

Bayesian Nonlinear Mixed Effects Models

Subjects: Computer Science, Artificial Intelligence | Statistics & Probability | Computer Science, Theory & Methods

Contributor: Se Yoon Lee

Nonlinear mixed effects models have become a standard platform for analysis when data is in the form of continuous and repeated measurements of subjects from a population of interest, while temporal profiles of subjects commonly follow a nonlinear tendency. While frequentist analysis of nonlinear mixed effects models has a long history, Bayesian analysis of the models has received comparatively little attention until the late 1980s, primarily due to the time-consuming nature of Bayesian computation. Since the early 1990s, Bayesian approaches for the models began to emerge to leverage rapid developments in computing power, and have recently received significant attention due to (1) superiority to quantify the uncertainty of parameter estimation; (2) utility to incorporate prior knowledge into the models; and (3) flexibility to match exactly the increasing complexity of scientific research arising from diverse industrial and academic fields.

Keywords: Bayesian Nonlinear Mixed Effects Models

1. Introduction

One of the common challenges in biological, agricultural, environmental, epidemiological, financial, and medical applications is to make inferences on characteristics underlying profiles of continuous, repeated measures data from multiple individuals within a population of interest ^{[1][2][3][4]}. By 'repeated measures data' researchers mean the data type generated by observing a number of individuals repeatedly under differing experimental conditions, where the individuals are assumed to constitute a random sample from a population of interest. A common type of repeated measures data is longitudinal data such that the observations are ordered by time ^{[5][6]}.

Linear mixed effects models for repeated measures data have become popular due to their straightforward interpretations, flexibility allowing correlation structure among the observations, and utility accommodating unbalanced and multi-level data structure (i.e., clustered designs that vary among individuals) ^{[7][8]}. The modeling framework is also intuitively appealing: the central idea that individuals' responses are governed by a linear model with slope or intercept parameters that vary among individuals seems to be appropriate in many scientific problems (for, e.g., see ^{[9][10]}). It also allows practitioners to test and evaluate multivariate causal relationships by conducting regression analysis at the population level. By preserving the multi-level structure in a single model, estimation or prediction for the analyses can take advantage of information borrowing ^[11].

For many applications, researchers often want to theorize that time courses of individual response commonly follow a certain nonlinear function dictated by a finite number of parameters ^[12]. These nonlinear functions are based on reasonable scientific hypotheses, typically represented as a differential equation system. By tuning the parameters, the shape of the function in terms of curvature, steepness, scale, height, etc., may change, which is used as the rationale behind describing heterogeneity between subjects. Nonlinear mixed effects models, also referred to as hierarchical nonlinear models, have gained broad acceptance as a suitable framework for these purposes ^{[13][14][15]}. Analyses based on this model are now routinely reported in various industrial problems, which is, in part, enabled by the breakthrough development of software ^{[16][17][18][19][20]}. The excellent books and review papers were published by ^{[14][15][21]}. Although their works were published more than 20 years ago, they still provide statisticians, programmers, and researchers with many pedagogical insights about the modeling framework, implementations, and practical applications of using the nonlinear mixed effects models.

While frequentist analysis of nonlinear mixed effects models has a long history, Bayesian analysis for the models was a relatively dormant field until the late 1980s. This is due primarily to the time-consuming nature of the calculations required for Bayesian computation to implement a Bayesian model ^[22]. Since the early 1990s, Bayesian approaches began to re-emerge, motivated both by exploitation of rapid developments in computing power and by the growing desire to quantify the uncertainty associated with parameter estimation and prediction ^{[23][24][25]}. Since then, Bayesian nonlinear mixed effects models, also called Bayesian hierarchical nonlinear models, have been extensively used in diverse industrial and

academic researches, endowed with new computational tools providing a far more flexible framework for statistical inference exactly matching the increasing complexity of scientific research [26][27][28][29][30][31].

2. Trends and Workflow of Bayesian Nonlinear Mixed Effects Models

2.1. Rise in the Use of Bayesian Approaches for the Nonlinear Mixed Effects Models

As of January 1970 to December 2021, PubMed.gov (<https://pubmed.ncbi.nlm.nih.gov/>, accessed on 23 February 2022) searched of “nonlinear mixed-effect models” yielded 6288 publications. Among the published articles, nearly 94% of works used frequentist approaches (5929 articles), while only 6% of works adopted Bayesian approaches (359 articles).

The dormancy of the Bayesian approaches until the late 1980s is mainly due to the time-consuming nature of the calculations based on a sampling scheme, which was previously impossible by the limitation of computing power. Fortunately, a breakthrough in computer processors (for, e.g., Am386 in 1991, Pentium Processor in 1993, etc.) took place in the early 1990s, driving the computing revolution to solve computationally intense problems, and this has given the statistical community the ability to solve statistical questions by using Bayesian methods. This timeline is also aligned with the widespread Markov chain Monte Carlo (MCMC) sampling techniques in the Bayesian community [32][33]. Since then, the Bayesian community has been gradually gaining the momentum to leverage the rapidly growing developments of computing power, and now, assorted Bayesian software packages (e.g., JAGS [34], BUGS [35], and Stan [17]) are available for researchers to answer scientific questions arising from industrial and academic research.

To understand the rise of the Bayesian approaches, it should be understood what will be some of the advantages of using Bayesian methods over frequentist methods in the context of nonlinear mixed effects models. As the primary focus is to provide the readers with some insight on methodologies and practical implementation of using Bayesian approaches, the comparison and exposition below are described from an operational viewpoint.

The stark difference between using frequentist and Bayesian approaches may be the procedure of describing an uncertainty underlying the parameter estimation for the nonlinear mixed effects models. Here, the parameter which is of primary interest is the population-level parameters (also called fixed effects), typical values for the individual-level parameters. In many cases, frequentist

confidence intervals for the parameters of the models are constructed by assuming that asymptotic normality of maximum likelihood estimator holds in a finite sample study, which is actually the most accurate in large sample scenario [36]. Most frequentist software packages, such as NONMEM [13][37], Monolix [38], and nlmixr [18], by default, may print out a 95% confidence interval of the form, “*Estimate* $\pm 1.96 \times$ *Standard Error*”, or some transformation of the lower and upper bounds, if necessary, such that the *Standard Error* is calculated by using (observed) Fisher information matrix [39][40][41]. Using such a scheme in small sample studies is highly likely to overlook the gap between the reality of the data and the idealistic asymptotic situation.

In contrast, as for the Bayesian approaches, the large-sample theory is not needed for the uncertainty quantification, and the procedure to obtain 95% posterior credible intervals is a lot easier than obtaining 95% confidence intervals (See Chapter 4 of [42]). Furthermore, Bayesian credible intervals based on percentiles of posterior samples allow for a strongly skewed distribution, wherein frequentist confidence intervals (based on large-sample theory) may induce a non-negligible approximating error due to the deviation from the asymptotic normality. Along with that, Bayesian methods are highly appreciated when researchers wish to incorporate prior knowledge from previous studies into the model so that posterior inference provides the researchers with an updated view on the problem, possibly with a more accurate estimation. Using prior information would be particularly useful in small-sample contexts [43][44]. For example, in medical device clinical trials, some opportunities and challenges in developing a new medical device are: (i) there is often a great deal of prior information for a medical device; (ii) a medical device evolves in relatively small increments from previous generations to a new generation; (iii) there are only a few numbers of patients for the trials; and (iv) companies need to make a rational decision promptly to reduce cost. In those settings, Bayesian methods have been demonstrated to be suitable, and their proper use is guided by Food and Drug Administration [45][46].

2.2. Bayesian Workflow

Researchers outline the first two steps in the Bayesian workflow of using Bayesian nonlinear mixed effect models described in **Figure 1**. The panel includes some mathematical notations that are consistently used throughout the paper. These notations will be clearly understood later. The aim of explanation at this point is to provide readers with a blueprinted plan to implement Bayesian modeling strategies for the nonlinear mixed effects models. Researchers assume

that readers are familiar with basic concepts and generic workflow in Bayesian statistics; see [42][47][48] for those basic concepts and refer to the review paper by [49] and references therein for detailed concepts and general terminologies used in workflow, such as prior and posterior predictive checks, and prior elicitation, etc.

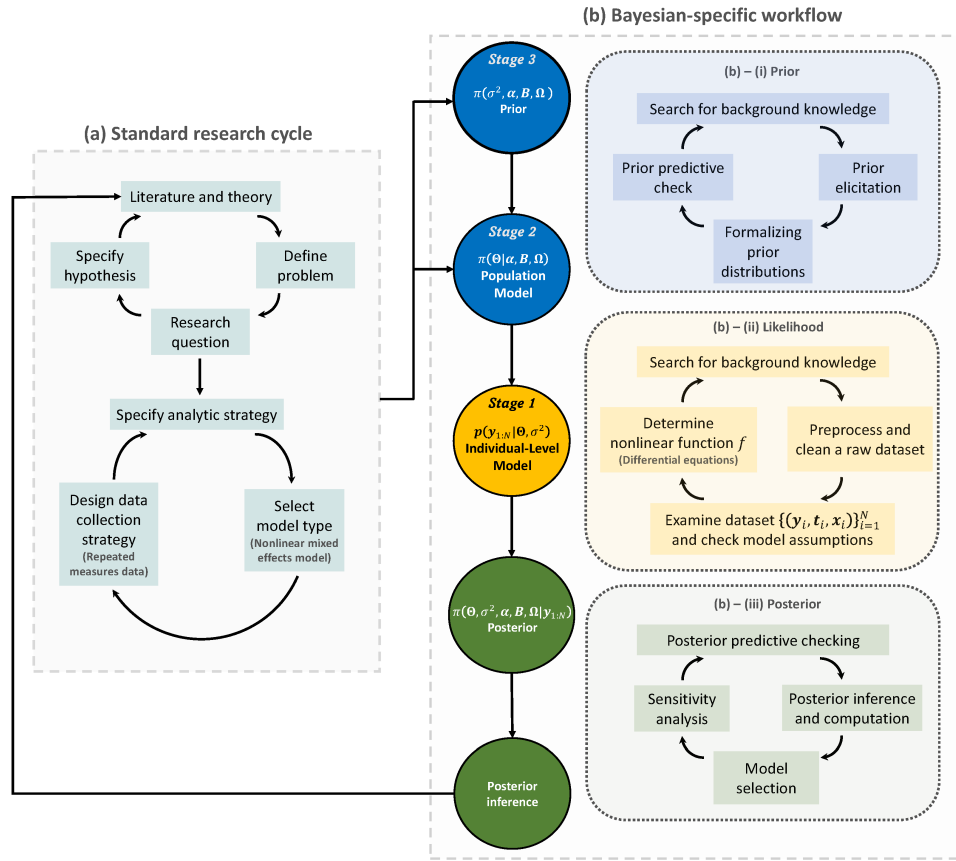


Figure 1. The Bayesian research cycle. A research cycle using Bayesian nonlinear mixed effects model comprises two steps: (a) standard research cycle and (b) Bayesian-specific workflow. Standard research cycle involves literature review, defining a problem and specifying the research question and hypothesis. Bayesian-specific workflow comprises three sub-steps: (b)–(i) formalizing prior distributions based on background knowledge and prior elicitation; (b)–(ii) determining the likelihood function based on a nonlinear function f ; and (b)–(iii) making a posterior inference.

The first step of the Bayesian research cycle is (a) standard research cycle [50][51]. Some early activities at this step involve reviewing literature, defining a problem, and specifying a research question and a hypothesis. After that, researchers specify which analytic strategy would be taken to solve the research question and suggest possible model types, followed by data collection. The data type arising in this process may include a response variable and some covariates that are grouped longitudinally, which then formulates repeated measures data of a population of interest. Furthermore, if there appears to be some nonlinear temporal tendency at each subject, then a possible model type for the analysis is a nonlinear mixed effects model [15][52].

The second step of the Bayesian research cycle is (b) Bayesian-specific workflow. Logically, the first thing to do at this step is to determine prior distributions (see Step (b)–(i) in **Figure 1**). The selection of priors is often viewed as one of the most crucial choices that a researcher makes when implementing a Bayesian model, as it can have a substantial impact on the final results [53]. As exemplified earlier in the context of Bayesian medical device trials, using a prior in small sample studies may improve the estimation accuracy, but an unthoughtful choice of priors would lead to a significant bias in estimation. Prior elicitation effort would require Bayesian expertise to formulate domain expert's knowledge in a probabilistic form [54]. Strategies for prior elicitation include asking domain experts to provide suitable values for the hyperparameters of the prior [55][56]. After prior is specified, one can check the appropriateness of the priors through prior predictive checking process [57]. For almost all practical problems, prior distribution of Bayesian nonlinear mixed effect models can be hierarchically represented as follow: (1) a prior for the parameters used in likelihood, often called 'population-level model' in the literature of mixed effects modeling; and (2) a prior for the parameters used in the population-level model and for the parameters describing the residual errors used in likelihood. It is important to note that the former type of prior distribution (that is, (1)) is also a requirement to implement frequentist approaches for the nonlinear mixed effects model, as a name of 'distribution for random effects'.

The second task is to determine the likelihood function (see Step (b)–(ii) in the panel). At this time, the raw dataset collected in (a) standard research cycle should be cleaned and preprocessed. Before embarking on more serious statistical modeling, it is a common practice to get some insight about the research question via exploratory data analysis and have a discussion with domain experts such as clinical pharmacologists, clinicians, physicians, engineers, etc. To some extent, eventually, all these efforts are to determine a nonlinear function that best describes the temporal profiles of all subjects. This nonlinear function is a *known* function because it should be specified by researchers. In other words, the branch of the nonlinear mixed effects models belongs to parametric statistics. However, one technical challenge is that, in many problems, such a nonlinear function is represented as a solution of a differential equation system [58][59], and therefore there is no guarantee that people can conveniently work with a closed-form expression of the nonlinear function. For example, if researchers wish to work with nonlinear differential equations [60][61], then some approximation via differential equation solver [62][63] may be needed to calculate the nonlinear function. As such, most software packages dedicated to implementing a nonlinear mixed effect model, or, more generally, a Bayesian hierarchical model, are equipped with several built-in differential equation solvers [17][37][64]. For instance, visit the website (https://mc-stan.org/docs/2_29/stan-users-guide/ode-solver.html, accessed on 20 February 2022) to see some functionality supported in Stan [17].

Finally, the likelihood is combined with the prior to form the posterior distribution (see Step (b)–(iii) in the panel). Given the important roles that the prior and the likelihood have in determining the posterior, this step must be conducted with care. The implementational challenge at this step is to construct an efficient MCMC sampling algorithm. The basic idea behind MCMC here is the construction of a sampler that simulates a Markov chain that is converging to the posterior distribution. One can use software packages if prior distributions to be implemented in Bayesian models exist in the list of prior options available in the packages.

References

1. Sterba, S.K. Fitting nonlinear latent growth curve models with individually varying time points. *Struct. Equ. Model. Multidiscip. J.* 2014, 21, 630–647.
2. McArdle, J.J. Latent variable growth within behavior genetic models. *Behav. Genet.* 1986, 16, 163–200.
3. Cook, N.R.; Ware, J.H. Design and analysis methods for longitudinal research. *Annu. Rev. Public Health* 1983, 4, 1–23.
4. Mehta, P.D.; West, S.G. Putting the individual back into individual growth curves. *Psychol. Methods* 2000, 5, 23.
5. Zeger, S.L.; Liang, K.Y. An overview of methods for the analysis of longitudinal data. *Stat. Med.* 1992, 11, 1825–1839.
6. Diggle, P.; Diggle, P.J.; Heagerty, P.; Liang, K.Y.; Zeger, S. *Analysis of Longitudinal Data*; Oxford University Press: Oxford, UK, 2002.
7. Demidenko, E. *Mixed Models: Theory and Applications with R*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
8. Snijders, T.A.; Bosker, R.J. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*; Sage: Los Angeles, CA, USA, 2011.
9. Goldstein, H. *Multilevel Statistical Models*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 922.
10. Raudenbush, S.W.; Bryk, A.S. *Hierarchical Linear Models: Applications and Data Analysis Methods*; Sage: Thousand Oaks, CA, USA, 2002; Volume 1.
11. Efron, B. The future of indirect evidence. *Stat. Sci. A Rev. J. Inst. Math. Stat.* 2010, 25, 145.
12. Sheiner, L.B.; Rosenberg, B.; Melmon, K.L. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput. Biomed. Res.* 1972, 5, 441–459.
13. Lindstrom, M.J.; Bates, D.M. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990, 46, 673–687.
14. Davidian, M.; Giltinan, D.M. Nonlinear models for repeated measurement data: An overview and update. *J. Agric. Biol. Environ. Stat.* 2003, 8, 387–419.
15. Davidian, M.; Giltinan, D.M. *Nonlinear Models for Repeated Measurement Data*; Routledge: New York, NY, USA, 1995.
16. Beal, S. The NONMEM System. 1980. Available online: <https://iconplc.com/innovation/nonmem/> (accessed on 20 February 2022).
17. Stan Development Team. *RStan: The R Interface to Stan*. R Package Version 2.21.3. 2021. Available online: <https://mc-stan.org/rstan/> (accessed on 20 February 2022).

18. Fidler, M.; Wilkins, J.J.; Hooijmaijers, R.; Post, T.M.; Schoemaker, R.; Trame, M.N.; Xiong, Y.; Wang, W. Nonlinear mixed-effects model development and simulation using nlmixr and related R open-source packages. *CPT Pharmacometr. Syst. Pharmacol.* 2019, 8, 621–633.
19. Wang, W.; Hallow, K.; James, D. A tutorial on RxODE: Simulating differential equation pharmacometric models in R. *CPT Pharmacometr. Syst. Pharmacol.* 2016, 5, 3–10.
20. Stegmann, G.; Jacobucci, R.; Harring, J.R.; Grimm, K.J. Nonlinear mixed-effects modeling programs in R. *Struct. Equ. Model. Multidiscip. J.* 2018, 25, 160–165.
21. Vonesh, E.; Chinchilli, V.M. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*; CRC Press: Boca Raton, FL, USA, 1996.
22. Lee, S.Y. *Structural Equation Modeling: A Bayesian Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
23. Dellaportas, P.; Smith, A.F. Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *J. R. Stat. Soc. Ser. C* 1993, 42, 443–459.
24. Bush, C.A.; MacEachern, S.N. A semiparametric Bayesian model for randomised block designs. *Biometrika* 1996, 83, 275–285.
25. Zeger, S.L.; Karim, M.R. Generalized linear models with random effects; a Gibbs sampling approach. *J. Am. Stat. Assoc.* 1991, 86, 79–86.
26. Brooks, S.P. Bayesian computation: A statistical revolution. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 2003, 361, 2681–2697.
27. Bennett, J.; Wakefield, J. A comparison of a Bayesian population method with two methods as implemented in commercially available software. *J. Pharmacokinet. Biopharm.* 1996, 24, 403–432.
28. Wakefield, J. The Bayesian analysis of population pharmacokinetic models. *J. Am. Stat. Assoc.* 1996, 91, 62–75.
29. Gelman, A.; Bois, F.; Jiang, J. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Am. Stat. Assoc.* 1996, 91, 1400–1412.
30. Lee, S.Y.; Lei, B.; Mallick, B. Estimation of COVID-19 spread curves integrating global data and borrowing information. *PLoS ONE* 2020, 15, e0236860.
31. Lee, S.Y.; Mallick, B.K. Bayesian Hierarchical Modeling: Application Towards Production Results in the Eagle Ford Shale of South Texas. *Sankhya B* 2021, 1–43.
32. Hammersley, J. *Monte Carlo Methods*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
33. Green, P.J.; Łatuszyński, K.; Pereyra, M.; Robert, C.P. Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.* 2015, 25, 835–862.
34. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, 20–22 March 2003; Volume 124, pp. 1–10.
35. Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. The BUGS project: Evolution, critique and future directions. *Stat. Med.* 2009, 28, 3049–3067.
36. Chernoff, H. Large-sample theory: Parametric case. *Ann. Math. Stat.* 1956, 27, 1–22.
37. Beal, S.L.; Sheiner, L.B.; Boeckmann, A.; Bauer, R.J. *NONMEM Users Guides*; NONMEM Project Group, University of California: San Francisco, CA, USA, 1992.
38. Lavielle, M. *Monolix User Guide Manual*. 2005. Available online: <https://monolix.lixoft.com/> (accessed on 20 February 2022).
39. Wand, M. Fisher information for generalised linear mixed models. *J. Multivar. Anal.* 2007, 98, 1412–1416.
40. Kang, D.; Bae, K.S.; Houk, B.E.; Savic, R.M.; Karlsson, M.O. Standard error of empirical bayes estimate in NONMEM® VI. *Korean J. Physiol. Pharmacol.* 2012, 16, 97–106.
41. Breslow, N.E.; Clayton, D.G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 1993, 88, 9–25.
42. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; Chapman and Hall/CRC: London, UK, 2004.
43. Smid, S.C.; McNeish, D.; Miočević, M.; van de Schoot, R. Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Struct. Equ. Model. Multidiscip. J.* 2020, 27, 131–161.
44. Rupp, A.A.; Dey, D.K.; Zumbo, B.D. To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Struct. Equ. Model.* 2004, 11, 424–451.

45. Bonangelino, P.; Irony, T.; Liang, S.; Li, X.; Mukhi, V.; Ruan, S.; Xu, Y.; Yang, X.; Wang, C. Bayesian approaches in medical device clinical trials: A discussion with examples in the regulatory setting. *J. Biopharm. Stat.* 2011, 21, 938–953.
46. Campbell, G. Bayesian methods in clinical trials with applications to medical devices. *Commun. Stat. Appl. Methods* 2017, 24, 561–581.
47. Hoff, P.D. *A First Course in Bayesian Statistical Methods*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 580.
48. O'Hagan, A. Bayesian statistics: Principles and benefits. *Frontis* 2004, 3, 31–45.
49. van de Schoot, R.; Depaoli, S.; King, R.; Kramer, B.; Märtens, K.; Tadesse, M.G.; Vannucci, M.; Gelman, A.; Veen, D.; Willemsen, J.; et al. Bayesian statistics and modelling. *Nat. Rev. Methods Prim.* 2021, 1, 1–26.
50. Blaxter, L.; Hughes, C.; Tight, M. *How to Research*; McGraw-Hill Education: New York, NY, USA, 2010.
51. Neuman, W.L. *Understanding Research*; Pearson: New York, NY, USA, 2016.
52. Pinheiro, J.; Bates, D. *Mixed-Effects Models in S and S-PLUS*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
53. Gelman, A.; Simpson, D.; Betancourt, M. The prior can often only be understood in the context of the likelihood. *Entropy* 2017, 19, 555.
54. Garthwaite, P.H.; Kadane, J.B.; O'Hagan, A. Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.* 2005, 100, 680–701.
55. O'Hagan, A.; Buck, C.E.; Daneshkhah, A.; Eiser, J.R.; Garthwaite, P.H.; Jenkinson, D.J.; Oakley, J.E.; Rakow, T. *Uncertain Judgements: Eliciting Experts' Probabilities*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2006.
56. Howard, G.S.; Maxwell, S.E.; Fleming, K.J. The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychol. Methods* 2000, 5, 315.
57. Levy, R. Bayesian data-model fit assessment for structural equation modeling. *Struct. Equ. Model. Multidiscip. J.* 2011, 18, 663–685.
58. Wang, L.; Cao, J.; Ramsay, J.O.; Burger, D.; Laporte, C.; Rockstroh, J.K. Estimating mixed-effects differential equation models. *Stat. Comput.* 2014, 24, 111–121.
59. Botha, I.; Kohn, R.; Drovandi, C. Particle methods for stochastic differential equation mixed effects models. *Bayesian Anal.* 2021, 16, 575–609.
60. Fucik, S.; Kufner, A. *Nonlinear Differential Equations*; Elsevier: Amsterdam, The Netherlands, 2014.
61. Verhulst, F. *Nonlinear Differential Equations and Dynamical Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
62. Cohen, S.D.; Hindmarsh, A.C.; Dubois, P.F. CVODE, a stiff/nonstiff ODE solver in C. *Comput. Phys.* 1996, 10, 138–143.
63. Dormand, J.R.; Prince, P.J. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* 1980, 6, 19–26.
64. Margossian, C.; Gillespie, B. Torsten: A Prototype Model Library for Bayesian PKPD Modeling in Stan User Manual: Version 0.81. Available online: <https://metrumresearchgroup.github.io/Torsten/> (accessed on 20 February 2022).