Web-based Supplementary Materials for:

'Bayesian analysis of non-linear differential equation models with application to a gut microbial ecosystem'

by Daniel John Lawson, Grietje Holtrop and Harry Flint.

This document provides web-based supplementary material (Web Appendix A-F), giving a detailed definition of the model, as well as details of the data and parameter values considered for both the inference from microbiology data and simulated data. It also gives the simulation results for the one-and-two bacteria cases (Web Figure 1-2), and a list of all variables used (Web Table 1).

A. Web Appendix A: Data used for inference

Data taken from (Walker et al., 2005) for the Peptide 0.6% data (Donor 1), and converted into grams. SCFA are converted from Moles and Bacterial mass is assumed to be $1.6 \times 10^{-12} g$ per cell based on the order of magnitude of E. coli bacterial mass (Pelczar et al., 1993). Bacterial groupings are based on Fluorescent In Situ Hybridisation (FISH) groupings: 'Bacteroides' are the Bac303 probe, 'Rrec' is the whole Erec482 probe (of which the true Rrec group is a subset), and 'AcProd' is all other bacteria as determined by Eubacteria (all bacteria, probe Eub358) subtracting the above probes. $\sigma(.)$ refers to the standard deviation of the estimation, and is the measurement error as reported in (Walker et al., 2005) (rounded up in 0.5 mMols).

Experimental Data for pH 5.5:

Time (hr)	Ac.	Bu.	σ(Ac.)	σ(Bu.)
0	2.16	0.352	0.060	0.044
48	1.44	2.464	0.030	0.044
96	0.90	2.376	0.024	0.044
144	0.72	2.376	0.030	0.044

Time (hr)	Bac	Rrec	AcProd	σ(Bac)	σ(RRec)	σ(AcProd)
0	0.0357	0.2074	0.6137	0.0085	0.017	0.085
144	0.6970	1.0285	0.4505	0.0340	0.051	0.085

Experimental Data at pH 6.5:

Time	Ac.	Bu.	σ(Ac.)	σ(Bu.)
0	0.78	1.936	0.18	0.088
24	1.20	1.320	0.03	0.044
72	1.44	0.792	0.06	0.044
120	1.68	0.704	0.12	0.044
168	1.62	0.704	0.03	0.044

Time	Bac	Rrec	AcProd	σ(Bac)	σ(RRec)	σ(AcProd)
0	0.697	1.0285	0.4505	0.034	0.034	0.085
168	3.757	0.2414	0.4896	0.170	0.017	0.085

Table 1. Explanation of all variables in the differential equation model for Gut Bacteria growth.

Index	Description
i	A type of Bacteria.
j	A type of Substrate.
k	A type of SCFA.
Variable	**Description**
t	Time.
$B_i(t)$	Bacterial concentration of type i.
$S_j(t)$	Substrate concentration of type j.
$A_k(t)$	SCFA (acid) concentration of type k.
Parameter	**Description**
F	Flow rate through the fermenter.
E_{ij}^S	Efficiency of usage of substrate j by bacteria i.
E_{ik}^A	Efficiency of usage of SCFA k by bacteria i.
G_{ij}^S	Maximum growth rate of bacteria i on substrate j.
G_{ik}^A	Maximum growth rate of bacteria i on SCFA k.
K_i^S	Michaelis-Menten factor for bacteria i on substrate, giving substrate abundance for which the growth rate is reduced by half from its maximum value.
K_i^A	Michaelis-Menten factor for bacteria i on SCFA.
O_{ik}	Output of SCFA k by bacteria i on unit growth.
R_k	Host absorption rate of SCFA k.
Intermediate	**Description**
$B_{ij}(t)$	Amount of Bacteria i adhered to substrate j.
$U_S(i,j;t)$	Efficient usage rate by bacteria i of substrate j.
$U_A(i,k;t)$	Efficient SCFA usage by bacteria i of SCFA k.
$P_A(i,k;t)$	Production of SCFA k by bacteria i.
$H_A(k;t)$	Absorption by the host of SCFA k.

B. Web Appendix B: Detailed definition of the model

C. Web Appendix C: Mapping Matrices

The mapping matrices X_a are constructed by starting with a $p \times p$ identity matrix and removing rows for parameters not present in a given experiment.

To illustrate our use of mapping matrices, assume that we have two experiments for inference of two different growth rates G_1 and G_2 with a constant Michaelis Menten Factor M. Experiment 1 has $\theta_1 = (M_1, G1)^T$ and experiment 2 has $\theta_2 = (M_2, G2)^T$. Let $\mu = (\bar{M}, G_1, G_2)^T$. Then $X_1 = [1\,0\,0\,;\,0\,1\,0]$ and $X_2 = [1\,0\,0\,;\,0\,0\,1]$ ensure that the appropriate elements of μ are mapped to the θ_a.

In addition, consider

$$\Sigma = \begin{pmatrix} \sigma^2_{MM} & 0 & 0 \\ 0 & \sigma^2_{G1} & 0 \\ 0 & 0 & \sigma^2_{G2} \end{pmatrix},$$

then

$$\Sigma_1 = X_1 \Sigma X_1^T = \begin{pmatrix} \sigma^2_{MM} & 0 \\ 0 & \sigma^2_{G1} \end{pmatrix},$$

and

$$\Sigma_2 = X_2 \Sigma X_2^T = \begin{pmatrix} \sigma^2_{MM} & 0 \\ 0 & \sigma^2_{G2} \end{pmatrix}.$$

D. Web Appendix D: Parameter values

The substrate is a composite of potato starch, amylopectin, xylan, pectin and arabinogalactan. For inference purposes these are combined together into Starch (potato starch and amylopectin) and NSP (xylan, pectin and arabinogalactan). It is assumed that the volume of the vessel is one litre and that the flow rate is equal to one complete replacement per day. Rates are specified in units of grammes per day. The input rate of substrates were 5.6g starch and 1.8g NSP per day.

For two substrate types (starch and NSP), three bacterial types (Acetate Producers, Bacteroides and Roseburia) and two SCFA types (Acetate and Butyrate), the following prior distributions are chosen as expert guesses taken from a wide range of literature. Priors for all parameters are assumed to follow a normal distribution, with mean μ and standard deviation σ specified using a normal distribution as $\mu \pm \sigma$.

Flow rates in the experimental conditions from Section 5 are 1 complete replacement per day. The input rate of substrates were $5.6g$ starch and $1.8g$ NSP per day. Uncertain parameters are specified as $\beta_{ab} \pm \tau_{ad}$. Bacterial rates are specified as (Bac, Rrec, AcProd) triplets. Maximum growth rate per hour on starch at pH 6.5: $(0.65, 0.4, 0.5) \pm (0.05, 0.05, 0.05)$, and on NSP: $(0.01, 0.015, 0.05) \pm (0.01, 0.01, 0.01)$. At pH 5.5 these are $(0.4, 0.4, 0.4) \pm (0.05, 0.05, 0.05)$ on starch and $(0.01, 0.015, 0.05) \pm (0.01, 0.01, 0.01)$ on NSP. Net Acetate production is $(2.5, 0, 2.5) \pm (0.2, 0, 0.2)$, and net Butyrate production is $(0, 1.5, 0) \pm (0, 0.2, 0)$. Only Rrec uses SCFA (Acetate) at a rate of $1/3 \pm 0.1$ for growth (with unknown efficiency). Michalis Menten Factors for substrate are specified by their inverse which takes values $(80, 80, 80) \pm (40, 40, 40)$ and the Michaelis-Menten factor for Rrec uptake of Acetate is also 80 ± 40 on its inverse. This is because the parameters are effectively unknown but a proper prior is required for satisfactory convergence.

The unknown parameters form part of a non-linear system of differential equations, and this complicates the inference problem. Therefore, to keep inference manageable, it is assumed that the variation within and variation between experiments is known. This was obtained from data from a second donor.

Initial conditions for the MCMC chain were set to the prior Mean where appropriate. Other values were: inefficiency of bacteria on starch: $(2, 4, 2)$ and on NSP: $(4, 8, 4)$. The Michaelis-Menten values for substrates were set to $(0.02, 0.017, 0.02)$ and the value for Rrec utilisation of Acetate was 5 at inefficiency 4/3. All other values are zero. Although these values give a reasonable approximation of the data, they are still assigned a likelihood numerically rounded to zero. Therefore MCMC was performed on the data with inflated prior standard deviations (by a factor of 5) for an initial 10000 MCMC steps. This yielded starting values suitable for the MCMC to begin. The first 500000 iterations were discarded as burn-in time, which was observed by eye as suitable since the likelihood and fit to the prior had stopped increasing by this point. The analysis was run twice more with all starting value perturbed by 10% to help detect MCMC convergence. The three chains were each run for 2 weeks yielding runs of length $(1043999, 1248671, 1216894)$ respectively; for calculation of $\sqrt{\hat{R}}$ and the potential scale reduction factor only the first 1000000 iterations are considered but all iterations are used in the final analysis.

The hyper-parameter for the covariance matrix Σ between parameters between experiments is $\Sigma_{ab} = 0$ except on diagonal elements for fundamental parameters, which are specified non-zero constants. These are specified to capture the level of variation (including uncertainty) in previous data. Therefore a Σ_{dd} value specifies the covariance between parameter d in all experiments in which d appears. For maximum growth rates on both starch and NSP the diagonal Σ elements take the values $(0.4, 0, 0.4)$ for bacterial triplets as above (note that Rrec do not vary in this parameter). Similarly for inefficiencies the elements of Σ are $(0.4, 0.05, 0.4)$. These values are large because the Bac group has been observed to be very pH sensitive in the region of interest (and so the growth rate may vary a lot) and the AcProd group is poorly defined. Other values are not used because they appear exactly in all experiments.

To test sensitivity to the choice of Σ, maximum aposteriori (MAP) estimation of the parameters was performed at $\Sigma^* = (\Sigma/2, \Sigma, 2\Sigma)$. Parameter estimates did not vary by greater than 1% from the results presented at $\Sigma^* = \Sigma$, implying that the MAP estimate is very robust to choices of Σ. The posterior does however become more disperse as Σ is increased, although due to computational constraints in running additional MCMC chains we have not examined the magnitude of this effect.

Impossible parameter sets are determined as follows. If in experiment a there exists i, j s.t. the inefficiency $E^S_{ij} > MP_{ij}$ then $V(\theta_a) = 0$, else $V(\theta_a) = 1$, where the maximum production possible MP_{ij} is defined by

$$MP_{ij} = \max_j \left\{ 1 + \sum_k P_{ik} + \sum_{k'} G^A_{ik'} \left(1 + \sum_k P_{ik} - E^A_{ik} \right) \right\}. \quad (1)$$

The mapping matrices X_a were constructed as described in Web Appendix C.

E. Web Appendix E: Model implementation

Due to the inherent non-linearities in the model acceptance rates are low (or step sizes small) if each parameter is treated independently, and block structure is difficult to identify

in advance.

For computational purposes, we define the vector of all experimental parameters $\vartheta = \{\theta_a\}$ which contains P parameters in total. A sample of these parameters with size $s \sim \text{Binomial}(P, p_m)$ is chosen to form a block vector B (also of length P) taking the value 1 for the s members of the block and 0 otherwise. Here, p_m is the average proportion of parameters to be chosen (with empty samples rejected). Therefore p_m controls the block size distribution between single parameter updates (for $p_m \to 0$) and full blocking (for $p_m = 1$).

The MCMC proposal is constructed from the sample parameter block according to

$$p(\vartheta'_e|\vartheta_e) = N(\vartheta'_e|\vartheta_e, B_e\delta_e^2), \tag{2}$$

which forms the proposal distribution. Here δ is the vector of step sizes defined by the modeller and the index $e = 1 \dots P$.

For the results on experimental data, the average proportion of parameters to be updated was $p_m = 0.1$ (i.e. on average 2.9 parameters per MCMC update).

For all studies, $\delta = 0.01$ with the following exceptions. In Study 1, $\delta = 0.05$ for the Michaelis-Menten factor K^S. In Study 2, $\delta = 0.05$ for growth rate on NSP G_2^S, and $\delta = 0.005$ for growth rate on starch G_1^S. For the real data study $\delta = 0.005$ for all Rrec growth rates $G_{2\cdot}^S$, $\delta = 0.05$ for Rrec SCFA inefficiency $G_{2\cdot}^S$, and $\delta = 0.1$ for all Michaelis-Menten factors for substrate K^S and SCFA K^A.

F. Web Appendix F: Simulation Study parameters

The simulated data from Section 4 was performed using a different set of parameters to those for the inference from real data. Growth rate per hour of 'Bac' was 0.5 at pH 6.5 and 0.05 at at the lower pH 5.5. Growth of 'Rrec' was 0.25 at all pH. All Michaelis-Menten factors were set to 1. Bacteroides substrate inefficiency was 3, Rrec substrate inefficiency was 10 and its SCFA inefficiency was 3. Rrec Acetate utilisation was 2, Bac Acetate production was 1 and Rrec Butyrate production was 2. The simulated flowrate was 2.4 turnovers per day, with $8g$ of starch being input per turnover. The mapping matrices X_a were constructed as described in Web Appendix C.

F.1. Effect of Σ

The results from Section 4.1 of the paper were obtained from assuming perfect correlation in the Michaelis Menten factor, inefficiency of uptake and SCFA production. Here we consider the effect of instead assuming a hierarchical correlation $\Sigma_{dd} = diag(s)$ for a range of values $s = (0.05, 0.20, 0.40, 0.60, 0.80, 0.95, 2.00, 5.00)$ in addition to perfect correlation ($s = 0$ in the figure). All values of s are centered on the correct value, since the data were generated under the correct model; however there are identifiability problems for large Σ due to the use of a uniform prior. The minimum sample size for $c = 0.8$ and larger was less than 50 and so these values will tend to be an underestimate.

References

Pelczar, M. J., E. C. S. Chan, and N. R. Krieg (1993). Microbiology: concepts and applications. New York: McGraw-Hill.

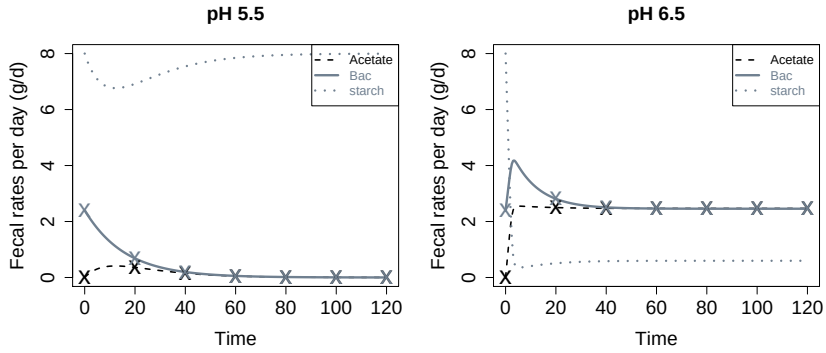

Fig. 1. Simulated time series for Study 1 using a single bacteria representing reasonable parameter values for the Bacteroides strain (solid grey line) and a single output SCFA of Acetate produced by Bacteroides (dotted black line), with the starch substrate levels (dotted grey line). Data points used for inference are shown as crosses. Experiment 1 (left) is the acidic pH 5.5 environment for which the growth rate is a factor of 10 slower than in experiment 2 at pH 6.5 (right). Parameters are taken from Web Appendix D.

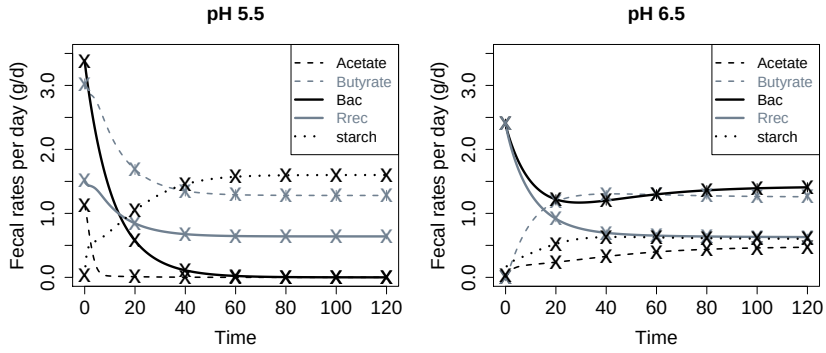

Fig. 2. Simulated time series for Study 2 with two experiments using two bacterial strains, Bac (solid black line) and Rrec (solid grey line). Also shown are the Acetate levels (black dashed line), Butyrate levels (grey dashed line) and starch substrate (black dotted line). Simulated data points used for inference are shown as crosses (with a 95% confidence interval given by a normally distributed error with standard deviation of 5% of the value). Experiment 3 (left) has high acidity (pH 5.5) so the growth rate for acid-vulnerable Bac is reduced by a factor of 10 compared to experiment 4 (right) at neutral acidity (pH 6.5).

Walker, A. W., S. H. Duncan, E. C. M. Leitch, M. W. Child, and H. J. Flint (2005). pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. Applied and Environmental Microbiology 71, 3692–3700.

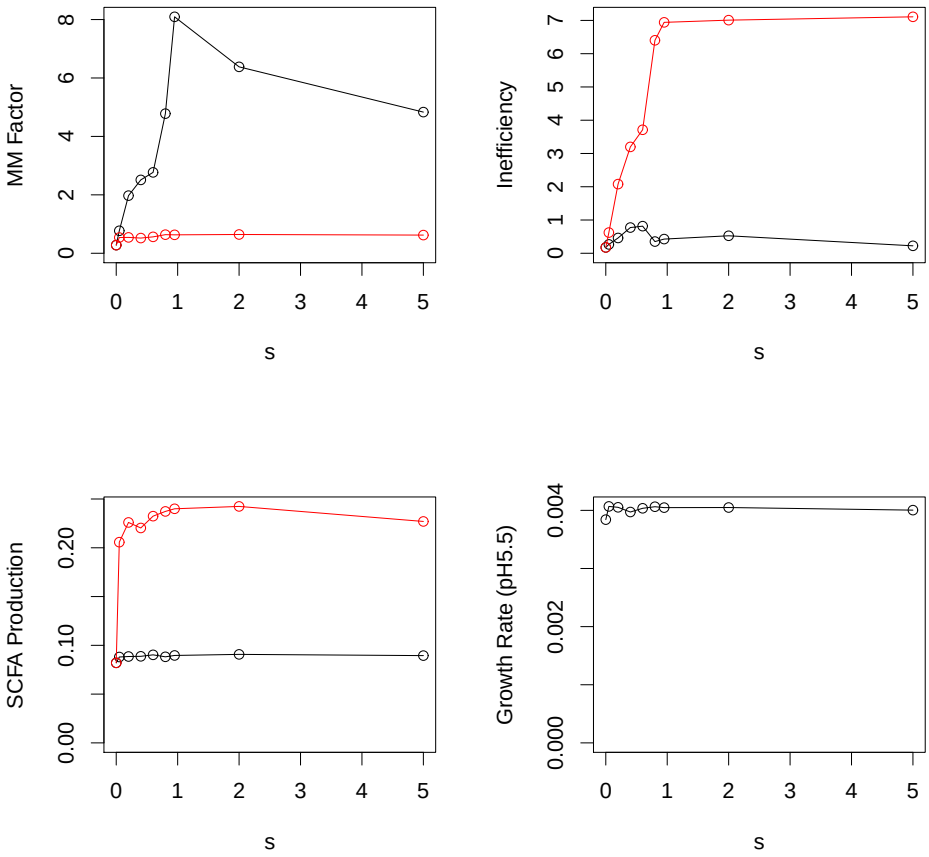

Fig. 3. Range of 95% confidence interval for parameters depending on the correlation between experiments. EXperiment 1 is shown in black and Experiment 2 is shown in red. The value $s = 0$ corresponds to the case in the main text.