly regularizing priors. The resulting pairs and posterior parameter density plots are in Figure 3A. These plots look alright but the rather strong correlation between parameter c_0 and k_r is noteworthy (ρ = - 0.941); this is partly due to the mathematical model formulation in equation (4). It did not have a major effect on parameter estimation because all the diagnostic measures were found to be OK.

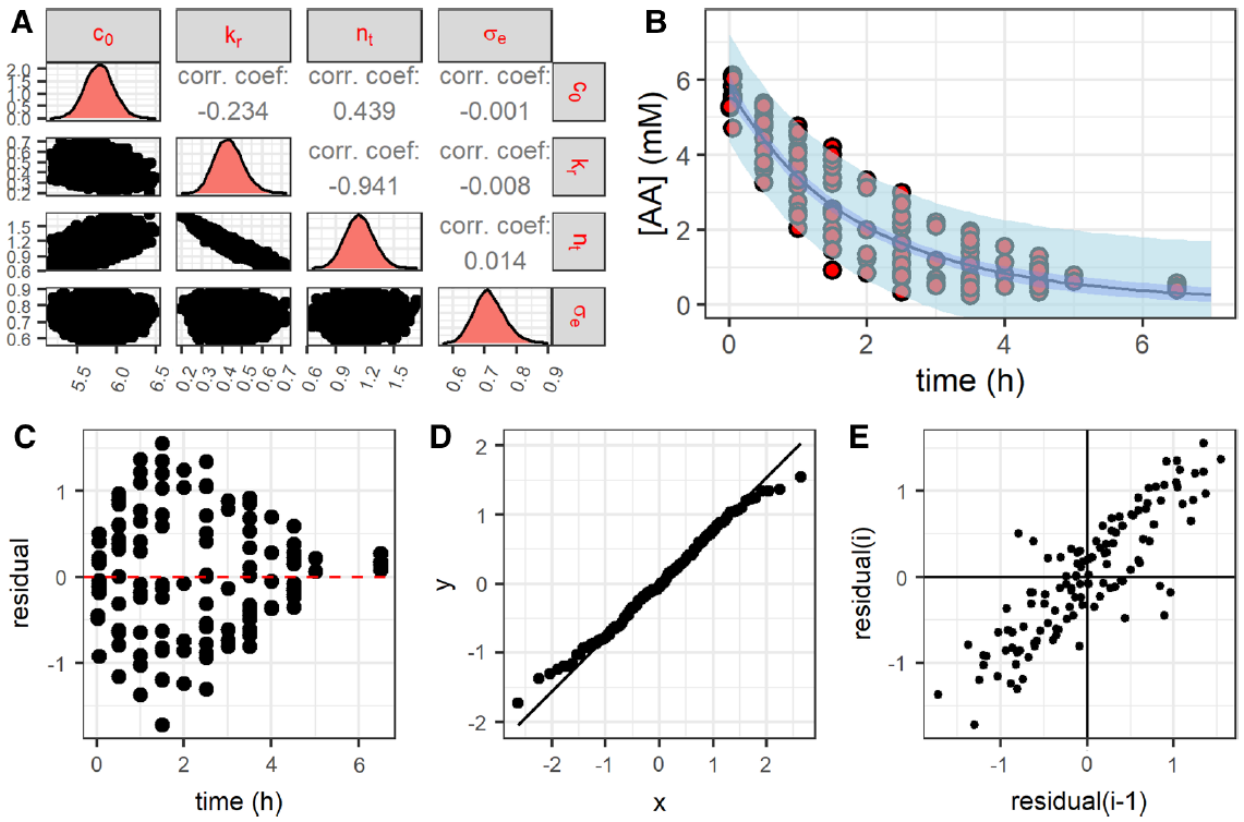

Figure 3. Bayesian regression results for the pooled (not averaged) ascorbic acid data heated at 70 °C using the n[th]-order model. Pairs and posterior parameter density plots and correlation coefficients (A), regression line, dark blue ribbon: 95% credible interval, light blue ribbon: 95% prediction interval (B), residuals (C), QQ plot (D) and lagplot (E)

Table 2 shows the numerical summary of the posterior. Because the order n_t appears to be somewhat higher than 1, the estimate for k_r is different from the one in Table 1 because of the strong correlation between these two parameters. It appears that, by allowing the parameters c_0 and n_t to vary, rather than fixing them, the posterior distribution of the parameter n_t does contain the value of 1 in its 95% credible interval, but the maximum density occurs at a somewhat higher value near 1.1 (Figure 3A). By having the initial concentration and order as parameters, the model has more freedom to choose the most likely estimates than when the order is fixed. The resulting fit, showing in how far the model is able to retrodict the data, is in Figure 3B. Just as with the normalized data, the resulting model is well able to describe the fate of the data overall and the 95% credible interval is again quite narrow (shown as a dark blue ribbon around the regression line), suggesting that the model is well able to predict the mean. The 95% prediction interval, on the other hand, is much wider than with the averaged data because of the scatter that is now 'seen' by the model, even predicting (impossible) negative values. In other words, the uncertainty in prediction is mainly due to the experimental variation characterized by the standard deviation σ_e rather than resulting from estimating the mean.

Table 2: Numerical summary of the posterior distribution resulting from nonlinear Bayesian regression of the n^{th}-order model for the pooled ascorbic acid data. SE= standard error, the lower and upper bound represent the 95% credible interval

Parameter	**mean**	**SE**	**lower bound**	**upper bound**
$\boldsymbol{c_0}$ (mM)	5.78	0.18	5.44	6.14
$\boldsymbol{k_r}$ $(\mathrm{h}^{\mathbf{-1}})$	0.43	0.07	0.31	0.58
$\boldsymbol{n_t}(-)$	1.14	0.15	0.86	1.44
$\boldsymbol{\sigma_e}$ (mM)	0.71	0.05	0.63	0.82

Next, in the model criticism phase, the residuals, QQ and lag plot are displayed in Figure 3C, D and E, respectively. The residuals look OK, indicating a better model fit than with the first-order model. The QQ plot looks also OK but the lag plot still shows rather strong signs of serial correlation. This is, in hindsight, to be expected because of the way the experiments were done: taking samples from the same solution in the same reactor in the same experimental run introduces dependencies in the data obtained for that run. Hence, one of the conditions for regression, independence between collected data, is violated, possibly leading to statistical bias.

So, by pooling (analyzing all the data together), the experimental scatter is taken into account but does not acknowledge that there may be dependencies in the data, resulting in too much trust in them. With complete pooling, the model is not informed that the experiments come from different runs. Variation is only accounted for in the residual standard deviation/variance, capturing the remaining variation not explained by the model. This is shown in Table 2 by σ_e, which is estimated as a parameter with its own uncertainty. The parameter estimates are sometimes mentioned as being at the "fixed effect level," fixed because they are supposed to describe the behaviour at the population level. The question is, though, whether or not fixed effect parameters obtained by complete pooling of data are a good indication for what is going on. That is what will be investigated next in this case study. The opposite of complete pooling is to use no pooling and to apply regression on each dataset separately. That is described in the next section.

3.3 Analysis of the kinetics of each run using the nth-order model (no pooling)

On the way to partial pooling, it may be instructive to compare complete pooling (as done in the previous section) with no pooling at all, i.e., studying the individual experiments separately. The parameters c_0, k_r and n_t are then estimated for every experiment separately, resulting in 15 estimates of the order of the reaction, the initial concentration and the rate constant, using again the n^{th}-order model (no-pooling was in fact already done with the individual runs using the first-order model as described in the Supplement (Figure S3) but the extra step taken here is to use the n^{th}-order model rather than fixing the order at 1). The same priors were used as displayed in equation (5). The pairs and parameter posterior density plots and parameter correlation coefficients for each run did not give reason for concern (Figure S4 in the Supplement shows run 1 by way of example, similar plots were obtained for the other runs). However, as found with complete pooling, parameters k_r and n_t are quite strongly negatively correlated, so a lower value of n_t is compensated by a higher value of k_r by the software to find the most likely fit; the strong correlation did not affect parameter estimation though. The individual fits with 95% prediction intervals are shown in Figure 4A and the residuals in Figure 4B. The fits look OK with well-behaved residuals. The resulting 15 posterior parameter distributions can be shown in several ways, one way is via so-called forest or ridgeline plots: see Figure 4C, D , E for the parameters c_0, n_t and k_r, respectively. Ridgeline plots compare posterior parameter densities for each category (runs in this case) and show in a glance how they differ.

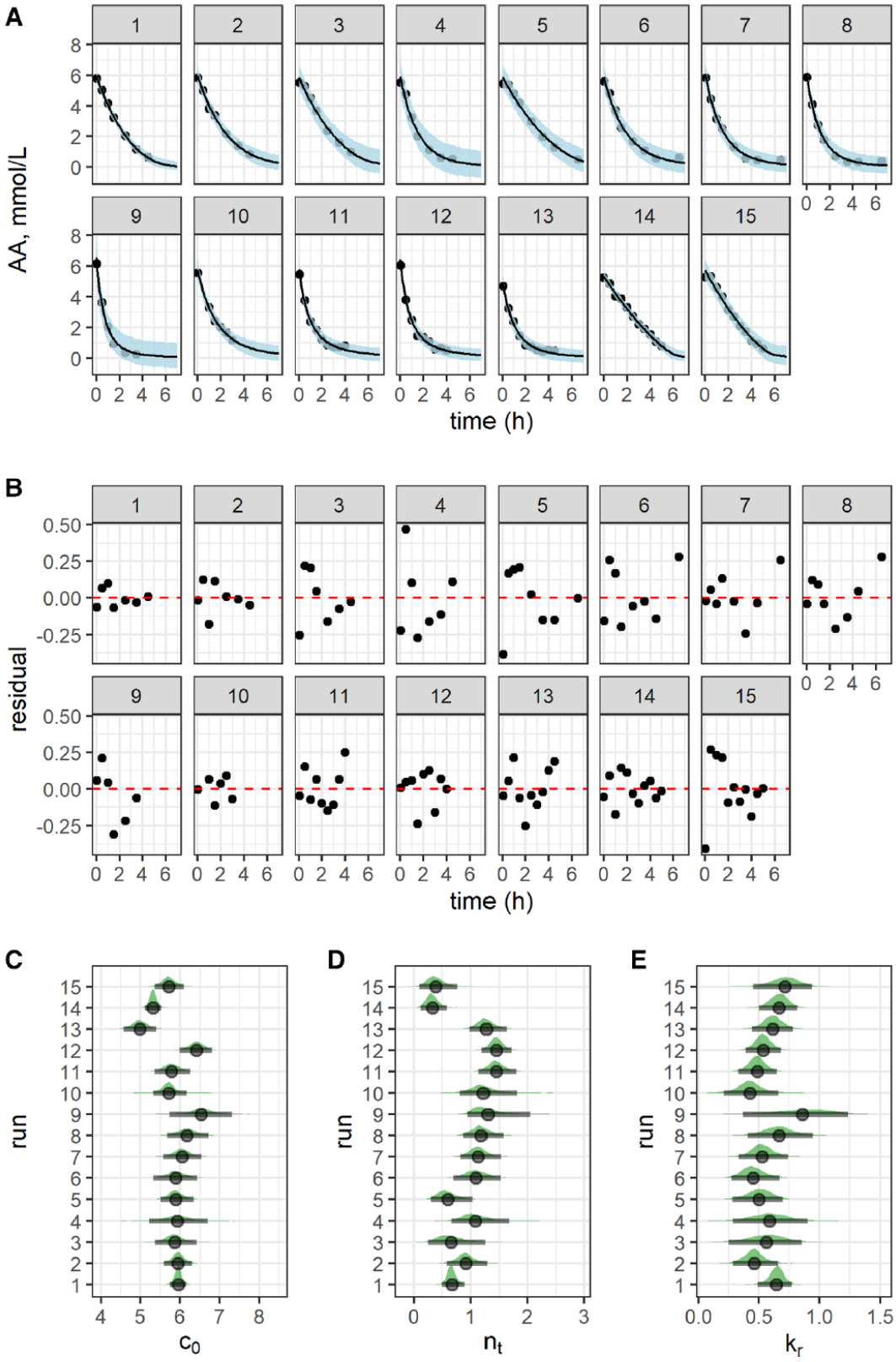

Figure 4. Bayesian regression of each run using the n^{th}-order model for the degradation of ascorbic acid (AA) at 70 °C (no pooling). A: Fits for each run (blue ribbon: 95% prediction interval); B: residuals for each run; C, D, E: ridgeline plots for parameters c_0, n_t, k_r for each run, green shapes: posterior parameter densities, dots and black lines: mean + 95% highest posterior density interval.

The variation in c_0 per run as displayed by the marginal posterior distribution is mainly due to variation in the way experiments were done: each repetition had a slightly different initial concentration, this is "real" variation. The estimation process leads to uncertainty in the value of the parameter and this uncertainty is also expressed as variation, and both types of variation cannot be separated. The result for n_t clearly shows the considerable variation in order per run, some very different from 1, but it also shows why on average an order around 1 was found for the pooled results. Furthermore, the posterior densities for parameter n_t show wide tails and skewed distributions. As for the parameter k_r, the distributions overlap but show also a wide variation.
However, in performing individual regressions like so, information that might be common between individual regressions is not shared: each regression starts without 'knowing' anything from the previous or next regression. With multilevel modeling on the other hand it is possible to incorporate such information. Moreover, it acknowledges the serial correlation within experiments where samples cannot be independent in the statistical sense (these are so-called longitudinal data in statistical jargon). A next logical step is therefore multilevel modeling.

3.3 Multilevel modeling: partial pooling using the nth-order model

Multilevel modeling means that there are, in this case, two levels (in other cases there might be more, see for instance Garre et al. (2020), Van Boekel (2021a), Van Boekel (2021b), Van Boekel (2022). Level 1 is at the level of individual measurements within each experiment (the samples taken at certain time points during each run), level 2 is formed by grouping of the runs (15 in total in this case). Each run can be seen as a subsample of the population of all possible runs. The parameters derived from each run can also be seen as a representation of the population of all possible parameter values. It is thus attempted to characterize the random variation within each run as well as between runs. That is why the term random effect (in contrast to a fixed effect) is sometimes used, while the term mixed effect refers to the fact that there is a mixture of random and fixed effects. The terminology is somewhat confusing but the approach is not. Its strength is that the levels inform each other; room is given for individual variation but it is also acknowledged that there are similarities. Multilevel modeling can be done both in the Bayesian and frequentist framework but the Bayesian approach is more natural where parameters are considered to be random anyway, whereas they are considered fixed in the frequentist way, so then there is a bit of a tension to consider them random. The frequentist method uses a specific maximum likelihood

method called restricted maximum likelihood (REML), interested readers are referred to Bates et al. (2015) and Pinheiro et al. (2017).

The pair plots and posterior densities are displayed in Figure 5A. The output shows, in addition to what was shown before, the random effects displayed as standard deviations and correlation coefficients with a posterior density. The plots look OK without serious parameter correlation. The numerical summary of the multilevel analysis is in Table 3.

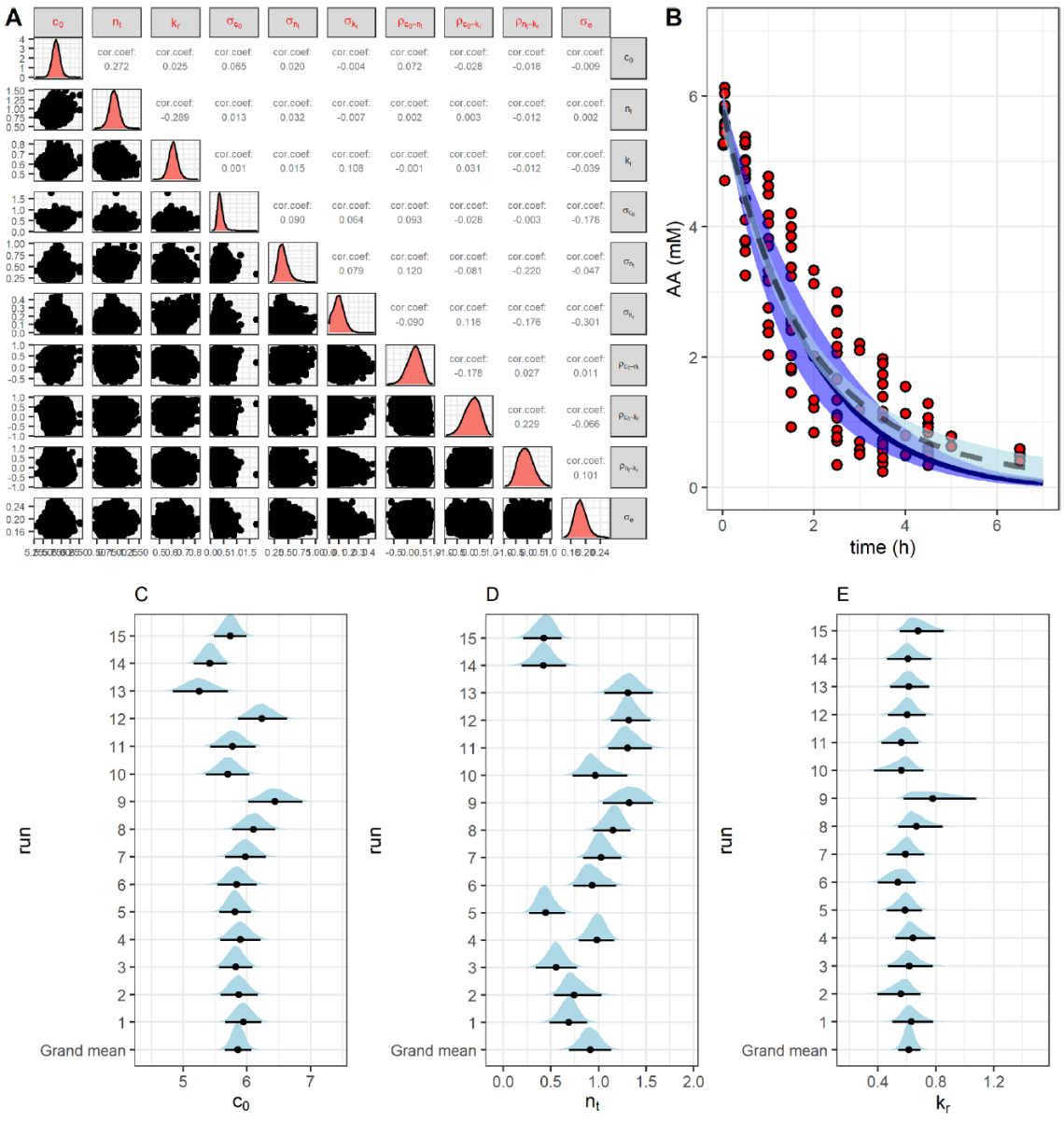

Figure 5. Multilevel modeling (partial pooling, n^{th}-order model) of the ascorbic acid data. A: Pairs and density plots + correlation coefficients; B: regression line (solid line) + 95% credible interval (dark blue ribbon), compared to regression line (dotted line) + 95% credible interval (light blue ribbon) from completely pooled data; C, D, E: ridgeline plots for parameters c_0, n_t, k_r, posterior densities: blue, black dots and lines: mean + 95% highest posterior density interval.

Table 3: Numerical summary of the posterior distribution resulting from bayesian multilevel modeling with varying initial concentration, reaction order, rate constant and correlation between them of the ascorbic acid data. SE = standard error, the lower and upper bound reflect the 95% credible interval

parameter	**mean**	**SE**	**lower bound**	**upper bound**
$\boldsymbol{c_0}$ (mM)	5.858	0.107	5.651	6.073
$\boldsymbol{n_t}$ $(-)$	0.913	0.113	0.686	1.134
$\boldsymbol{k_r}$ $\left(\mathrm{h}^{\boldsymbol{-1}}\right)$	0.615	0.039	0.539	0.697
$\boldsymbol{\sigma_u}$ (mM)	0.370	0.108	0.198	0.618
$\boldsymbol{\sigma_v}$ $(-)$	0.406	0.092	0.263	0.624
$\boldsymbol{\sigma_w}$ $\left(\mathrm{h}^{\boldsymbol{-1}}\right)$	0.098	0.055	0.007	0.218
$\boldsymbol{\rho_{u,v}}$	0.248	0.251	-0.278	0.685
$\boldsymbol{\rho_{u,w}}$	0.174	0.330	-0.511	0.749
$\boldsymbol{\rho_{v,w}}$	-0.067	0.323	-0.640	0.606
$\boldsymbol{\sigma_e}$ (mM)	0.186	0.016	0.158	0.221

Table 3 might require some further explanation. The parameter estimates for c_0, n_t, k_r are the overall estimates, which could also be called the population mean, or grand mean, resulting from multilevel modeling of all data sets (i.e., partial pooling). Note that the point estimate for the order n_t is now lower than 1 (though with a wide credible interval), in contrast to the result for completely pooled data. This is because the regressions have 'learned' from each other, information is shared, and some runs are more informative than others. The standard deviations $\sigma_u, \sigma_v, \sigma_w$ are the estimates for the random effects of the respective parameters, they estimate how much the parameters vary between repetitions (where the variation in c_0, in this case, is mainly due to slightly different experimental starting values). The correlation coefficients $\rho_{u,v}, \rho_{u,w}, \rho_{v,w}$ indicate the correlations between the random effect parameters. The correlation coefficients for the fixed effect parameters at the population level were already shown in the pairs and density plots in Figure 5. σ_e summarizes the residual standard deviation not explained by the model. An impression is thus obtained of the uncertainty in the random effect parameters characterizing the variation between repetitions. The correlation turns out to be rather low, but with quite some uncertainty. Also note that the residual standard deviation σ_e has decreased from 0.71 when the pooled data were analyzed (see Table 2) to 0.19 with the partially pooled data. This is because most of the variation is now accounted for in the standard deviations describing the random effects. This is one of the benefits of multilevel modeling: a better characterization of the sources of variation.

The resulting overall fit is in Figure 5B, compared to the fit obtained from completely pooled data with the 95% **credible** interval for the average (grand mean) effect. As a reminder, the credible interval reflects the uncertainty in estimating the mean at the population level. Three effects are

worth noting when comparing to the fit obtained from complete pooling. First, the regression line has shifted downwards going from regression using completely pooled to partially pooled data (also reflected in the different estimates for c_0, n_t and k_r in Table 3 as compared to Table 2). This is due to the 'borrowing effect,' experiments have 'informed' each other, resulting in a shift of parameter values. It illustrates the compromise that is reached between underfitting and overfitting: partial pooling combines all information that will give the best estimate of parameters. The second effect is that the 95% credible interval for the mean is much wider for regression with partially pooled data than for regression with completely pooled data. This does not mean that the situation has worsened because of applying multilevel modeling, it actually reveals that regression with completely pooled data underestimates the variation involved in estimating the mean. The reason behind that is that with completely pooled data dependencies are ignored, neglecting the fact that the data come from different repetitions and considering them all equal, thereby putting too much trust in the data. It is an important advantage of multilevel modeling that variation is much better characterized, and thus comes to more realistic estimates.

Figure 5C, D, E show the ridgeline plots for parameters c_0, n_t and k_r, respectively. These ridgeline plots can be compared to the ones obtained from no-pooling results (Figure 4D, E, F) and the differences are due to sharing information between runs. It is seen that especially the parameter n_t varies relatively more than the other two. The variation in estimated initial concentrations is rather small and fluctuates around the aimed value of 5.6 mM. The order of the reaction varied per experiment as shown in the n_t forest plot also after multilevel modeling. It should be realized that the order and rate constant are to some extent correlated, so variation in one parameter has an effect on variation in the other. The k_r forest plot shows the variation in the rate constant per run. If the reaction is exactly the same in each run, one would expect no variation here. But of course there is variation and it goes to show that there are, most likely, slightly uncontrolled (or uncontrollable) different varying conditions per run, which are reflected back in the value of the rate constant. It is known that ascorbic acid is quite sensitive to conditions like presence of transition metals, pH, and also oxygen, of course, though the oxygen concentration was kept constant in these experiments at 21%. Note, however, that the variation in k_r is less when applying multilevel modeling than when analyzing each experiment separately (compare Figure 5C, D, E with Figure 4C, D, E).

So, the lesson learnt here is that the rate constant as well as the order of reaction are quite sensitive to slightly varying conditions that apparently happen within each run. Even though the nature of these variations is not known exactly, there is at least a quantitative impression of how sensitive

the parameters are and it does pinpoint more clearly the type of variation in the repetitions (note the decrease in unexplained variance σ_e in going from the completely pooled data to partially pooled data). In passing, it may be good to realize what the difference is between the standard **errors** for the parameters and the values of the random effect standard **deviations** (note the difference in terminology). The standard errors and their associated credible intervals indicate the uncertainties in parameter values (given the data and the model), they do not indicate the actual variation. However, the standard deviations for the random effects attempt to estimate the actual variation in the population of parameters. It is a rather subtle but very important distinction. Note also that these standard deviations have a standard error of their own, indicating the uncertainty with which these parameters are estimated. What has been achieved with this multilevel modeling exercise is that the variation observed in a plot like Figure 1 is quantitatively and explicitly characterized.

As discussed above, variation in the value of the reaction order n_t has its consequences also for the value of the rate constant k_r because the two parameters are correlated. The results so far suggest that, overall, the order of reaction is not too far from $n_t = 1$ and it could therefore be interesting to compare the performance of the first-order model to that of the n^{th}-order model also with complete pooling and with multilevel modeling. Fixing the order eliminates its correlation with the rate constant and will therefore show the effect of no-pooling, complete pooling and partial pooling on the rate constant more clearly (the price to be paid for that is a lesser fit with trends in the residuals). The complete analysis is described in the Supplement: Figure S5 shows the results for complete pooling and Figure S6 for partial pooling with pairs and posterior parameter density plots, fits, and forest/ridgeline plots for the partially pooled results. Studying the multilevel modeling results using the first-order model, the variation in parameter c_0 is comparable to that with the nth-order model but the one for k_r is, obviously, different because of fixing the order at $n_t = 1$, showing how the rate constant parameter varies between runs at a fixed order $n_t = 1$. This variation is seen to be considerable, almost a factor 2, although it is also clear that run no. 9 is deviating strongly from the rest. The variation seen in k_r reflects the variation by unexplainable causes as it is 'felt' by and reflected in the rate constant. It is also striking to see that higher values of k_r go along with wider posterior densities. A higher k_r values reflects a higher reaction rate, so if the reaction goes faster it is apparently more difficult to estimate k_r. Perhaps, this could be remedied by adjusting the experimental design, such as taking more samples at the very beginning of the reaction.

3.4 Posterior predictive checks, model comparison and predictive accuracy

3.4.1 Posterior predictive checks.

Several models have now been tested and evaluated: the completely-pooled, single level n^{th}-order model, the partially pooled, multilevel n^{th}-order model, the completely pooled single level first-order model, and the partially pooled multilevel first-order model. It might be interesting to compare their performances. Posterior predictive checks basically measure how well predicted responses based on the posterior distribution correspond to what was actually measured: see Figure 6 where the empirical cumulative distribution function (ecdf) is compared between the actual data and simulated predictions. From these plots, the best predicting model seems to be the multilevel n^{th}-order model, followed by the first-order multilevel model while the difference between the two single-level models does not look too big. In other words, these latter two models, based upon complete pooling, seem to be less capable in predicting new values as compared to their multilevel companions.

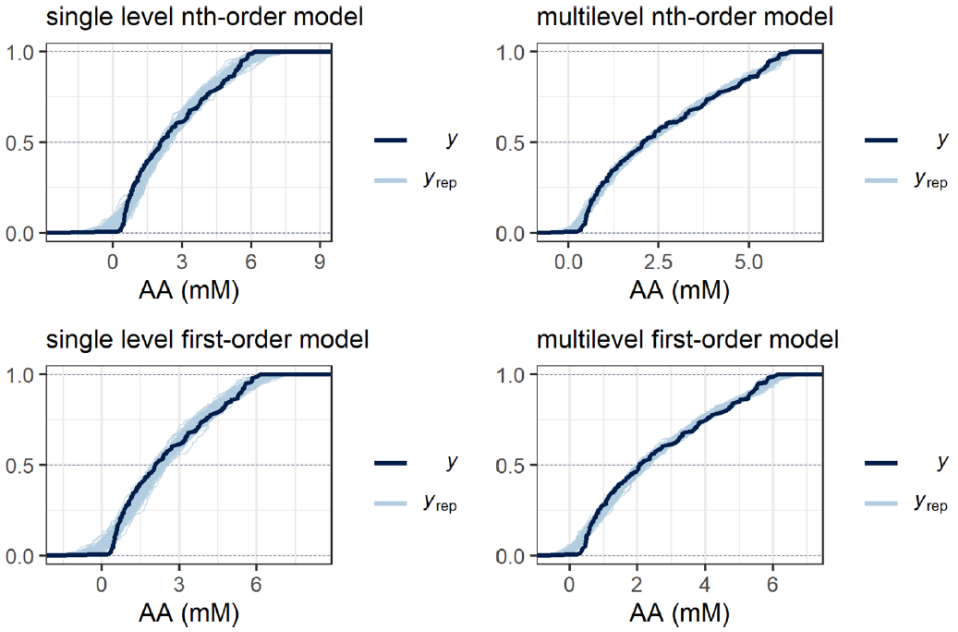

Figure 6. Posterior predictive checks of the 4 models describing degradation of ascorbic acid at 70 °C: 100 draws from the posterior predictive distribution (*yrep*=replicated from posterior distribution) providing empirical cumulative distribution function results (light blue lines, y-axis: density) compared to that of the data (dark blue lines, *y*).

3.4.2 Model comparison.

Besides posterior predictive checks, models can be compared in their predictive performance. The recommended technique by statisticians is to use loo-cv (leave-one-out-cross-validation) (Vehtari,

Gelman, & Gabry, 2017). This is a technique in which one sample is left out from the data set, the model is then refit without this one sample and it is evaluated how well the left-out observation is predicted. This is repeated for each data point. Because this is computationally expensive, an approximation is used called 'Pareto-smoothed importance sampling cross-validation' (PSIS in short). Details can be found in Gabry, Simpson, Vehtari, Betancourt, and Gelman (2019) and Vehtari et al. (2017). This technique was applied here, the software for doing this (the R package `loo`) is taken up in the R package `brms`. It is actually a test in how far models are able to predict new values and the errors they make in doing that. The metric output is the elpd-value ('expected log predictive density,' a measure of predictive accuracy, Bürkner, Gabry, and Vehtari (2021)), the model yielding the higher elpd value is performing better, according to this criterion; in addition, a standard error on the elpd values is also provided so that it can be judged in how far various elpd values really differ. Here, they are made visible in a plot, see Figure 7A. This analysis shows that the n^{th}-order multilevel model performs the best, followed by the multilevel first-order model, while the two less performing models are the single-level models based on completely pooled data. That the n^{th}-order model performs best is probably due to the fact that it allows more flexibility to find the most likely fit and prediction because it has one parameter more (n_t), and is therefore better able to predict future observations. This result also goes to show that the well-known modeling rule called Ockham's razor (cut down the number of parameters as much as possible) may not always be true in the case of multilevel models. According to McElreath (2020), the most important aspect to consider is finding the right balance between over- and underfitting, and multilevel modeling does exactly that.

A further advantage of the loo-cv-criterion is that it alerts for possible "problematic observations"; this happens if the "pareto-shape parameter k" is higher than 0.7. What this basically means is that the software is able to pinpoint observations that are very influential on the outcome. When applying this to the 4 models tested, the output for the two single-level, completely pooled models did not show a warning and reported that "All Pareto k estimates are good (k<0.5)." However, for the two multilevel models there were warnings. These outputs report 12 "bad cases" for the multilevel n^{th}-order model and 6 for the multilevel first-order model. A graphical output is given in Figure 7B, C for the multilevel models. It is even possible to ask the software specifically for the "bad cases." When this is done, it is striking to see that most of these cases concern the first measured values. The question is what to do with this information. The higher the Pareto-k-values, the more influential the data point is. For k-values > 0.7 they are considered problematic. One

possibility is to omit the pinpointed data points but that is only recommended when there are very good reasons to do so, for instance, if it is clear that something went wrong with these data points. That information is not available, so that will not be done here. In any case, these "bad cases" make the results less trustworthy.

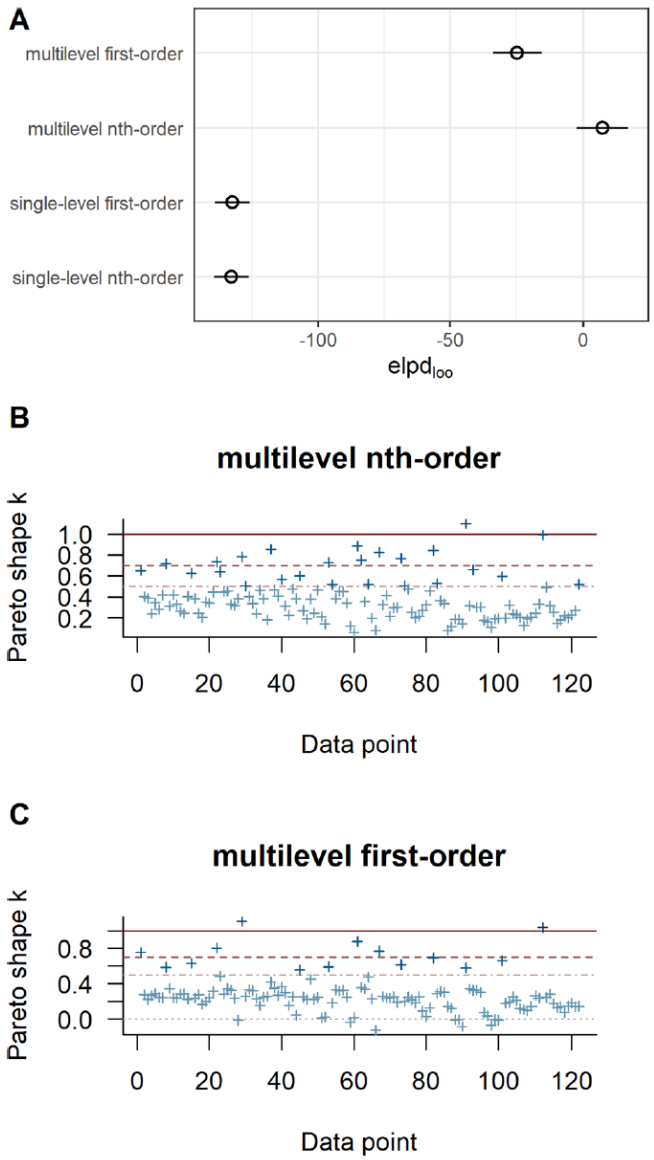

Figure 7. A: Plot of the elpd-loo values and their standard errors for the four ascorbic acid models. B: Diagnostic plot for the multilevel n^{th}-order model. C: Diagnostic plot for the multilevel first-order model. Plots B and C identify data that are troublesome according to the "Pareto Smoothed Importance Sampling (PSIS)" criterion. Values higher than 0.7 are considered problematic.

Yet another way of model comparison is to calculate relative weights according to WAIC (widely applicable information criterion). This kind of model comparison is based upon information theory; a well-known measure in the frequentist world is the Akaike Information Criterion (AIC). In the

Bayesian world, however, WAIC is the (more powerful) equivalent (McElreath, 2020). Weights can be used as a relative measure. When applying this method to the four models tested, the result is that the multilevel n^{th}-order model gets weight 1, and the other three get all weight 0, thus confirming the outcome of the loo-cv calculations. Despite the difficulties with the high pareto-k values in the loo-cv method, it can be concluded that the multilevel n^{th}-order model performs best. Incidentally, a less performing model should not be immediately discarded, it does not mean it is a bad model, only that it is less accurate in predicting.

3.4.3 Multilevel model prediction

An important goal of modeling is to make predictions, for instance, to predict process conditions and shelf life. Several graphs above showed already 95% prediction intervals. However, prediction becomes a bit more intricate in the case of multilevel models. It depends on which research question one wants to answer. Is it to predict new, not yet observed experiments? Or is it to predict the average effect on the population level rather than an individual experiment? Both possibilities exist. Figure 8A shows the 95% **prediction** interval at the population level and is compared to the one obtained from completely pooled data. The 95% prediction interval is much narrower with multilevel modeling using partially pooled data. This is due to the 'borrowing effect' where partial pooling has reduced the residual standard deviation (as already mentioned above: from 0.71 to 0.19) which consequently reduces the uncertainty in predicting new values at the population level. It is different, however, for predicting a **global mean value** (rather than new, not yet observed data). To illustrate that, a simulation is done to calculate the global mean ascorbic acid value after 1 h heating at 70 °C, with its uncertainty expressed as a probability density curve, using three models: based upon averaged-normalized data, completely pooled data and partially-pooled data. The results are shown in Figure 8B. It shows clearly that the uncertainty increases in going from averaged-normalized to completely pooled to partially pooled. It illustrates that the model based upon averaged-normalized data, and to a lesser extent the one based on completely pooled data, underfits; these two models do not use all the information that is present in the data. The partially pooled model leads to the highest uncertainty for estimating **the population mean** because it makes optimal use of all the information in the data and takes dependencies in the data into account. It thereby gives the most realistic impression of the uncertainty involved.

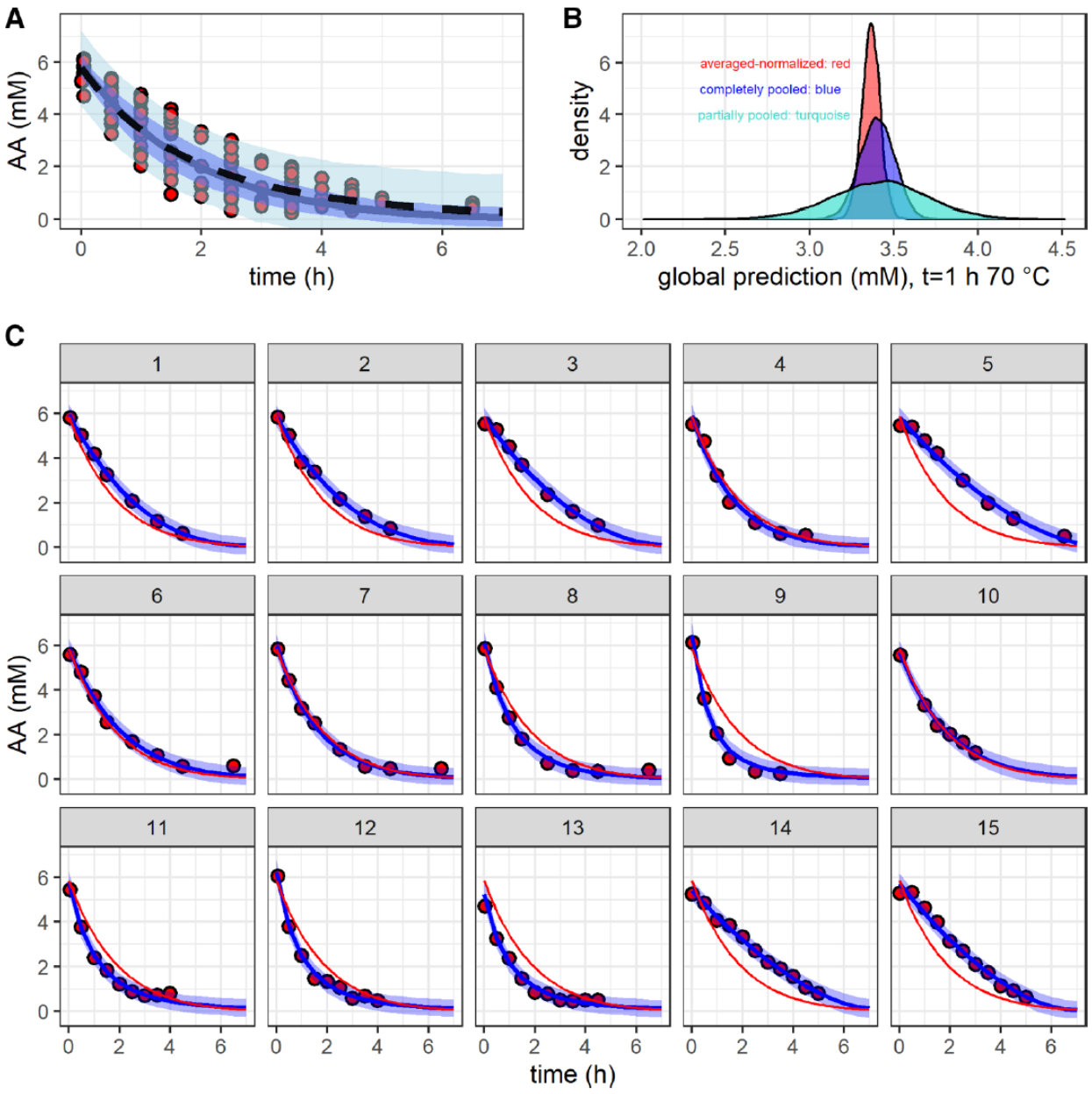

Figure 8. Multilevel modeling result (n^{th}-order model) for the ascorbic acid data. A: Partially pooled data, population level regression line (solid line) + 95% prediction interval (dark blue ribbon), compared to regression line (dotted line) + 95% prediction interval (lightblue ribbon) of completely pooled data; B: Global predicted value using models based upon averaged-normalized data (red), completely pooled data (blue), partially pooled data (turquoise). C: Regression result for the partially pooled individual level (blue lines) and population level (grand mean, red line).

Besides the fact that there is now detailed information about the population level, this is also the case for the individual runs. In other words, there are estimates for the three parameters available for each of the runs, based on sharing information between the runs. Since the individual parameter estimates are also available, the individual fits can be shown for each run: see Figure 8C. The blue ribbons reflect the 95% prediction intervals around the individual regression line. The individual

fits resulting from partial pooling look quite good, including their prediction intervals, but it also shows that some individual fits deviate substantially from the grand mean. It is instructive to see how far these fits deviate from the individual fits where the information between runs is not shared (no-pooling). This shows whether or not the phenomenon of "shrinkage" has occurred: the partially pooled regression line for an individual run is shifted ("shrunk") towards the grand mean. This is best shown for runs that were quite different from the grand mean, for instance run 5 and 9: see Figure 9.

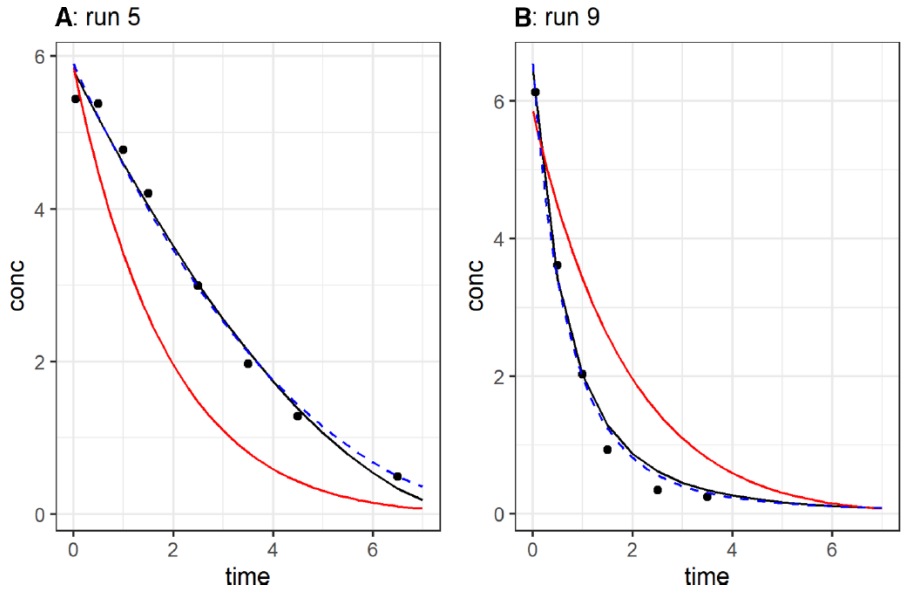

Figure 9. Regression lines according to the n^{th}-order model for experiment 5 (A) and 9 (B), showing the fits resulting from regression of the individual dataset (no-pooling, hyphenated line), the regression line resulting from the partially pooled data (solid black line) and the grand mean regression line resulting from the partially pooled data (red line).

In run 5, shrinkage is seen most clearly at the end as the partially pooled regression line (solid black) being pulled slightly down towards the grand mean regression line (red line) as compared to the individual regression line (black hyphenated). In experiment 9 the partially pooled regression line (solid black) is moved up a little towards the grand mean regression line (solid red line) as compared to the individual regression line (hyphenated blue). The shrinkage effect is not too strong in this particular case of ascorbic acid degradation because each run contains quite a number of data points, which will keep the partially pooled regression line close to the individual regression line. Should there have been runs with less data points, shrinkage would have been more clear for such runs. This shrinkage effect is desired because it counteracts the phenomenon of over- and underfitting. It illustrates that with multilevel modeling the emphasis is less on fitting

("retrodiction") because the global fit for each group is less in comparison to each individual fit, but it will be much more powerful in prediction. Shrinkage is the result of a trade-off of a poorer fit at the individual level and better predictive power at the global level (McElreath, 2020). Moreover, if so desired, fits at the individual level can also be obtained and they are shown to be quite good.

4 Conclusion

This article has illustrated the application of multilevel modeling to kinetic data analysis. The data sets available for ascorbic acid degradation are unique in the sense that 15 different experimental runs were available at the same conditions. In actual practice, that many runs will not be available often but a number of, say, 5-6 runs is already enough to do multiresponse modeling. The case study illustrates various aspects:

- kinetic analysis of averaged results is not recommended because it is like throwing away useful information with the result that actual variation is strongly underestimated
- analysis of completely pooled data also underestimates actual variation
- analysis of data that are not pooled overestimates actual variation
- multilevel modeling of partially pooled data allows to partition variance over parameters, thereby diminishing unexplained variance and characterizing variation of parameters in a realistic way
- predictions using multilevel modeling estimates lead to less uncertainty (better precision) for new, to be observed data because of reduction of unexplained variance but to more uncertainty (lower precision) for global mean values because of increased variance of parameters

These results have implications for calculations like shelf life estimations. Such predictions will be done for global mean values and then it is important to have a realistic impression of the uncertainties involved, an impression that is best obtained by multilevel modeling using partial pooling. It has been shown that models based on averaged-normalized data and completely pooled data do lead to underestimation of the uncertainties involved. Though the magnitudes of such

effects depend of course on the problem at hand, it is a principled approach to apply multilevel modeling whenever possible.

Clearly, multilevel modeling requires more work. A substantial amount of repetitions is needed (5-6 at least), the modeling exercise is more difficult and requires some programming expertise. However, the reward of investing in this is that much more insight is obtained and results become more useful and trustworthy. The insight in sources of variation resulting from multilevel modeling may help to design experiments to reduce such variation if that is required. The emphasis on predictive accuracy of models cannot be underestimated when modeling results need to be used in practice, such as shelf life modeling.

Furthermore, this paper has illustrated how Bayesian regression can help in visualizing posterior parameter distributions so that the researcher gets an impression of their behaviour. Also, working in the Bayesian way forces researchers to be explicit about their modeling assumptions, which would be a good development as these assumptions are usually not specified. Bayesian regression leads to better insight in the implications of the assumptions than the frequentist approach. Most importantly, predictive capacity and accuracy of models can be tested quantitatively, making it possible to go beyond fitting/retrodiction to which most food science publications seem to limit themselves. Uncertainties in parameters cannot only be characterized as such but also be used quantitatively in further calculations where these uncertainties are propagated.

There are, as yet, not many data published in food science literature that allow further exploration of the concept presented here. This article is meant to inspire researchers to pay attention to this important aspect of variation now that the modeling techniques are available. In the words of McElreath (2020), “When it comes to regression, multilevel regression deserves to be the default approach.”

Acknowledgement

The authors gratefully acknowledge the contribution of dr. Gomez Ruiz for performing the experimental analyses.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

M.A.J.S. van Boekel: Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing – original draft.

S. Roux: Conceptualization, Methodology, Visualization, Investigation, Resources, Data curation, Writing – review & editing.

Supplementary information

The Supplement (at the end of this document) gives additional information on the kinetic analysis of the ascorbic acid data.

The R code and data sets can be found at the GitHub page of the first author: https::\\github.com\TinyvanBoekel\FRI

References

Al Fata, N., Georgé, S., André, S., & Renard, C. M. G. C. (2017). Determination of reaction orders for ascorbic acid degradation during sterilization using a new experimental device: The thermoresistometer Mastia. *LWT - Food Science and Technology*, *85*, 487–492. https://doi.org/10.1016/j.lwt.2016.08.043

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). https://doi.org/10.18637/jss.v067.i01

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. Retrieved from http://arxiv.org/abs/1701.02434

Bürkner, P. C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. Retrieved from http://arxiv.org/abs/1705.11123

Bürkner, P. C., Gabry, J., & Vehtari, A. (2021). Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. *Computational Statistics*, *36*(2), 1243–1261. https://doi.org/10.1007/s00180-020-01045-4

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Garre, A., Zwietering, M. H., & Den Besten, H. M. W. (2020). Multilevel modelling as a tool to include variability and uncertainty in quantitative microbiology and risk assessment. Thermal inactivation of Listeria monocytogenes as proof of concept. *Food Research International*, *137*, 109374. https://doi.org/10.1016/j.foodres.2020.109374

Gelman, A., Carlin, J. B., Stern, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis 3rd Edition* (p. 675). Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (p. 651). Cambridge: Cambiridge University Press.

Gelman, A. Simpson, D., Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. Entropy, 19 (10), 555. https://doi.org/10.3390/e19100555

Giannakourou, M. C., & Taoukis, P. S. (2021). Effect of Alternative Preservation Steps and Storage on Vitamin C Stability in Fruit and Vegetable Products: Critical Review and Kinetic Modelling Approaches. *Foods*, *10*, 2630. https://doi.org/10.3390/foods10112630

Gómez Ruiz, B., Roux, S., Courtois, F., & Bonazzi, C. (2018). Kinetic modelling of ascorbic and dehydroascorbic acids concentrations in a model solution at different temperatures and oxygen contents. *Food Research International*, *106*, 901–908. https://doi.org/10.1016/j.foodres.2018.01.051

Hickman, D. A., Ignatowich, M. J., Caracotsios, M., Sheehan, J. D., & D'Ottaviano, F. (2018). Nonlinear mixed-effects models for kinetic parameter estimation with batch reactor data. *Chemical Engineering Journal*, *377*, 119817–119824. https://doi.org/10.1016/j.cej.2018.08.203

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis* (2nd ed., p. 759). Academic Press.

Lambert, B. (2018). *A student's guide to Bayesian statistics* (p. 498). London: SAGE publications Ltd.

Lazic, S. E., Clarke-Williams, C. J., & Munafò, M. R. (2018). What exactly is 'N' in cell culture and animal experiments? *PLoS Biology*, *16*(4). https://doi.org/10.1371/journal.pbio.2005282

Lazic, S. E., Mellor, J. R., Ashby, M. C., & Munafo, M. R. (2020). A Bayesian predictive approach for dealing with pseudoreplication. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-59384-7

Matheson, G. (2020). Nonlinear modelling using nls,nlme and brms, a.k.a. when straight lines don't provide enough of a thrill any longer. Retrieved from: Nonlinear Modelling using nls, nlme and brms-Granville Matheson's Blog

McElreath, R. (2020). *Statistical Rethinking. A Bayesian course with examples in R and Stan. Second Edition* (p. 612). Boca Raton: CRC Press.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., & R Core Team (2017). nlme: Linear and nonlinear mixed effects models. Retrieved from https:// CRAN.R-project.org/package=nlme

Shen, J., Griffiths, P. T., Campbell, S. J., Utinger, B., Kalberer, M., & Paulson, S. E. (2021). Ascorbate oxidation by iron, copper and reactive oxygen species: review, model development, and derivation of key rate constants. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-86477-8

Van Boekel, M. A. J. S. (2020). On the pros and cons of Bayesian kinetic modeling in food science. *Trends in Food Science & Technology*, *99*, 181–193. https://doi.org/10.1016/j.tifs.2020.02.027

Van Boekel, M. A. J. S. (2021a). Kinetics of heat-induced changes in foods: a workflow proposal. *Journal of Food Engineering*, *306*(April), 110634. https://doi.org/10.1016/j.jfoodeng.2021.110634

Van Boekel, M. A. J. S. (2021b). To pool or not to pool: That is the question in microbial kinetics. *International Journal of Food Microbiology*, (June), 109283. https://doi.org/10.1016/j.ijfoodmicro.2021.109283

Van Boekel, M. A. J. S. (2022). Kinetics of heat-induced changes in dairy products: developments in data analysis and modelling techniques. *International Dairy Journal*, *9*, 105187. https://doi.org/10.1016/j.idairyj.2021.105187

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Supplement to: Multilevel modeling in food science: a case study on heat-induced ascorbic acid degradation kinetics

M.A.J.S. van Boekel[1] & S. Roux[2]

[1] Food Quality & Design Group, Wageningen University & Research, Wageningen, the Netherlands

[2] Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 91300, Massy, France

Supplement to: Multilevel modeling in food science: a case study on heat-induced ascorbic acid degradation kinetics

Introduction

This Supplement gives additional information about the multilevel-modeling results of the kinetic analysis of ascorbic acid degradation. A schematic example of a nested hierarchical design is illustrated in Figure S1.

The raw data are shown in the main article in one graph (Figure 2), here the individual experiments are shown so that each run can be seen separately: see Figure S2. It shows similarities (nonlinear decrease over time) but also differences (in initial concentrations, shapes of the curves and rates are clearly different) between the runs.

First-order model of the individual data sets (no pooling).

The authors of the original article analyzed 9 data sets using a first-order model with the non-normalized raw data leading to 9 estimates of rate constants k_r and initial values c_0 (these results were reported in their Supplement to the article). This analysis is repeated here in the Bayesian way, including the 6 new additional data sets that became available. The numerical outcomes of the Bayesian regression correspond to those of the least-squares regression (results not shown). A representative example of a pairs and correlation plot can be found in Figure S3A, which all looks well-behaved. The first-order fits do not look bad at first sight (Figure S3B) but when studying the residuals, trends are visible in quite some cases, see Figure S3C, casting doubt on the assumption that degradation occurs according to a first-order model.

Individual fits (no-pooling) using the n^{th}-order model

The parameters c_0, k_r and n_t were estimated for each run separately, resulting in 15 estimates of the order of the reaction, the initial concentration and the rate constant using again the n^{th}-order model. A representative example of a pairs and posterior parameter density plots can be found in Figure S4, showing the rather strong negative parameter correlation between n_t and k_r, which is due to the mathematical formulation of the model; this parameter correlation is not that high that it disturbed parameter estimation. The skewed posterior parameter distributions are noteworthy. The resulting fits (retrodictions) can be found in Figure 4A in the main article with the residual plots are shown in Figure 4B. The fits look quite adequate and the residuals behave better than with the first-order model, confirming that a first-order model is not the most suitable one if judged from each individual run. The variation in the order of reaction per run is shown in the main article in Figure 4D.

Single level complete pooling, assuming first-order kinetics (no averaging, no normalization)

In the original article by Gómez Ruiz et al. (2018), first-order kinetics was assumed throughout, while analysis was done in the frequentist way (nonlinear least-squares regression). First-order kinetics analysis is repeated here but then in the Bayesian way and without normalizing and averaging concentrations. This will allow to compare the results from the n^{th}-order kinetics (shown in the main article) with the analysis assuming first-order kinetics. The resulting pairs and posterior density plots for the first-order analysis of the completely pooled data, the fit, the residuals, QQ-plot and lag plot are in Figure S5.

Multilevel modeling using the first-order model

Variation in the value of the reaction order n_t has its consequences also for the value of the rate constant k_r because the two are correlated. Despite the observed variation in the order of reaction per run, it could be interesting to fix the order at 1 (as was done in the original article) and to compare the performance of the first-order model with the n^{th}-order model when analyzed according to the multilevel analysis. The proposed likelihood and priors for this case are then as in equation (1):

$$\begin{aligned} c_t &\sim \mathcal{N}(\mu_i, \sigma_e) \\ \mu_i &= c_0 \cdot \exp(-k_r \cdot t) \\ c_0 &\sim \mathcal{N}(6, 1) \\ k_r &\sim \mathcal{N}(0.5, 0.3) \quad (\text{lb}{=}0) \\ \sigma_{c_0} &\sim \text{half-cauchy}\,(10) \\ \sigma_{k_r} &\sim \text{half-cauchy}\,(10) \\ \sigma_e &\sim \text{half-cauchy}(10) \\ \rho &\sim \text{LKJ}(2) \end{aligned} \tag{1}$$

The pairs and density plots for the multilevel first-order model can be found in Figure S6A. This looks well behaved and the correlations are quite low. The accompanying fits are shown in Figure S6B, while the resulting ridgeline plot for parameters c_0 and k_r are to be found in Figure S6C. While the pairs and density plots look OK, the fits show a mismatch, as is to be expected because it was already shown that the order is not exactly one. What is interesting, however, is to observe how the rate constant behaves as it is now no longer correlated to the order n_t. It is seen to vary between the individual runs by almost a factor two. This reflects the sensitivity of the degradation reaction to slightly changing conditions that is picked up by the rate constant.

References

Gómez Ruiz, B., Roux, S., Courtois, F., & Bonazzi, C. (2018). Kinetic modelling of ascorbic and dehydroascorbic acids concentrations in a model solution at different temperatures and oxygen contents. *Food Research International*, *106*, 901–908. https://doi.org/10.1016/j.foodres.2018.01.051

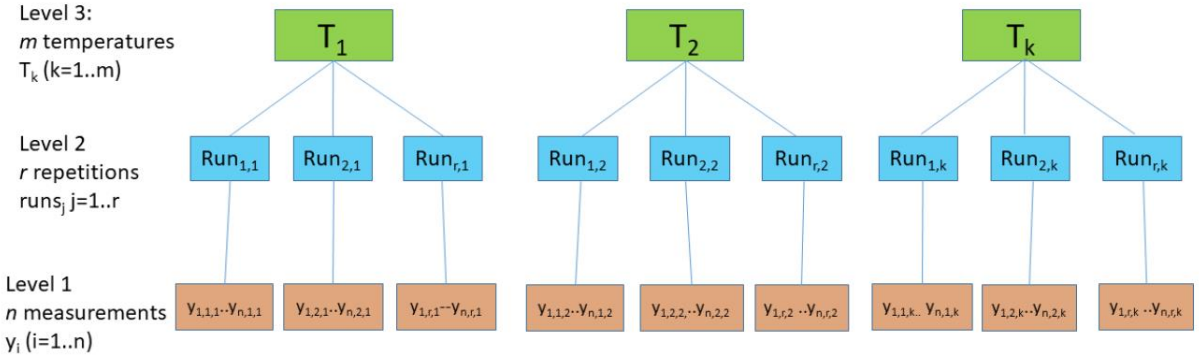

Figure S1. Schematic overview of a common experimental design with repetitions of isothermal experiments at m temperatures ($T \quad (1..m)$), various r repetitions at the same temperature (run $(1..r)$) and n measurements ($y \quad (1..n)$) in each run. The three levels that can be distinguished form an example of a nested, hierarchical design. (In the current article only level 1 and 2 are studied for experiments done at a single temperature of $T = 70$ °C.)

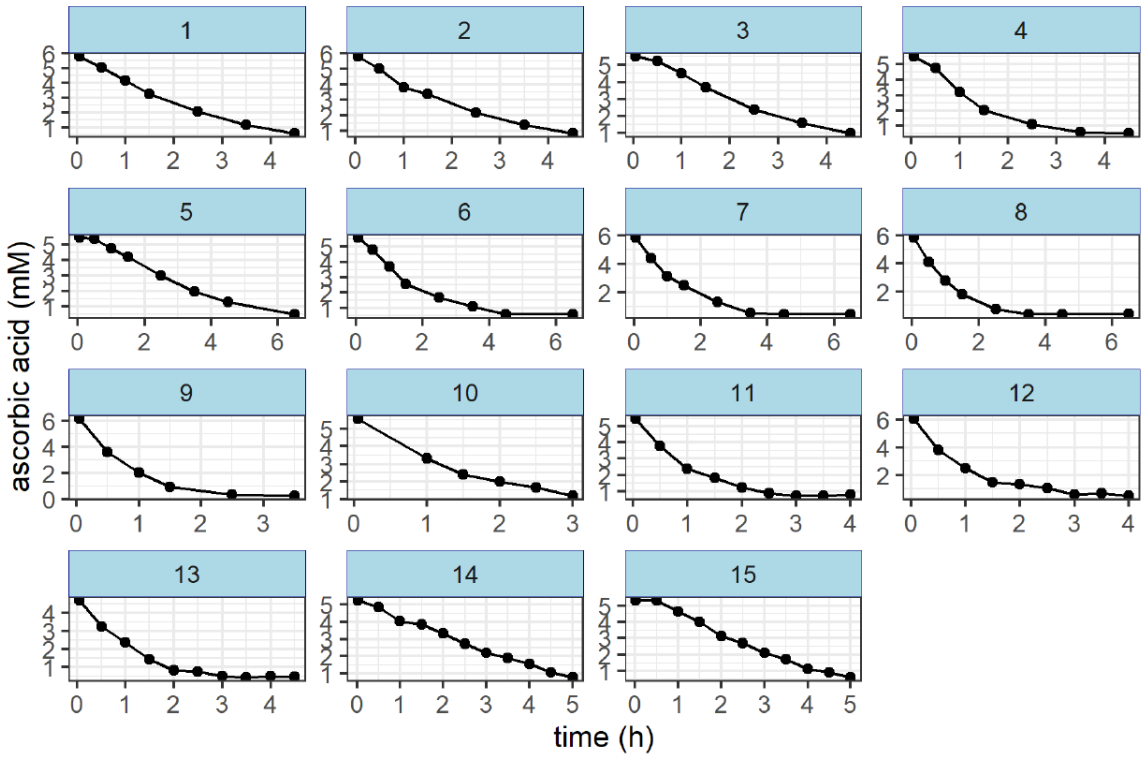

Figure S2. Plot of the change in ascorbic acid concentration during heating at 70 °C for each of the 15 runs (the lines are just to guide the eye). Source: Gómez Ruiz, Roux, Courtois, and Bonazzi (2018).

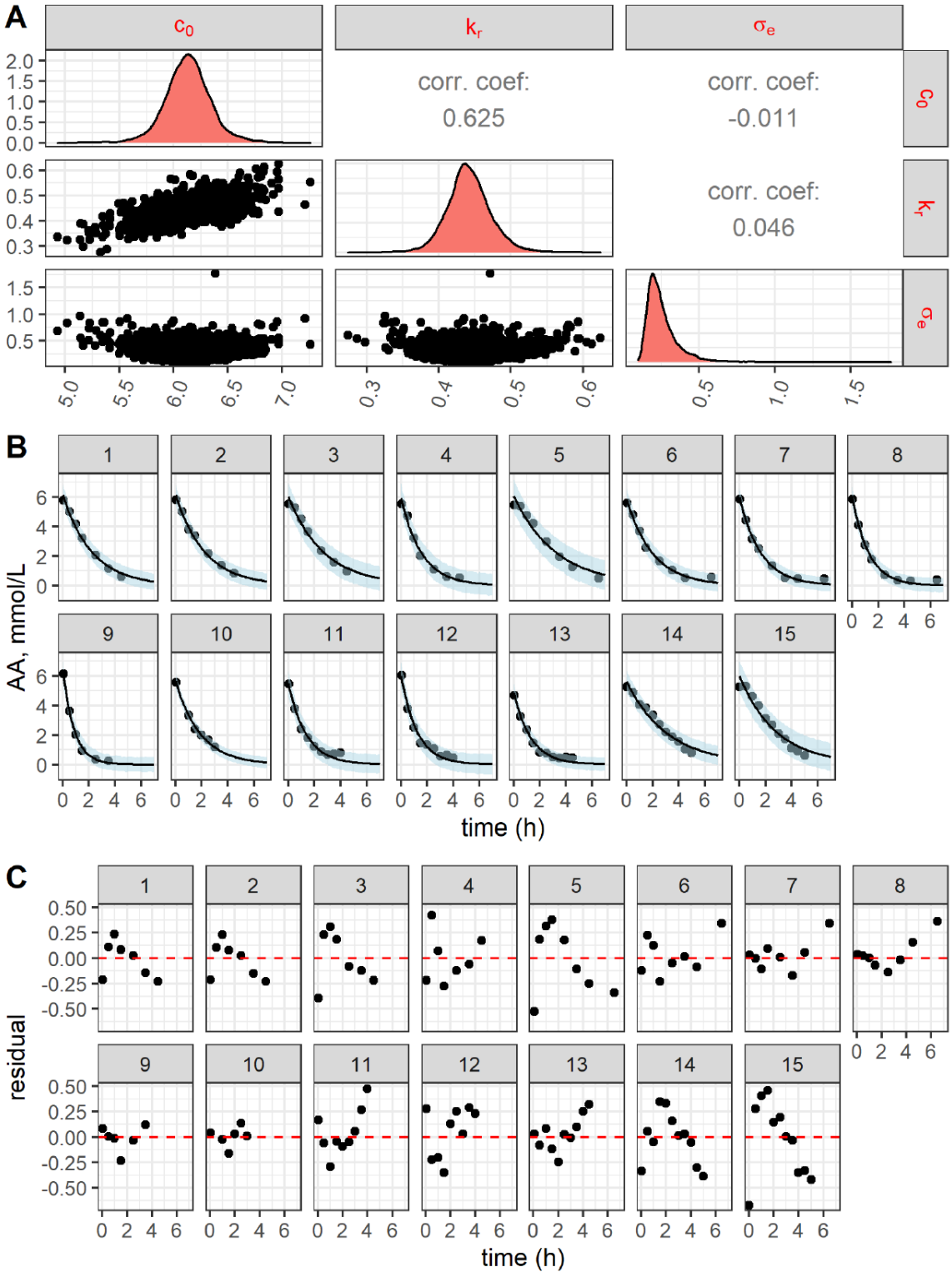

Figure S3. Bayesian regression results of each individual data set using the first-order model for the degradation of ascorbic acid (AA) at 70 °C (no pooling). A: posterior parameter densities for run 1 by way of example (other runs showed similar behavior); B: Fits for each run (blue ribbon: 95% prediction interval); C: residuals for each run.

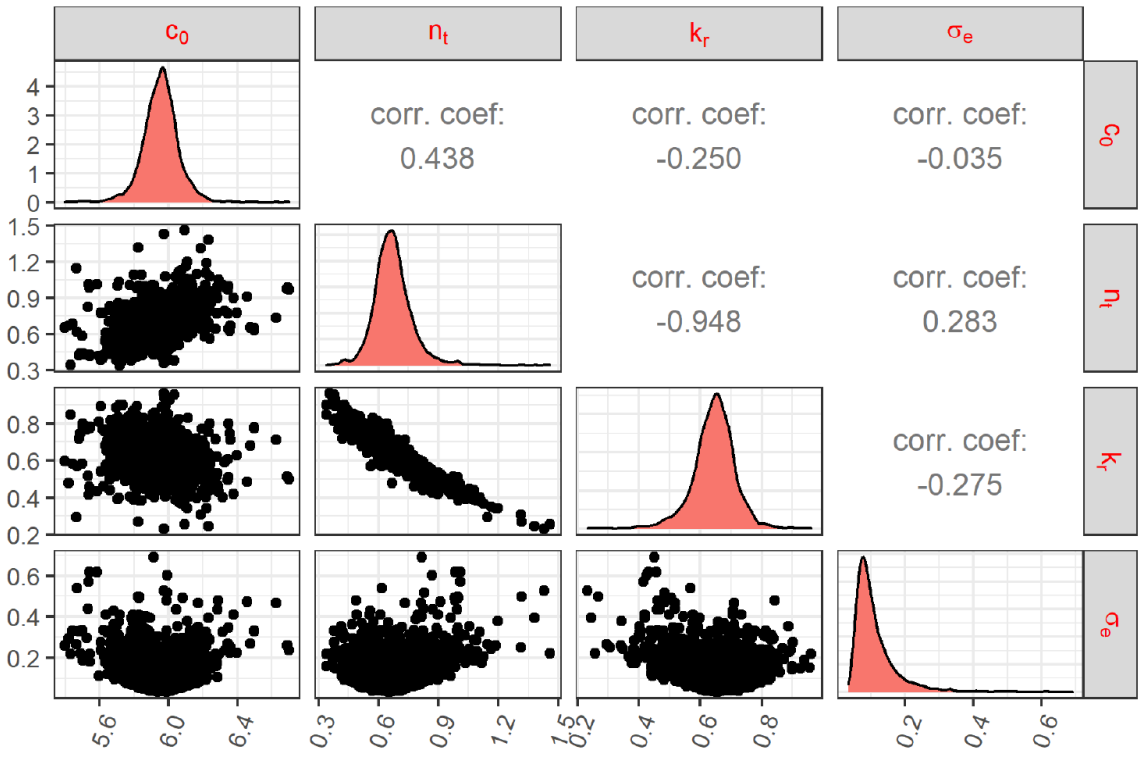

Figure S4. Pairs and density plots and correlation coefficients for run 1 of ascorbic acid degradation, using the n^{th}-order model. Similar plots were obtained for the other runs (results not shown).

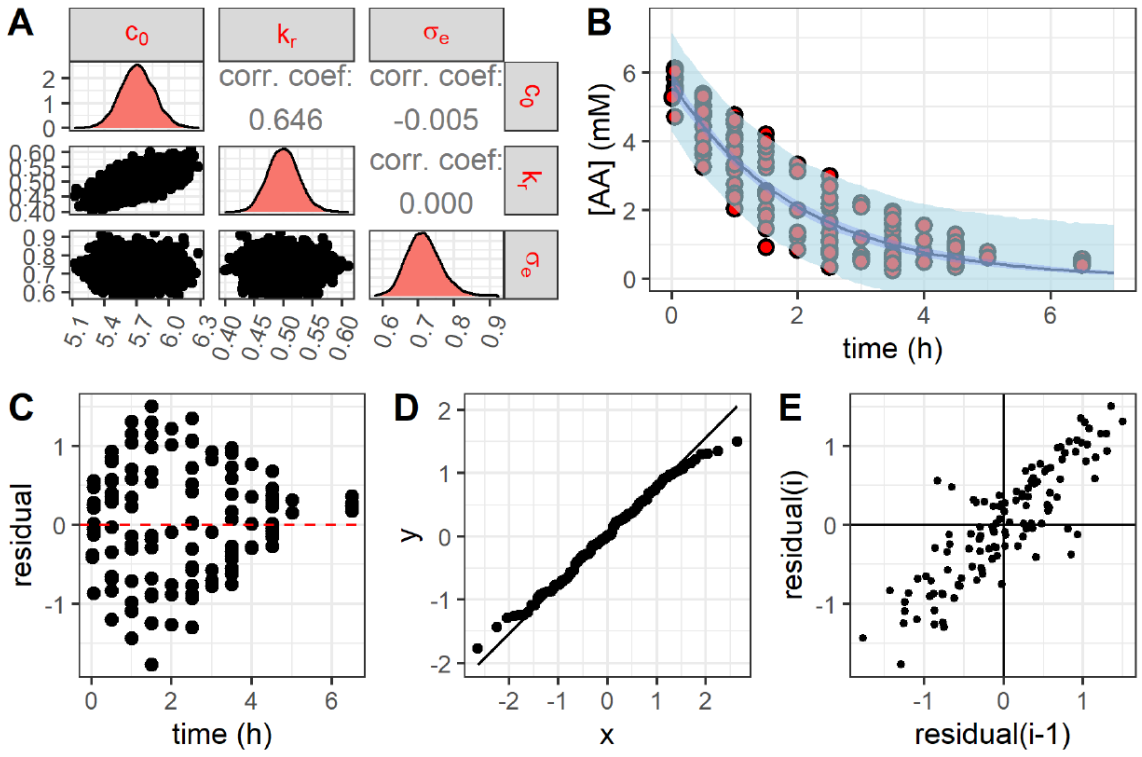

Figure S5. Bayesian regression results for the pooled (not averaged) ascorbic acid data heated at 70 °C using the first-order model. A; Pairs and posterior parameter density plots; B: regression line + 95% credible interval (dark-blue ribbon) + 95% prediction interval (light blue ribbon); C: residuals; D: QQ plot; E: lagplot.

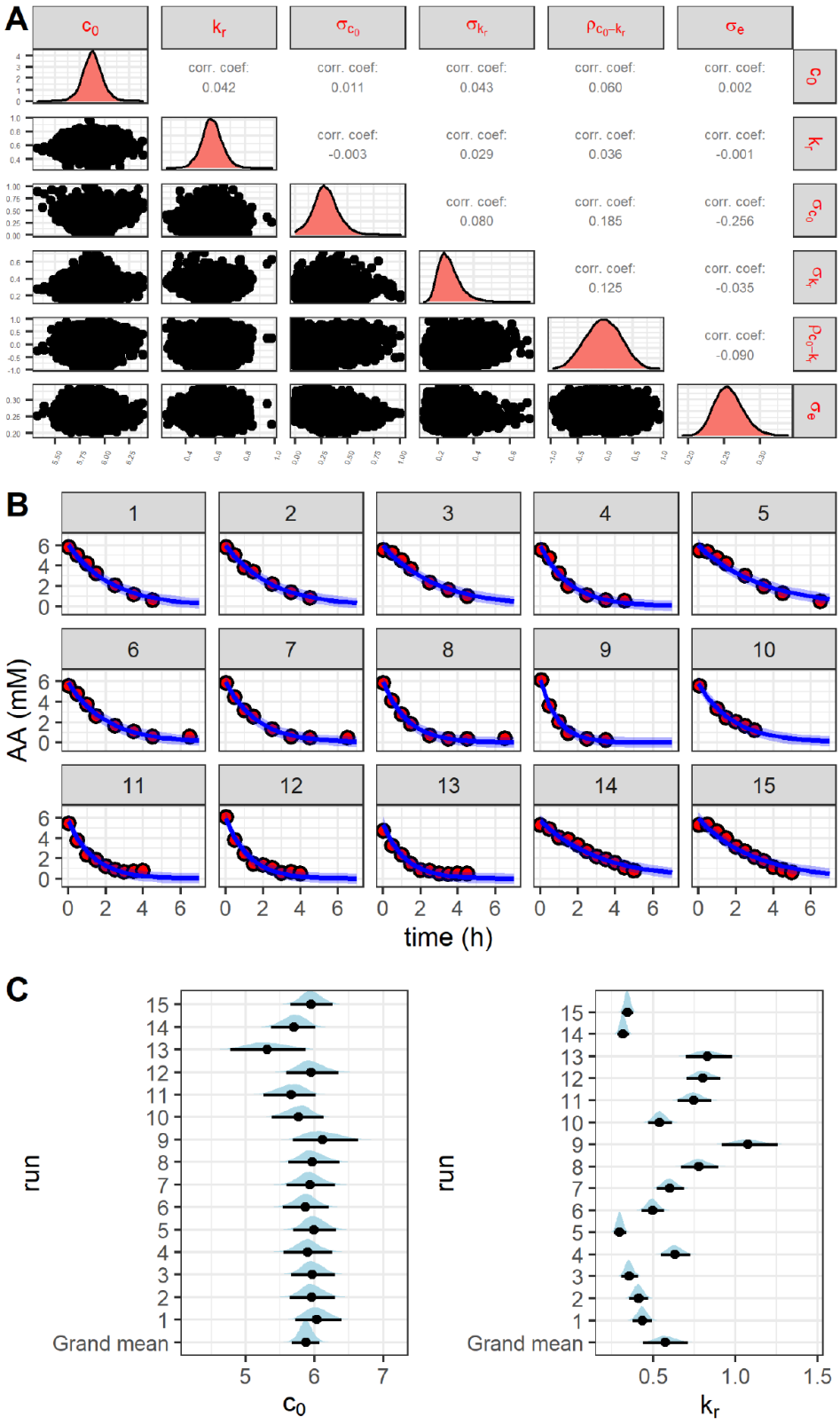

Figure S6. Bayesian multilevel regression (partial pooling, first-order model). A: posterior parameter density plots, pairs plots and correlation coefficients. B: Fits per run with 95% prediction interval (blue ribbon). C: forest/ridgeline plot for parameters c_0 and k_r, posterior parameter densities: blue, dots + black lines: mean + 95% highest density interval.