A mathematical model of cocoa bean fermentation

Mauricio Moreno-Zambrano, Sergio Grimbs, Matthias S. Ullrich and Marc-Thorsten Hütt

Department of Life Sciences and Chemistry, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany

Cocoa bean fermentation relies on the sequential activation of several microbial populations, triggering a temporal pattern of biochemical transformations. Understanding this complex process is of tremendous importance as it is known to form the precursors of the resulting chocolate’s flavour and taste. At the same time, cocoa bean fermentation is one of the least controlled processes in the food industry. Here, a quantitative model of cocoa bean fermentation is constructed based on available microbiological and biochemical knowledge. The model is formulated as a system of coupled ordinary differential equations with two distinct types of state variables: (i) metabolite concentrations of glucose, fructose, ethanol, lactic acid and acetic acid and (ii) population sizes of yeast, lactic acid bacteria and acetic acid bacteria. We demonstrate that the model can quantitatively describe existing fermentation time series and that the estimated parameters, obtained by a Bayesian framework, can be used to extract and interpret differences in environmental conditions. The proposed model is a valuable tool towards a mechanistic understanding of this complex biochemical process, and can serve as a starting point for hypothesis testing of new systemic adjustments. In addition to providing the first quantitative mathematical model of cocoa bean fermentation, the purpose of our investigation is to show how differences in estimated parameter values for two experiments allow us to deduce differences in experimental conditions.

1. Introduction

The fermentation of cocoa beans is recognized as a key step in cocoa processing in terms of the development of chocolate’s flavour and aroma [1,2]. It occurs mainly in the pulp, i.e. a white mucilaginous mass that surrounds the bean, where three major microbial groups drive mostly the whole process whose main activity occurs in a consecutively way (figure 1a), being metabolically dominated in earlier stages by yeast (Y) and...
subsequently surpassed by lactic acid bacteria (LAB), and after the decline of these two first groups, acetic acid bacteria (AAB) take over [4–7]. This so-called three-phase process, depending on the region and local farm practices, is expected to happen within a time frame of 2–10 days [2,3,8–10].

As a result of the fermentation, a series of biochemical reactions are triggered in the raw material, the qualitative characteristics of which have been exhaustively described in terms of the microbial groups involved and the associated metabolic alterations [9,11,12].

Despite the fact that this process is of high industrial relevance, there are hardly any attempts to construct a mathematical model of cocoa bean fermentation. So far, the existing modelling attempts are either focusing on specific post-fermentation steps such as drying kinetics [13,14], restricted to the sequential interaction of microbial communities using metabolic flux analysis [15,16] or kinetic approaches that ‘cannot explain the dynamics in microbial population’ [17].

The reasons for this are manifold, among them, the lacking of control over the fermentation process itself as well as the systemic complexity in terms of involved microbial communities. On the one hand, cocoa bean fermentation is conducted in a spontaneous way unlike the majority of other food fermentation processes [18] with a huge diversity of techniques and devices, e.g. heaps, boxes, baskets, trays, sacks and platforms [2,8,18]. Owing to the lack of control, it is difficult to identify the crucial parameters and key variables required for the formulation of an appropriate model.

On the other hand, in contrast to other relevant industrial fermentation processes such as those of beer and wine, the fermentation in cocoa involves microbial community dynamics of three major microbial groups, i.e. Y, LAB and AAB, which are, in turn, represented by several different strains [3,4,6,7]. Hence, the complexity is precisely one of the biggest challenges to overcome since growth modelling of microorganisms has been classically applied under much more controlled conditions that involve single strain cultures where mortality phenomena have been scarcely considered [19]. Consequently, this needs to be taken into account for an approximation of the cocoa bean fermentation process.

Cocoa bean fermentation is a prototypical situation for the application of modelling using coupled nonlinear ordinary differential equations (ODE): The initial situation displays a rich diversity with a multitude of influencing factors and the result of the dynamic process, the fermented cocoa bean, is of high relevance for the subsequent industrial processing steps and for the quality of the final product, chocolate.

Here, we present a one-compartment model for the cocoa bean fermentation process using the mathematical concepts of the Monod [20] and Contois [21] equations, assuming that single strain kinetic modelling techniques can describe the growth of mixtures of microbial species belonging to different microbial groups in the same environment. Moreover, microbial death processes are handled by the use of the Chick–Watson mortality law [22].

While conceptionally the modelling approach presented here is rather in the tradition of theoretical biology, the way to analyse and apply the model differs from e.g. a traditional linear stability analysis, as the purpose of the model is predominantly to describe transients in a batch culture [23], rather than asymptotic states as would be expected in continuous cultures.

With the model constructed along these lines, we were able to describe three datasets corresponding to two different cocoa-producing countries where two different fermentation methodologies were implemented. In that way, our model also can interpret differences in the experimental set-up of the

![Figure 1. Typical time series for community dynamics and metabolite kinetics during cocoa bean fermentation. (a) Community dynamics; yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). (b) Metabolite kinetics; glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac). Both counts and concentration are shown in arbitrary units. (After [3]; use permitted under the Creative Commons Attribution License CC-BY 4.0 number 4354810766457).](image_url)
two trials conducted under the same methodology, in terms of significant changes in the estimated parameters. This approach serves as a source of elucidation of possible hypotheses on how these parameters are affected by slight changes within a particular region where the fermentation took place.

2. Material and methods

2.1. Experimental data

The experimental data used in this study were reported in Camu et al. [4] and Papalexandratou et al. [24]. In both instances, the predominant cocoa hybrids harnessed by the chocolate industry, Criollo and Forastero, were used as the source of raw material. In the study of Camu et al. [4], the beans were fermented by the heaps method, while for Papalexandratou et al. [24], wooden boxes were used as fermenting devices. The data of Camu et al. [4] were collected in Ghana from seven trials in two field experiments and data of one representative trial were published. The data include measurements of microbial counts of Y, LAB, AAB and total aerobic bacteria. Metabolite time series measured both in pulp and bean are available for glucose (Glc), fructose (Fru), sucrose, lactic acid (LA), acetic acid (Ac), ethanol (EtOH), mannitol, citric acid and succinic acid.

The data reported by Papalexandratou et al. [24] were collected in Brazil from two trials in two field experiments, of which both trials were published as ‘box 1’ and ‘box 2’. The conditions in which both trials were conducted differed slightly. On the one hand, the fermenting mass of box 1 was placed under a metal roof to protect it from weather conditions. On the other hand, the fermentation for box 2 was carried out in a fermentary room. The data include measurements of microbial counts of Y, LAB, AAB and total aerobic bacteria. Metabolite time series measured both in pulp and bean are available for Glc, Fru, sucrose, LA, Ac, EtOH, mannitol and gluconic acid.

In both collections, the fermentation trials took place in a time frame of 6 days with measurements performed at 17 time points for Camu et al. [4] and 14 time points for Papalexandratou et al. [24]. Abiotic factors, i.e. temperature and pH, were measured as well. Cell counts were done by means of malt extract agar, Man–Rogosa–Sharpe agar and deoxycholate-mannitol-sorbitol agar for Y, LAB and AAB, respectively, from both data sources. As metabolite time series, we considered Glc, Fru, EtOH, LA and Ac.

2.2. Microbial count units transformation

One of the most common forms of quantifying microbial growth is the count of colony forming units (CFU), specially when dealing with mixtures of microorganisms as the microbial successions are reported in the original works of Camu et al. [4] and Papalexandratou et al. [24]. In this sense, the vast majority of studies involving single-strained microbial growth are reported in terms of dry biomass as well as their dependent constants, i.e. maximum growth rates, mortality rates and yield coefficients.

In order to obtain comparable estimates to those available for species of these microbial groups in the literature, a conversion from CFU to dry biomass units was conducted based on available knowledge as well as geometric deductions for the microbial group involved in the process.

For species within the microbial group of Y, we used the conversion factor that one CFU of \( \text{S. cerevisiae} \) is equivalent to 15 picograms (pg), as assumed by Schwabe & Bruggeman [25]. For LAB and AAB, since such conversion factors have not been reported yet, values were inferred by taking into account a geometric approximation based on the usual dimensions of the cells belonging to the genus of \( \text{Acetobacter} \) and to the species of \( \text{Lactobacillus plantarum} \), respectively, according to the \textit{Bergey's Manual of Systematic Bacteriology} [26,27] and assuming their shape given by a spherocylinder. Thus, using as reference a density value derived from the dry weight of a cell of \( \text{E. coli} \) of 0.28 pg [28] per micro cubic metre (\( \mu \text{m}^3 \)) [29], the conversion factor between CFU to dry biomass of LAB and AAB were determined as 1.25 and 0.28 pg CFU\(^{-1}\), respectively (see electronic supplementary material, S1).

2.3. Model development

2.3.1. Biochemical background

The fermentation of cocoa beans has been described in detail with respect to its microbial dynamics and metabolite kinetics in both pulp and bean [3,4,6–8]. From such descriptions, the fermentation process in
Cocoa can be understood as an overlapping succession of microbial activities that mostly occur in the pulp, where three core processes are easily identifiable. These are the conversion of Glc and Fru into EtOH by Y, Glc into LA by LAB and EtOH into Ac by AAB (figure 1). Further processes such as the conversions of Glc into Ac by LAB and LA into Ac by AAB, have also been described [30]. The interpretation of these processes in a network diagram covering the pulp only, is shown in figure 2, where microbial growth rate is taken into account represented as the uptake of the respective substrates as well as the mortality rates for Y, LAB and AAB.

In our model, we consider a simultaneous growth of the three major microbial groups. The sequential dominance in the process is then emerging from the availability pattern of their respective substrates without taking into account abiotic factors.

2.3.2. Mathematical representation

The process of cocoa bean fermentation can be seen as a batch process that can involve the usual phases of microbial growth, i.e. lag, exponential, stationary and death phase. However, it is known that microbial growth occurring in natural environments might show different patterns [31]. In this sense, the two collections of experimental data on microbial successions expressed as the log CFU showed basically phases that resembled the exponential and death phases without noticeable stationary or lag phases. From the mathematical perspective, such phenomena can be expressed as an ODE for each of the state variables involved in such a way that the growth of microorganisms is dependent on the availability of their respective substrates, together with mortality equations to capture the inherent decay of the populations along time.

The two major effects on population size considered here, exponential growth and death phase, were modelled by different approaches. On the one hand, we use the classical Monod [20] and Contois [21] equations to describe the growth of groups of microorganisms belonging to a same microbial group (namely, Y, LAB and AAB), instead of the common use of these terms for single strain cultures. Accordingly, the growth rates of Y, \( v_1 \) on Glc and \( v_2 \) on Fru, of LAB, \( v_3 \), and of AAB on ethanol \( v_4 \) and acetic acid \( v_5 \) are represented as straight dashed arrows. The mortality rates of yeast \( v_6 \), lactic acid bacteria \( v_7 \) and acetic acid bacteria \( v_8 \) are represented as zigzag dashed arrows. Straight dashed arrows pointing from products to mortality rates represent product influence on mortality rates. Solid straight arrows show the direction in which the conversion of metabolites occurs.

![Network diagram of the cocoa bean fermentation model. Microbial groups: yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB) are represented as circles. Metabolites: glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac) are represented as squares. The growth rates of yeast on glucose (\( v_1 \)) and fructose (\( v_2 \)), of lactic acid bacteria (\( v_3 \)), and of acetic acid bacteria on ethanol (\( v_4 \)) and acetic acid (\( v_5 \)) are represented as straight dashed arrows. The mortality rates of yeast (\( v_6 \)), lactic acid bacteria (\( v_7 \)) and acetic acid bacteria (\( v_8 \)) are represented as zigzag dashed arrows. Straight dashed arrows pointing from products to mortality rates represent product influence on mortality rates. Solid straight arrows show the direction in which the conversion of metabolites occurs.](rsos.royalsocietypublishing.org R. Soc. open sci. 5: 180964)
of their corresponding metabolite products upon themselves, as shown in Table 1. Together, all equations in Table 1 comprise 13 parameters: (i) five maximum specific growth rates, (ii) five substrate saturation constants, and (iii) three mortality rate constants.

From the set of growth and mortality rate equations defined in Table 1, a system of ODEs can be established in order to mathematically express the given network considering the necessary eleven yield coefficients to take into account the amounts of biomass that can be obtained from substrate as well as the amounts of produced metabolites, as shown in the system of ODEs in equations (2.1) to (2.8) that represent Glc, Fru, EtOH, LA, Ac, Y, LAB and AAB, respectively. A complete interpretation of the 24 estimated parameters is given in Table 2.

Table 1. Growth and mortality rate equations for the cocoa bean fermentation process. Microbial groups are represented as yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). Metabolites are represented as glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac). Biomass and concentration of metabolites are both shown within square brackets [ ]. Maximum specific growth rates $\mu_{\text{max}}$ are shown of the form $\mu_{\text{max},i}$, where $i$ can be either Y, LAB and AAB, and $n$ refers to whether $\mu$ corresponds to the maximum specific growth of Y on either Glc or Fru, or AAB on either EtOH or LA. Substrate saturation constants for the growth of Y, LAB and AAB are shown of the form $K_{j,m}$, where $j$ can be either Y or LAB and $m$ can be either Glc, Fru, EtOH and LA. Constant mortality rates are shown of the form $k_i$, where $i$ can be either Y, LAB or AAB.

| growth rate equation | mortality rate equation |
|----------------------|------------------------|
| $v_1 = \frac{\mu_{\text{max},Y}}{[\text{Glc}]}$ | $v_6 = k_Y [\text{EtOH}]$ |
| $v_2 = \frac{\mu_{\text{max},Y}}{[\text{Fru}]}$ | $v_7 = k_{LAB} [\text{LAB}]$ |
| $v_3 = \frac{\mu_{\text{max},LAB}}{[\text{Glc}]}$ | $v_8 = k_{AAB} [\text{Ac}]$ |
| $v_4 = \frac{\mu_{\text{max},E_{\text{EtOH}}}}{[\text{EtOH}]}$ | $v_9 = k_{LAB} [\text{LAB}]$ |
| $v_5 = \frac{\mu_{\text{max},A_{\text{LA}}}}{[\text{LA}]}$ | |

The proposed model (as described in equations (2.1)–(2.8)) relies on three simple general assumptions: (i) relationships between Y and AAB, as well as of LAB and AAB, are of a pure commensalistic nature since there is no competition between them for any substrate, i.e. Glc and Fru,
and there is no direct effect upon the growth either of Y or LAB by the uptake of its main products, LA and EtOH, by AAB, respectively [33]; (ii) relationship between Y and LAB is a resource-type competition because both microbial groups share Glc as a main limiting substrate and they do not excrete metabolites affecting each other’s growth [34,35]; and (iii) no impact of chemical and physical effects such as temperature and pH on the set of kinetic parameters.

2.4. Parameter estimation

2.4.1. Bayesian framework

In order to estimate the 24 parameters of the proposed model, a Bayesian framework was considered by fitting it to the experimental data. In this way, these parameters can be composed in a vector \( \mathbf{u} \) of the form
\[
\mathbf{u} = [u_1, u_2, \ldots, u_k].
\]
If it is assumed that \( \mathbf{u} \) has given rise to the data \( \mathcal{D} \), the problem can be solved by inferring the posterior probability of \( \mathbf{u} \) given \( \mathcal{D} \),
\[
P(\mathbf{u} \mid \mathcal{D}) \propto \prod_{i=1}^{T} \prod_{j=1}^{N} P(D_{ij} \mid \theta)P(\theta),
\]
where \( D_{ij} \) is the data point from time series \( i \) measured at time \( j \).

### Table 2. Parameters of the cocoa bean fermentation model and their interpretation. Microbial groups: yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). Metabolites: glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac).

| parameter | unit | interpretation |
|-----------|------|----------------|
| \( m_{YG_{max}} \) | h\(^{-1}\) | maximum specific growth rate of Y on Glc |
| \( m_{YF_{max}} \) | h\(^{-1}\) | maximum specific growth rate of Y on Fru |
| \( m_{LAB_{max}} \) | h\(^{-1}\) | maximum specific growth rate of LAB on Glc |
| \( m_{AAB_{max}} \) | h\(^{-1}\) | maximum specific growth rate of AAB on EtOH |
| \( m_{AAB_{LA}} \) | h\(^{-1}\) | maximum specific growth rate of AAB on LA |
| \( K_{YG} \) | mg(Glc) g(pulp)\(^{-1}\) | substrate saturation constant of Y growth on Glc |
| \( K_{YF} \) | mg(Fru) g(pulp)\(^{-1}\) | substrate saturation constant of Y growth on Fru |
| \( K_{LAB} \) | mg(Glc) g(pulp)\(^{-1}\) | substrate saturation constant of LAB growth on Glc |
| \( K_{AAB} \) | mg(EtOH) g(pulp)\(^{-1}\) | substrate saturation constant of AAB growth on EtOH |
| \( K_{LA} \) | mg(LA) g(pulp)\(^{-1}\) | substrate saturation constant of AAB growth on LA |
| \( k_{Y} \) | h\(^{-1}\) | mortality rate constant of Y |
| \( k_{LAB} \) | h\(^{-1}\) | mortality rate constant of LAB |
| \( k_{AAB} \) | h\(^{-1}\) | mortality rate constant of AAB |
| \( Y_{G_{max}} \) | Y mg(Glc) mg(Y)\(^{-1}\) | Y-to-Glc yield coefficient |
| \( Y_{LAB} \) | Y mg(Glc) mg(LAB)\(^{-1}\) | Y-to-LAB yield coefficient |
| \( Y_{F_{max}} \) | Y mg(Fru) mg(Y)\(^{-1}\) | Y-to-Fru yield coefficient |
| \( Y_{E_{max}} \) | Y mg(EtOH) mg(Y)\(^{-1}\) | Y-to-EtOH from Glc yield coefficient |
| \( Y_{E_{max}} \) | Y mg(EtOH) mg(Y)\(^{-1}\) | Y-to-EtOH from Fru yield coefficient |
| \( Y_{E_{max}} \) | Y mg(EtOH) mg(AAB)\(^{-1}\) | Y-to-EtOH from LAB yield coefficient |
| \( Y_{LA} \) | Y mg(LA) mg(LAB)\(^{-1}\) | Y-to-LA yield coefficient |
| \( Y_{LA} \) | Y mg(LA) mg(AAB)\(^{-1}\) | Y-to-LA yield coefficient |
| \( Y_{A_{max}} \) | Y mg(Ac) mg(LAB)\(^{-1}\) | Y-to-Ac yield coefficient |
| \( Y_{A_{max}} \) | Y mg(Ac) mg(AAB)\(^{-1}\) | Y-to-Ac yield coefficient |
| \( Y_{E_{max}} \) | Y mg(Ac) mg(AAB)\(^{-1}\) | Y-to-Ac from EtOH yield coefficient |
| \( Y_{E_{max}} \) | Y mg(Ac) mg(AAB)\(^{-1}\) | Y-to-Ac from LA yield coefficient |
The deterministic model proposed in equations (2.1)–(2.8), can take the general form:

\[ \frac{dx_i}{dt} = f(x_{ij}, \theta), \quad (2.10) \]

where \( x_i \) denotes each of the state variables and \( f(x_{ij}, \theta) \) the model prediction of the change of \( x_i \) at time \( j \) as a function of the parameter vector \( \theta \). Each experimental observation \( D_{ij} \) can be considered as drawn from a normal distribution whose mean is equal to the model prediction \( f(x_{ij}, \theta) \), with a necessary standard deviation, \( \sigma \), that accounts for both experimental measurement error of the real observations and misprediction of the model. In this framework, each observation \( D_{ij} \) is drawn from a sampling distribution of the form

\[ D_{ij} \sim N(f(x_{ij}, \theta), \sigma), \quad (2.11) \]

thus allowing to reformulate the total posterior probability distribution in equation (2.9) as

\[ P(\theta | D) \propto \prod_{i=1}^{T} \prod_{j=1}^{N} N(f(x_{ij}, \theta), \sigma)P(\theta). \quad (2.12) \]

Using this form of the total posterior probability, in addition to the 24 parameters from equations (2.1) to (2.8), the total standard deviation, \( \sigma \), has also been estimated.

### 2.4.2. Variable scaling

For both collections of experimental data, the concentrations of microorganisms and metabolites differ by several orders of magnitude. As an example, after the transformation of CFU to biomass units in the experimental data of Camu et al. [4], the maximum concentration of AAB is approximately 0.0019 mg g\(^{-1}\) pulp\(^{-1}\), while the maximum concentration of the main substrate of AAB, EtOH, is approximately 22.4920 mg g\(^{-1}\) pulp\(^{-1}\).

These different orders of magnitude between the state variables can lead to numerical issues during optimization. In order to reduce such issues, all state variables were scaled by dividing each of the time series in the experimental data by its own maximum value. Consequently, possible large differences between the parameters to be estimated are avoided and, most importantly, the search space can be constrained.

Hence, in a first step, the parameters were estimated using time series with a maximum value of 1 and, in a second step, re-scaled to their original physical units through conversion factors derived from equations (2.1) to (2.8) (see electronic supplementary material, S3).

### 2.4.3. Priors

By scaling the system to allow maximum values for each time series equal to unity, the large differences in orders of magnitude between the parameter estimates, e.g. maximum specific growth rates in the boundaries of fractions of milligrams with respect to yield coefficients that might take values of hundreds, are regularized; in this case, by introducing a scale that needs a prior distribution to be sampled within values between 0 and 1. In that way, an independent normal distribution with mean 0.5 and a standard deviation of 0.3 as prior choice for each \( k \) element of \( \theta \) represents a weakly informative prior by introducing scale information of the original units in which the parameters of the model are originally measured. For the standard deviation \( \sigma \), a Cauchy distribution \( C \) with location and scale parameters of 0 and 1, respectively, was used as prior distribution. This choice follows the same reasoning as depicted for the \( k \) independent priors for \( \theta \), with the addition that the heavy tails of \( C \) allow for the sampling of extreme values which would account for outlying observations in the original data. To avoid the estimation of negative parameters, both priors are constrained to take values in the positive set of real numbers and are mathematically expressed as

\[ \{ \theta_k \sim N(0.5, 0.3), \quad \theta_k > 0 \} \]

and

\[ \{ \sigma \sim C(0, 1), \quad \sigma > 0 \} \quad (2.13) \]

### 2.4.4. Implementation

The Bayesian parameter estimation framework was performed with Stan [36], using the RStan interface package for R [37,38]. The model was solved as an initial value problem, where the initial concentrations
for the eight state variables in equations (2.1) to (2.8) were provided as they were reported in the original works of Camu et al. [4] and Papalexandratou et al. [24]. Sampling for obtaining the posterior distributions of the unknown parameters as well as the model predictions was conducted using full Bayesian inference through the Markov chain Monte Carlo (MCMC) No-U-Turn sampler method [39]. The ODEs were specified and solved by the built-in mechanism of Stan \texttt{rk45}, which provides a fourth- and fifth-order Runge–Kutta method for solving non-stiff systems [40,41]. All datasets were fitted by running four parallel Markov Chains of 3000 iterations each, 1000 of which were used as warm-up. Convergence of the sampling was determined by examining the $\hat{R}$ statistics computed by Stan.

2.5. Statistical analyses

Once the parameters of equations (2.1)–(2.8) were estimated, their posterior distributions obtained from fitting the model to each of the two trials, i.e. box 1 and box 2, reported by Papalexandratou et al. [24] were compared between each other. For doing so, an effect size statistic was used as proposed by Cohen [42], as a measure of the magnitude of either their relationship or difference. In that way, the effect size expressed as the standardized mean difference ($d$) of two independent continuous distributions was computed as

$$d = \frac{\bar{\theta}_1 - \bar{\theta}_2}{S_{\text{ws}}},$$

(2.14)

where $\bar{\theta}_1$ and $\bar{\theta}_2$ are the sampled means corresponding to the parameter $k$ obtained from box 1 and box 2, respectively, and $S_{\text{ws}}$ is the pooled within-groups standard deviation corresponding to the parameter $k$, which for the case of groups of equal sample sizes (in this study represented as the number of iterations of the MCMC sampler) is given by

$$S_{\text{ws}} = \sqrt{\frac{S_{k,1}^2 + S_{k,2}^2}{2}},$$

(2.15)

where $S_{k,1}$ and $S_{k,2}$ are the standard deviations of the posterior distribution of parameter $k$ for box 1 and box 2, respectively. We used a threshold of $|d| > 1.2$, in order to identify significant differences between parameters [43].

3. Results

3.1. Model’s diagnostics

In all three datasets, the proposed model was fitted without major issues. The calculated $\hat{R}$ statistic was 1 for all three cases (see electronic supplementary material, tables S4–S6), showing that convergence of the MCMC sampler was accomplished. Such a behaviour is also noticeable in the obtained traceplots (see electronic supplementary material, figures S6–S8), that show the typical ‘caterpillar’ shape as probe of a good mixing of the MCMC sampler along the exploration of the parameter space. Asymptotically, the ODE system converges to a stable fixed point (see electronic supplementary material, S5).

3.2. Metabolite and microbial population dynamics

The proposed model, as described in equations (2.1)–(2.8), fits each of the collections of data reported previously by Camu et al. [4] (figure 3) and Papalexandratou et al. [24] (figures 4 and 5) remarkably well. In each of the data collections, despite the noisy nature of the experimental data and the low sampling rate, the corresponding simulations show the microbial succession previously reported by several studies for Y, LAB and AAB that emerges from the interplay of metabolites and microbial communities. Thus, most of the experimental observations reported by Camu et al. [4] and Papalexandratou et al. [24] fall within the computed 95% credible intervals of the simulations indicating that the model is capable of predicting the dynamics of metabolites in all cases, even in those where theoretical knowledge is not fully reflected on the data. This is the case for the time series of LA in the data reported by Camu et al. [4], where its concentrations seem to stay steady after 48 h of fermentation, which would contradict its consumption by AAB as reported by Pereira et al. [30].
### 3.3. Parameter estimates

It was found that among the 24 parameters of the model, there are similar reported values for 14 of them for single strain cultures. This means that in our parameter estimation framework, for each dataset 14 parameter estimates can be compared with their counterparts in the literature. In this way, from a total of 42 parameter estimates, 17 (40.48%) of them are in accordance to the referenced ranges, 20 (47.62%) are out of the referenced range by less than the estimate divided or multiplied by 10, and 5 (11.90%) are out of the referenced range by more than the estimate divided or multiplied by 10.

---

**Figure 3.** Simulation results of the cocoa bean fermentation model for the data reported by Camu *et al.* [4]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines show the simulations of the model, while black points denote the experimental data of Camu *et al.* [4]. The red dashed lines represent the 95% credible interval of the model predictions.

**Figure 4.** Simulation results of the cocoa bean fermentation model for the data reported for box 1 by Papalexandratou *et al.* [24]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines show the simulations of the model, while black points denote the experimental data for box 1 of Papalexandratou *et al.* [24]. The red dashed lines represent the 95% credible interval of the model predictions.
3.3.1. Maximum specific growth rates

With respect to the maximum specific growth rates, the estimated values mostly fall within their reported values in the literature, with the exception of three estimates belonging to particular trials. Specifically, for the maximum specific growth rate of Y on Glc ($\mu_{\text{YGlc}}^{\text{max}}$), the obtained parameter values ranged from 0.06 to 0.37 h$^{-1}$. These values agree with the reported ones for species of this microbial group, between 0.0781 and 0.53 h$^{-1}$ [44–51] for Camu et al. [4] and box 2 [24] (green cells in table 3), while the estimate for box 1 [24] is far from the range in less than the estimate divided by 10 (orange cell in table 3). In a similar fashion, the estimated parameters for the maximum growth rate of LAB ($\mu_{\text{LAB}}^{\text{max}}$) varied across the three datasets between 0.36 and 0.5 h$^{-1}$, falling within the range of reported values in the literature between 0.0072 and 1.41 h$^{-1}$ [54–59] (all parameters are within green cells in table 3). By contrast, the maximum specific growth rate of Y on Fru ($\mu_{\text{YFru}}^{\text{max}}$) and AAB on EtOH ($\mu_{\text{AABEtOH}}^{\text{max}}$), do not agree completely with reported values in the literature. In the first case, only the estimate obtained from box 1 conducted by Papalexandratou et al. [24] falls within the range of 0.01–0.166 h$^{-1}$ [49,51–53] (green cell in table 3). In the latter case, the only estimate that does not fall within the range of 0.0106–0.25 h$^{-1}$ [60–64] is the one obtained from the data reported by Camu et al. [4] (orange cell in table 3). For the maximum specific growth rate of AAB on LA ($\mu_{\text{AABLA}}^{\text{max}}$), no values were reported in the literature for this microbial group on LA as carbon source. The estimated values for this parameter, were between 0.01 and 0.02 h$^{-1}$.

3.3.2. Substrate saturation constants

About the substrate saturation constants, here denoted by $K$, their estimated values for the three datasets agree with reported ones in the literature in one out of three instances. For these comparisons, in several occurrences a unit transformation was necessary from their original units in which they were reported to milligrams of substrate per millilitre of medium (mg(substrate) ml$^{-1}$), and assuming that one gram of pulp is equivalent to one millilitre of medium, since our estimates are given in mg(substrate) mg(pulp)$^{-1}$.
On the one hand, values reported in the literature for these parameters were found for the substrate saturation constants of Y on Glc ($K_{Y_{Glc}}$) [46,48], Y on Fru ($K_{Y_{Fru}}$) [53] and LAB on Glc ($K_{LAB_{Glc}}$) [57–59]. The obtained estimated values ranged from 30.01 to 35.32, 25.02 to 41.39 and 19.27 to 37.97 mg(substrate) ml$^{-1}$, respectively. Estimated values for $K_{Y_{Glc}}$ are far from the reported ranges by less than the

| parameter        | Camu | P. box 1 | P. box 2 | effect size $|d|$ | reported values | reference          |
|------------------|------|---------|---------|----------------|-----------------|--------------------|
| $\mu_{Y_{Glc}}$  | 0.253(0.094) | 0.063 (0.025) | 0.368 (0.131) | 3.23* | 0.0781 – 0.53 | [44 – 51]          |
| $\mu_{Y_{LAB}}$  | 0.359 (0.106) | 0.083 (0.03) | 0.572 (0.151) | 4.49* | 0.01 – 0.166 | [49,51–53]         |
| $\mu_{Y_{AAB_{max}}}$ | 0.358 (0.067) | 0.414 (0.096) | 0.499 (0.152) | 0.67 | 0.0072 – 1.41 | [54 – 59]          |
| $\mu_{Y_{EtOH}_{max}}$ | 0.380 (0.092) | 0.150 (0.051) | 0.168 (0.052) | 0.35 | 0.0106 – 0.25 | [60 – 64]          |
| $\mu_{Y_{EtOH}_{max}}$ | 0.008 (0.012) | 0.025 (0.017) | 0.022 (0.016) | 0.18 | n.a. | n.a. |
| $\mu_{Y_{EtOH}_{max}}$ | 35.322 (13.826) | 34.366 (14.964) | 30.015 (10.887) | 0.33 | 9.73 $\times 10^{-4}$ – 0.5 | [46,48] |
| $\mu_{Y_{EtOH}_{max}}$ | 35.492 (15.253) | 25.015 (14.187) | 41.386 (17.259) | 1.04 | 582 – 772 | [53] |
| $\mu_{Y_{EtOH}_{max}}$ | 37.966 (12.37) | 31.664 (13.882) | 19.272 (10.599) | 1.00 | 0.790 – 178.0 | [57–59]          |
| $\mu_{Y_{EtOH}_{max}}$ | 16.056 (5.646) | 3.818 (1.637) | 4.056 (1.98) | 0.13 | n.a. | n.a. |
| $\mu_{Y_{EtOH}_{max}}$ | 2509.622 | 312.051 | 81.981 (41.365) | 2.05* | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 0.0333 (0.0051) | 0.0517 (0.0119) | 0.092 (0.0248) | 2.07* | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 0.0054 (0.0014) | 0.0637 (0.0183) | 0.0686 (0.025) | 0.22 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 0.0069 (0.0014) | 0.0004 (0.0002) | 0.0004 (0.0002) | 0.00 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 33.400 (11.255) | 240.926 (59.899) | 119.672 (35.155) | 2.47* | 1.56 – 66.67 | [44 – 46,30, 51,65,66] |
| $\alpha_{Y_{max}}$ | 29.259 (10.852) | 20.217 (13.24) | 3.323 (3.417) | 1.75* | 4.50 – 200 | [59,67] |
| $\alpha_{Y_{max}}$ | 41.105 (11.215) | 244.153 (62.958) | 232.383 (62.958) | 0.21 | 43.48 – 200 | [51,52] |
| $\alpha_{Y_{max}}$ | 7.436 (4.526) | 11.941 (6.061) | 8.201 (4.503) | 0.70 | 1.39 – 21.49 | [46,49,50,68] |
| $\alpha_{Y_{max}}$ | 5.927 (3.351) | 11.195 (5.005) | 6.008 (3.882) | 1.16 | 5.7878 | [49] |
| $\alpha_{Y_{max}}$ | 1298.070 | 378.452 | 170.44 (35.155) | 1.81* | 8.06 – 166.67 | [61,63,69] |
| $\alpha_{Y_{max}}$ | 10.617 (1.54) | 2.785 (0.664) | 2.138 (0.799) | 0.88 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 1928.619 | 287.511 | 62.9 (38.531) | 2.17* | 7.94 | [63] |
| $\alpha_{Y_{max}}$ | 5.612 (0.919) | 3.279 (1.125) | 2.899 (1.898) | 0.24 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 104.056 (98.355) | 576.471 | 321.858 (132.72) | 1.19 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 1427.225 | 666.903 | 385.172 (177.474) | 1.00 | n.a. | n.a. |
| $\alpha_{Y_{max}}$ | 0.149 (0.010) | 0.168 (0.014) | 0.168 (0.013) | 0.00 | n.a. | n.a. |

On the one hand, values reported in the literature for these parameters were found for the substrate saturation constants of Y on Glc ($K_{Y_{Glc}}$) [46,48], Y on Fru ($K_{Y_{Fru}}$) [53] and LAB on Glc ($K_{LAB_{Glc}}$) [57–59]. The obtained estimated values ranged from 30.01 to 35.32, 25.02 to 41.39 and 19.27 to 37.97 mg(substrate) ml$^{-1}$, respectively. Estimated values for $K_{Y_{Glc}}$ are far from the reported ranges by less than the
estimate multiplied by 10 (orange cells in table 3); while for \( K_{\text{Fru}}^L \), the estimates were further from the referenced range than the estimate multiplied by 10 (red cells in table 3). From these estimates, the ones corresponding to \( K_{\text{Glc}}^L \) were those which fall within the reported range of 0.79–178.0 mg(Glc) ml\(^{-1}\) [57–59] (green cells in table 3).

On the other hand, no values of substrate saturation constants were reported either for AAB on EtOH (\( K_{\text{EtOH}}^L \)) or AAB on LA (\( K_{\text{LA}}^{L} \)). For these parameters, their estimated values ranged from 3.81 to 16.06 and 81.981 to 2509.62 mg(substrate) g(pulp)\(^{-1}\), respectively. For \( K_{\text{LAB}}^{L} \), this considerably higher value for the upper limit of the range was obtained for the Camu et al. [4] data. This inflation of values was observed for other parameters, i.e. AAB-to-EtOH yield coefficient (\( Y_{\text{EtOH/AAB}}^L \)), AAB-to-LA yield coefficient (\( Y_{\text{LAB/AAB}}^L \)) and AAB-to-Ac from LA yield coefficient (\( Y_{\text{Ac/AAB}}^L \)) of this dataset as well.

### 3.3.3. Mortality rate constants

For the mortality rate constants, \( k \), no values were reported in the literature. Here, their estimated values for \( Y \) (\( k_Y \)), LAB (\( k_{\text{LAB}} \)) and AAB (\( k_{\text{AAB}} \)) were in the ranges of 0.033–0.092 mg(EtOH)\(^{-1}\) h\(^{-1}\), 0.0054–0.067 mg(LA)\(^{-1}\) h\(^{-1}\) and 0.0004–0.0069 mg(Ac)\(^{-2}\) h\(^{-1}\), respectively. These estimates varied considerably between the data of Camu et al. [4] and Papalexandratou et al. [24]. In the latter, the only estimate that differed much between box 1 and box 2 was the one corresponding to \( k_Y \).

### 3.3.4. Yield coefficients

Finally, a higher variability among the obtained parameter estimates was found for the yield coefficients, \( Y \). These differences were notable between the two studies of Camu et al. [4] and Papalexandratou et al. [24], as well as the two trials (boxes 1 and 2) of the latter. In more detail, the estimated yield coefficient of Y-to-EtOH from Glc (\( Y_{\text{EtOH/Glc}}^L \)) was the only one that agreed in all three datasets with those reported in the literature (green cells in table 3). Their estimated values ranged between 7.44 and 11.94 mg(EtOH) mg(Y)\(^{-1}\), falling in the referenced range of 1.39–21.49 mg(EtOH) mg(Y)\(^{-1}\) [46,49,50,68]. For the yield coefficient of LAB-to-Glc (\( Y_{\text{Glc/LAB}}^{L} \)), the fits corresponding to the data of Camu et al. [4] and box 1 [24] were contained in the referenced range of 1.56–66.67 mg(Glc) mg(LAB)\(^{-1}\) [59,67] (green cells in table 3) with values of 29.23 and 20.22 mg(Glc) mg(LAB)\(^{-1}\), respectively. The remaining estimate \( Y_{\text{Glc/LAB}}^{L} \) for box 2 was far from the reported range in less than the estimate multiplied by 10, with a mean estimated of 3.32 mg(Glc) mg(LAB)\(^{-1}\) (orange cell in table 3).

By contrast, there are estimated yield coefficients that do not agree completely with previously reported values. On the one hand, the estimated value of 33.4 mg(Glc) mg(Y)\(^{-1}\) (green cell in table 3) for the yield coefficient of Y-to-Glc (\( Y_{\text{Glc/Y}}^{L} \)), in the dataset of Camu et al. [4] only, agree with the ranges of 1.56–66.67 mg(Glc) mg(Y)\(^{-1}\) [44–46,50,51,65,66]. Their counterparts from boxes 1 and 2 reported by Papalexandratou et al. [24], are away from the reported range in less than the estimate multiplied by 10 (orange cells in table 3) with values of 240.93 and 119.67 mg(Glc) mg(Y)\(^{-1}\), respectively. For the yield coefficient of Y-to-Fru (\( Y_{\text{Fru/Y}}^{L} \)), all estimated parameters are far from the referenced range of 43.48–200 mg(Fru) mg(Y)\(^{-1}\) [51,52] in less than the estimate divided by 10 (orange cells in table 3) with values between 41.11 and 244.15 mg(Fru) mg(Y)\(^{-1}\). A similar situation is observed for coefficients Y-to-EtOH from Fru (\( Y_{\text{Fru/EtOH}}^{L} \)) and AAB-to-EtOH (\( Y_{\text{EtOH/AAB}}^{L} \)) (orange cells in table 3), with estimated parameters between 5.927 and 11.195 mg(EtOH) mg(Y)\(^{-1}\) and a single referenced value of 5.7878 mg(EtOH) mg(Y)\(^{-1}\) [49] for \( Y_{\text{Fru/EtOH}}^{L} \). As estimated parameters between 170.44 and 1298.070 mg(EtOH) mg(AAB)\(^{-1}\) with a referenced range of 8.06–166.67 mg(EtOH) mg(AAB)\(^{-1}\) [61,63,69] for \( Y_{\text{EtOH/AAB}}^{L} \).

Moreover, the estimated yield coefficients for AAB-to-LA (\( Y_{\text{LAB/AAB}}^{L} \)) do not agree in any of the data collections with the reference value of 7.94 mg(LA) mg(AAB)\(^{-1}\) [63] with values far from the reference more than 10 times the parameter for the datasets of Camu et al. [4] and box 1 of Papalexandratou et al. [24] (red cells in table 3) and one value far from the reference in less than the estimate divided by 10 for box 2 (orange cell in table 3).

For the rest of yield coefficients: LAB-to-LA (\( Y_{\text{LAB/LAB}}^{L} \)), LAB-to-Ac (\( Y_{\text{Ac/LAB}}^{L} \)), AAB-to-Ac from EtOH (\( Y_{\text{EtOH/AAB}}^{L} \)) and AAB-to-Ac from LA (\( Y_{\text{Ac/AAB}}^{L} \)), no values were reported in the literature. Here their estimated values were in the ranges of 2.14–10.62 mg(LA) mg(LAB)\(^{-1}\), 2.89–5.61 mg(Ac) mg(LAB)\(^{-1}\), 104.06–576.47 mg(Ac) mg(AAB)\(^{-1}\) and 385.17–1427.23 mg(Ac) mg(AAB)\(^{-1}\), respectively.

Special attention needs to be given to the values of \( Y_{\text{EtOH/AAB}}^{L} \), \( Y_{\text{LAB/AAB}}^{L} \) and \( Y_{\text{Ac/AAB}}^{L} \) for the data of Camu et al. [4], which showed inflated values that are not biologically plausible.

The effect of the measurement errors is discussed in electronic supplementary material, S6.
3.4. Statistical comparison of fermentation trials

The statistical comparison of the parameter estimates between the two fermentation trials conducted by Papalexandratou et al. [24] showed that significant differences exist among them (table 3), even though these were done under slightly different conditions in the same region. In this respect, the parameter estimates that showed such a significantly large difference depending on which trial they were derived from, correspond to \( \mu_{AAB}^\text{m} \), \( \mu_{AAB}^\text{k} \), \( K_{LA}^\text{AAB} \), \( K_{Y}^\text{AAB} \), \( Y^\text{Glc}_{\text{LAB}} \), \( Y^\text{EtOH}_{\text{AAB}} \) and \( Y^\text{LA}_{\text{AAB}} \). In the comparison of all these parameters, the computed absolute value of the standardized mean difference (effect size \( d \)) was greater than the threshold of 1.2 suggested by Sawilowsky [43]. Such a result leads to hypothesize that minor changes in both, methodologies and regions, affect the parameters of the model as it will be addressed in the discussion section.

4. Discussion

4.1. Model fitting

As shown in the aforementioned results, our current model for cocoa bean fermentation is capable of reproducing each of the datasets with high accuracy. This means that the mechanistic assumptions made here are in accordance with the available biological knowledge to a considerably good degree, as it has been reflected in the conducted simulations. In this sense, our model represents a mechanistic approach which allows for a deep understanding of the transient responses of the process in a dynamic way, as opposed to current metabolic flux analyses [15,16] that assume steady-state metabolic conditions. Moreover, it represents a fully working kinetic model as opposed to a previous attempt which is capable of simulating metabolites and products time-courses only [17].

However, a detailed analysis of the resulting fits provides insight into the validity of some of the regulatory assumptions underlying the model, the relevance of additional effects not included in the present version of the model, as well as differences between the experimental set-ups behind the datasets.

4.2. Regulatory assumptions

By analysing the parameter estimates obtained here, important features of the model can be explored for its enhancement in future iterations. A Bayesian framework for parameter estimation, as used here, provides a scheme to investigate their uncertainty and determine their possible uniqueness. Therefore, it serves as a descriptive source for deriving the plausibility of the regulatory assumptions of the model.

In this sense, one of the particularities of the fitted model is the presence of strongly elevated estimates for the data of Camu et al. [4]. Specifically, the parameters showing such values were: (i) the substrate saturation constant \( K_{LA}^\text{AAB} \) and (ii) the yield coefficients \( Y^\text{EtOH}_{\text{AAB}} \), \( Y^\text{LA}_{\text{AAB}} \) and \( Y^\text{Ac}_{\text{AAB}} \). Looking at the standard deviations (table 3, electronic supplementary material, S4, figures S9–S11) of their corresponding posterior distributions, it can be noted that there is a huge uncertainty in their values (large errors). These uncertainties reveal that the parameters cannot be uniquely estimated from these particular data, suggesting a practical non-identifiability of the parameters with the data reported by Camu et al. [4]. The reason for this characteristic can be threefold. Firstly, noise in the experimental data prevents a unique determination of the model’s parameters because of an insufficient signal-to-noise response [70]. Secondly, the data may be incongruent particularly with the model mechanisms of growth of AAB on LA and the interactions AAB–EtOH. Finally, the estimated parameters might be correlated.

In our opinion, all elevated parameters related to the growth of AAB on LA, i.e. \( K_{LA}^\text{AAB} \), \( Y^\text{LA}_{\text{AAB}} \) and \( Y^\text{Ac}_{\text{AAB}} \) can be a result of noise in experimental data. Thus, as revealed by visual inspection of figure 3d where the real data does not reflect an accentuated decrease in the concentration of LA as opposed to the data reported by Papalexandratou et al. [24] (figures 4d and 5d), where such decrease exists after 72 h of the fermentation process. For the remaining elevated parameter which is related to the consumption of EtOH by AAB (\( Y^\text{EtOH}_{\text{AAB}} \)), a straight interpretation of its value of \( \approx 1300 \text{ mg(Glc)} / \text{mg(Y)} \) would imply that 1300 mg of EtOH are consumed by 1 mg of AAB; or in other words, that for generating 1 mg of AAB, 1300 mg of EtOH is required. Obviously, a yield coefficient of this order of magnitude is biologically implausible and for this reason it could be argued that the proposed model is not entirely capturing all the inherent mechanisms of the AAB–EtOH interaction in the fermentation process. A possible explanation of this specific inflated parameter value, is that \( Y^\text{EtOH}_{\text{AAB}} \) is not only capturing the consumption of EtOH by AAB, but also possible physical processes such as evaporation of this
metabolite. Temperature data, so far not implemented in the model, gave the reason for this hypothesis. More precisely, the data of Camu et al. [4] shows higher temperatures of the fermentation mass much earlier in the process compared to the data reported by Papalexandratou et al. [24]. In the first case, a temperature above 35°C was reached right after 30 h and its maximum of ≈45°C at 70 h of the fermentation process. In the latter, similar temperatures were reached at 40 and 80 h of the process. This disparity might explain why the estimated parameter values for $Y_{EOH\ AAB}$ in the Papalexandratou et al. [24] data collection are between 3 to 5 times less compared to the value obtained for the data of Camu et al. [4]. Finally, correlation between the estimated parameters might be playing an important role in the non-identifiability of these inflated parameters. In other words, interdependencies between different sets of parameters limit the MCMC sampler to freely explore the solution space. In that sense, from a simple correlation analysis (see electronic supplementary material, S7), we did not find remarkable patterns among the posterior probabilities of the parameter estimates, not even in the inflated ones.

According to these hypotheses, further iterations of the model should include additional physical effects, especially temperature. Moreover, pH conditions during the conduction of the fermentation should be also taken into account, as well as a non-dimensionalization of the model to identify correlated estimates and reduce their number.

4.3. Parameter conformance to values in the literature

The parameter ranges indicated as reported values in table 3 are in many cases referring to different experimental conditions and/or a specific microorganismal strain and might therefore not be directly comparable to the biological situation discussed here. We resorted to these values, whenever we failed to identify parameter values directly applicable to cocoa bean fermentation, in order to at least provide an order-of-magnitude estimate.

Hence, among the different estimated parameters, few did not agree with previously reported values in the literature as pointed out in §3.3. For the substrate saturation constants $K_{Y \text{Glc}}$ and $K_{Y \text{Fru}}$, such a difference can be explained by the fact that the reported values correspond to the growth of $Y$ under aerobic conditions, as opposed to the anaerobic earlier stage of cocoa bean fermentation where $Y$’s growth takes place. Consequently, higher values for these parameters as reported here, reflect a slow uptake rate of Glc and Fru by $Y$ under anaerobic conditions. Finally, from a general point of view, the various discrepancies between the obtained parameter estimates for the growth rates and yield coefficients with their values reported in the literature for single species confirms the high growth rates and yields coefficients that mixtures of microorganisms might show in fermentation processes [71].

4.4. Comparison of parameter estimates

As mentioned in §2.1, the study performed by Papalexandratou et al. [24], involved two fermentation trials conducted in two cocoa-producing farms in Brazil belonging to the same region. These trials denoted as ‘box 1’ and ‘box 2’, differed between each other in minor aspects. After we identified eight significantly different estimates between these trials (table 3), in the following paragraphs, three main differences were taken into account in order to formulate hypotheses on how the parameter estimates are affected when applying the same fermentation method, i.e. wooden boxes, under similar environmental conditions in distinct fermentations. The three main differences between the trials are: (i) initial concentration of microorganisms, (ii) concentration ratios of initial substrates, and (iii) evaporation rates.

4.4.1. Initial concentration of microorganisms

The initial concentrations of microorganisms between boxes 1 and 2 did not differ much, with the exception of $Y$. For $Y$, the initial concentrations in boxes 1 and 2 were equal to 0.18 and 0.005 mg g$^{-1}$, respectively. This difference as well as the microbial diversity that has been seen along different fermentation trials [2] determined that the estimates of these growth rates differ between each fermentation trial; with higher values of the growth rates $\mu_{\text{Glc}}$ and $\mu_{\text{Fru}}$ for box 2 than for box 1. A similar effect is evident in the yield coefficient related to the growth of $Y$ on Glc ($Y_{\text{GlcY}}$), where box 1 showed a higher value than the one obtained for box 2 as a consequence of the higher initial amount of $Y$ in box 1. In other words, a higher initial concentration of $Y$ determines the estimation of lower maximum specific growth rates as well as higher estimates for the yield coefficient of $Y$ on Glc. This can be explained that given a high initial microbial population, it needs a lower cell division rate to reach the maximum described by the observed data and an increased rate of uptake of its substrate.
4.4.2. Concentration ratios of initial substrates

Worthy of attention, was the difference in the initial concentrations of the main substrates Glc and Fru which might play an important role in the growth of Y and LAB. For box 1, the initial concentrations of Glc and Fru are 55.482 and 49.669 mg g\(^{-1}\), respectively; while for box 2, these are 42.936 and 67.249 mg g\(^{-1}\), respectively. This ratio would explain the significant difference in the yield coefficient of the growth of LAB on Glc (\(Y_{\text{Glc|LAB}}\)) between boxes 1 and 2. In box 2, this yield coefficient was estimated significantly lower than for box 1 which leads to the hypothesis that this phenomena might be the result of a growing population of Y restraining the access to Glc to the LAB microbial group. This final hypothesis coincides with our assumption of resource-type competition between Y and LAB [34,35], which determined a less successful \(Y_{\text{Glc|LAB}}\) for LAB under a lower initial concentration of its main substrate Glc in box 2.

4.4.3. Evaporation rates

Among the subtle differences between boxes 1 and 2 reported by Papalexandratou et al. [24], the final one to be considered is the possible uneven evaporation rates between them. With this in mind, it is not deceitful to expect a higher evaporation rate of volatile metabolites in a fermenting mass protected only by a metal roof, as reported for box 1, than in a fermenting mass held within a fermentary room, as reported for box 2. In this respect, the higher estimated yield coefficient of growth of AAB on EtOH (\(Y_{\text{EtOH|AAB}}\)) in box 1 than in box 2, might be explained for a possibly greater evaporation rate of EtOH in box 1, similar to the possible explanation of its inflated counterpart determined for the Camu et al. [4] dataset. A similarly higher evaporation rate for LA and Ac in box 1 could be the explanation for the differences among the remaining parameter estimates, i.e. the Contois substrate saturation constant for the growth of AAB on LA (\(K_{\text{LA|AAB}}\)), the yield coefficient of consumption of LA by AAB (\(Y_{\text{LA|AAB}}\)) as well as the large variances of the yield coefficients of the production of Ac from EtOH and LA by AAB (\(Y_{\text{EtOH|AAB}}\) and \(Y_{\text{LA|AAB}}\), respectively). This hypothesis can be also extended to the lower value in the mortality rate term of Y (\(k_Y\)) observed in box 1 than in box 2, where EtOH loses its effect on decreasing Y’s population due to a higher evaporation. In other words, the mortality rate of Y might be lowered in the presence of an increasing evaporation rate of EtOH in the fermenting mass.

5. Conclusion

The model presented here is a first biochemically plausible, ODE-based kinetic model of cocoa bean fermentation capable of reproducing the known sequential activation of microbial communities and capable of fitting available experimental data to an acceptable degree. However, it is necessarily a simplification of the diverse biological processes involved in cocoa bean fermentation. The remaining discrepancies between model prediction and experimental data, as well as those parameter values outside the biologically plausible ranges, point to the fact that relevant aspects of the processes have not been taken into account.

Based on the model features, we can hypothesize that the following regulatory mechanisms might exist: (i) resource-type competition between Y and LAB, (ii) microbial death is determined to a good degree by direct action of fermentation products upon their respective producing microorganisms and (iii) chemical and physical factors intervene in the decrement of volatile products, i.e. EtOH and LA, rather than microbial activities only.

This mathematical model allows relating observed microbial population sizes and concentrations of the five chemical compounds considered here, i.e. Glc, Fru, EtOH, LA and Ac, during the time course of fermentation with growth rates, mortality rates, substrate saturation constants and yield coefficients as intrinsic systemic parameters.

Additionally, the capability of the model to ‘reverse-engineer’ differences from the observed time courses of two trials conducted in the same region under the same methodology showed how these systemic parameters might be affected by minor changes between one and other fermentation trial. The cocoa and chocolate markets require a steady flow of high-quality raw material resulting from diverse types of fermentation. Although fermentation practices will remain locally determined, model-based recommendations to farmers on practices and use of specific starter cultures might help increase the quality of cocoa bean raw material prior to shipment. This will ultimately increase the sustainability of cocoa bean supply.
Subsequent versions of the model should include additional chemical and physical effects, such as temperature and pH dependence of kinetic parameters, the spatial heterogeneity of a fermentation pile, impact of additional (commonly occurring) microorganisms, as well as a further compartmentalization including the inner bean and the incorporation of sucrose as an additional carbon source, serving also as a (time-delayed) source of glucose and fructose.

Besides the extension by further chemical and physical effects, eventually such a kinetic model needs to be interfaced with the more microscopic, metabolic perspective put forward in other studies [15,16]. Should high-quality genome-scale metabolic models be available, flux balance analysis [72] may provide a suitable theoretical framework for such an approach on three levels: (i) A metabolic pathway analysis of synergies and competitions (e.g. using the methodology from [73]) may point to additional modes of interaction among the species involved. (ii) A detailed exploration of the emerging pattern of chemical compounds as a function of the fermentation time course may become feasible by incorporating the biochemical interactions of the cocoa bean and the microorganisms. (iii) The relevance of a larger diversity in microorganisms (e.g. different yeast strains) can be assessed. With the availability of genome-scale metabolic models currently developing rapidly [74], we expect this avenue of research to become feasible in the very near future.

The recent finding [75] about the metabolic interplay of yeast and LAB is an example of the richness of this metabolic foundation underlying the dynamics leading to the successful fermentation of a cocoa bean.

Data accessibility. The datasets and codes supporting this article are deposited at Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.321d33v [76].

Authors’ contributions. M.H. and S.G. conceived the original idea. M.M., S.G. and M.H. developed the model. M.M. conducted simulations and analyses. M.U. contributed and endorsed the biological hypotheses derived from the model. All contributed to writing and editing of the manuscript.

Competing interests. The authors declare no competing interests.

Funding. This work was funded by Barry-Callebaut through the Cocoa Metabolomics (COMETA) project, driven by Jacobs University Bremen.

Acknowledgements. We gratefully acknowledge Dr Anne Grimbs for her excellent and valuable advice and comments. We are also grateful to two anonymous reviewers, who provided comments that considerably improved the manuscript. M.M. acknowledges Silvana Beltrán Torres for her intellectual assistance.

References

1. Lopez AS, Dimick PS. 1995 Cocoa fermentation. In Biotechnology (eds G Reed, TW Napadowithana), 2nd edn, vol. 9, ch. 14, pp. 561–577. Weinheim, Germany: VCH Verlagsgesellschaft mbH.
2. Schwan RF, Pereira GVdM, Fleet GH. 2014 Microbial activities during cocoa fermentation. In Cocoa and coffee fermentations (eds RF Schwan, GH Fleet), ch. 4, pp. 129–192. Boca Raton, FL: CRC Press.
3. De Vuyst L, Weckx S. 2016 The cocoa bean fermentation process: from ecosystem analysis to starter culture development. J. Appl. Microbiol. 121, 5–17. (doi:10.1111/jam.13045)
4. Camu N, Verbrugghe K, Cleenwerck I, Vandamme P, Takrama JS, Vancanneyt M. 2007 Dynamics and biodiversity of populations of lactic acid bacteria and acetic acid bacteria involved in spontaneous heap fermentation of cocoa beans in Ghana. Appl. Environ. Microbiol. 73, 1809–1824. (doi:10.1128/AEM.02189-06)
5. Lefebre T, Gobert W, Wanden G, Camu N. 2011 Dynamics and species diversity of communities of lactic acid bacteria and acetic acid bacteria during spontaneous cocoa bean fermentation in vessels. Food Microbiol. 28, 457–464. (doi:10.1016/j.fm.2010.01.010)
6. Pereira GVdM, Magalhaes KT, de Almeida EG, da Silva Coelho I, Schwan RF. 2013 Spontaneous cocoa bean fermentation carried out in a novel-design stainless steel tank: influence on the dynamics of microbial populations and physical-chemical properties. Int. J. Food Microbiol. 161, 121–133. (doi:10.1016/j.ijfoodmicro.2012.11.018)
7. Miguel IMdV, Duarte MGdCP, Moreira WF, Dias DR, Schwan RF. 2013 Microbial succession and the dynamics of metabolites and sugars during the fermentation of three different cocoa (Theobroma cacao L.) hybrids. Food Res. Int. 54, 9–17. (doi:10.1016/j.foodres.2013.06.001)
8. De Vuyst L, Weckx S. 2016 The functional role of lactic acid bacteria in cocoa bean fermentation. In Biotechnology of lactic acid bacteria (eds F Mozzi, RR Raya, GM Vignolo), 2nd edn, ch. 16, pp. 246–278. Hoboken, NJ: Wiley Blackwell.
9. Schwan RF, Wheals AE. 2004 The microbiology of cocoa fermentation and its role in chocolate quality. Crit. Rev. Food Sci. Nutr. 44, 205–221. (doi:10.1080/10408699409464104)
10. Lima L, Almeida MH, Nout MJR, Zwiering MH. 2011 Theobroma cacao L., ‘The Food of the Gods’: quality determinants of commercial cocoa beans, with particular reference to the impact of fermentation. Crit. Rev. Food Sci. Nutr. 51, 731–761. (doi:10.1080/10408690809979913)
11. Ho VTT, Zhao J, Fleet G. 2014 Yeasts are essential for cocoa bean fermentation. Int. J. Food Microbiol. 174, 72–87. (doi:10.1016/j.ijfoodmicro.2013.12.014)
12. Ho VTT, Zhao J, Fleet G. 2015 The effect of lactic acid bacteria on cocoa bean fermentation. Int. J. Food Microbiol. 205, 54–67. (doi:10.1016/j.ijfoodmicro.2015.03.031)
13. MacManus Chimney N, Ogunkunle A, Olukunle O. 2010 Cocoa bean (Theobroma cacao L.) drying kinetics. Chiném J. Agri. Res. 70, 633–639. (doi:10.4067/S0718-58392010000400004)
14. Hii C, Law C, Suzannah S. 2012 Drying kinetics of the individual layer of cocoa beans during heat pump drying. J. Food Eng. 108, 276–282. (doi:10.1016/j.jfoodeng.2011.08.017)
15. Adler P, Bolten CJ, Doehn K, Hansen CE, Wittmann C. 2013 Core fluxome and metabolome of lactic acid bacteria under simulated cocoa pulp fermentation conditions. Appl. Environ. Microbiol. 79, 5670–5681. (doi:10.1128/AEM.01483-13)
16. Adler P, Frey LJ, Berger A, Bolten CJ, Hansen CE, Wittmann C. 2014 The key to acetate: metabolic fluxes of acetic acid bacteria under cocoa pulp fermentation-simulating conditions. Appl. Environ. Microbiol. 80, 4702–4716. (doi:10.1128/AEM.01048-14)
32. Moens F, Lefebre T, De Vuyst L. 2014 Oxidation of metabolites highlights the microbial interactions and role of Acetobacter pasteurianus during cocoa bean fermentation. Appl. Environ. Microbiol. 80, 1848 – 1857. (doi:10.1128/AEM.03443-14)

33. Reilly PJ. 1974 Stability of commensalistic systems. Biotechnol. Bioeng. 16, 1373 – 1392. (doi:10.1002/bit.260161006)

34. Fredrickson AG. 1977 Behavior of mixed cultures of microorganisms. Annu. Rev. Microbiol. 31, 63 – 88. (doi:10.1146/annurev.mi.31.100177.000431)

35. Fredrickson A, Stephanopoulos G. 1981 Microbial competition. Science 213, 972 – 979. (doi:10.1126/science.6728409)

36. Carpenter B et al. 2017 Stan: a probabilistic programming language. J. Stat. Softw. 76, 1 – 32. (doi:10.18637/jss.v076.i01)

37. R Core Team. 2017 R: a language and programming environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

38. Stan Development Team. 2016 RStan: the R interface to Stan. R package version 2.14.1. (doi:10.5281/zenodo.78717)

39. Homan MD, Gelman A. 2014 The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Monte Carlo. J. Mach. Learn. Res. 15, 1593 – 1623.

40. Dorand H, Prince P. 1960 A family of embedded Runge-Kutta formulas. J. Comput. Appl. Math. 6, 19 – 26. (doi:10.1016/0771-050X(80)90013-3)

41. Ahnert K, Mulansky M. 2011 Odeint—solving ordinary differential equations in C++. AIP Conf. Proc. 1389, 1586 – 1589. (doi:10.1063/1.3637934)

42. Goh J. 1988 Statistical power analysis for the behavioral sciences, 2nd edn. Hillsdale, NJ: Erlbaum Associates.

43. Sawilowsky S. 2009 New effect size rules of thumb. J. Modern Appl. Stat. Methods 8, 467 – 474. (doi:10.22237/jmasm/1257035100)

44. Bideaux C, Alfenne S, Camelye X, Molina-Jouve C, Urbelaresa J-L, Guillouet E. 2006 Minimization of glycerol production during the high-performance fed-batch ethanol fermentation process in Saccharomyces cerevisiae, using a metabolic model as a prediction tool. Appl. Environ. Microbiol. 72, 2134 – 2140. (doi:10.1128/AEM.72.3.2134-2140.2006)

45. Alfenne S, Camelye X, Benbadis L, Bideaux C, Urbelaresa J-L, Goma G, Molina-Jouve C, Guillouet SE. 2004 Aeration strategy: a need for very high ethanol performance in Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 63, 537 – 542. (doi:10.1007/s00253-003-1393-5)

46. Sonnleitner B, Koppelin O. 1966 Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. Biotechnol. Bioeng. 28, 927 – 937. (doi:10.1002/bt.260280620)

47. Kaspar von Meyenburg H. 1969 Energetics of the budding cycle of Saccharomyces cerevisiae during glucose limited aerobic growth. Archiv für Mikrobiologie 66, 289 – 303. (doi:10.1007/ BF00448365)

48. Snoep JL, Milvebi M, Schuurmans JM, Rohwer JM, Teiseira de Mattos MI. 2009 Control of specific growth rate in Saccharomyces cerevisiae. Microbiology 155, 1699 – 1707.

49. Wang D, Xu Y, Hu J, Zhao G. 2004 Fermentation kinetics of different sugars by apple wine yeast Saccharomyces cerevisiae. J. Institute Brewing 110, 340 – 346. (doi:10.1002/jibf.200630.x)

50. Aguilar-Uscanga M, Garcia-Alvarado Y, Gomez-Rodriguez J, Pihler T, Delia M, Sreihana P. 2011 Modelling the growth and ethanol production of Brettanomyces bruxellensis at different glucose concentrations. Lett. Appl. Microbiol. 53, 141 – 149. (doi:10.1111/j.1365-2672.2011.05181.x)

51. Lee SK, Skotnicki ML, Tribe DE, Rogers PL. 1981 The kinetics of ethanol production by Zymomonas mobilis on fructose and sucrose media. Biotechnol. Lett. 3, 207 – 212. (doi:10.1007/BF00154646)

52. Tomar-Diaz I, Delecon C, Burati J. 1983 The kinetics of ethanol production by Zymomonas mobilis on fructose medium. Biochim. Biophys. Acta 705, 409 – 412. (doi:10.1016/0005-276X(82)90126-3)

53. Arroyo-López P, Queral A, Barrio E. 2009 Application of a substrate inhibition model to the effect of fructose concentration on the growth of diverse Saccharomyces cerevisiae strains. J. Ind. Microbiol. Biotechnol. 36, 663 – 669. (doi:10.1007/s10295-009-0535-x)

54. Berry AR, Franco CM, Zhang W, Middelberg AP. 1999 Growth and lactic acid production in batch culture of Lactobacillus rhamnosus in a defined medium. Biotechnol. Lett. 21, 163 – 167. (doi:10.1023/A:1002653609605)

55. Yoo I-K, Chang H-N, Lee E-G, Chang Y-K, Moon S-H. 1996 Effect of pH on the production of lactic acid and secondary products in batch cultures of Lactobacillus casei. J. Microbiol. Biotechnol. 6, 482 – 486.

56. Siebold M, Freiling P, Joppin R, Rindfleisch D, Schougel K. 1995 Comparison of the production of lactic acid by three different lactobacilli and its recovery by extraction and electrodeysis. Proc. Biochem. 30, 81 – 95. (doi:10.1016/S0032-9592(95)07011-3)

57. Germain A, Pave A. 1983 On parameter estimation of Monod’s bacterial growth model from batch culture data. J. Gen. Appl. Microbiol. 29, 91 – 101. (doi:10.2323/jgam.29.91)

58. Äkerberg C, Hofvendahl K, Zacchi G, Hahn-Hägerdal B. 1998 Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis sp. lactis ATCC 19435 in whole-wheat flour. Appl. Microbiol. Biotechnol. 49, 682 – 690. (doi:10.1007/s002530051222)

59. Fu W, Mathews A. 1999 Lactic acid production from lactose by Lactobacillus plantarum: kinetic model and effects of pH, substrate, and oxygen. Biochem. Eng. J. 3, 163 – 170. (doi:10.1006/bieng.1999.0139)

60. Park YS, Oikawa H, Fukaya M, Okumura H, Kawamura Y, Toda K. 1989 Enhancement of acetic acid production in a high cell density culture of Acetobacter aceti. J. Fermentation Bioeng. 68, 315 – 319. (doi:10.1016/0922-338X(89)90044-2)

61. Bar R, Ganter JL, Kirwan DJ. 1987 An unusual pattern of product inhibition: batch acetic acid fermentation. Biotechnol. Bioeng. 29, 796 – 798. (doi:10.1002/bit.260290625)
62. Park YS, Toda K, Fukaya M, Okumura H, Kawamura Y. 1991 Production of a high concentration acetic acid by Acetobacter aceti using a repeated fed-batch culture with cell recycling. Appl. Microbiol. Biotechnol. 35, 149 – 153. (doi:10.1007/BF00184678)

63. Salgueiro Machado S, Luttik MAH, van Dijken JP, Jongejan JA, Pronk JT. 1995 Regulation of alcohol-oxidizing capacity in chemostat cultures of Acetobacter pasteurianus. Appl. Microbiol. Biotechnol. 43, 1061 – 1066. (doi:10.1007/BF00166926)

64. Luttik MAH, van Spanning R, Schipper D, van Dijken J, Pronk JT. 1997 The low biomass yields of the acetic acid bacterium Acetobacter pasteurianus are due to a low stoichiometry of respiration-coupled proton translocation. Appl. Environ. Microbiol. 63, 3345 – 3351.

65. Babel W, Müller RH, Markuske KD. 1983 Improvement of growth yield of yeast on glucose to the maximum by using an additional energy source. Arch. Microbiol. 136, 203 – 208. (doi:10.1007/BF00409045)

66. Verdun C, Stouthamer AH, Scheffers WA, van Dijken JP. 1991 A theoretical evaluation of growth yields of yeasts. Antonie van Leeuwenhoek 59, 49 – 63. (doi:10.1007/BF00582119)

67. Mostafa HM, Collins EB. 1968 Molar growth yields of certain lactic acid bacteria as influenced by autolysis. J. Bacteriol. 96, 117 – 125.

68. Cheng NG, Hasan M, Ling C, Tham M. 2009 Production of ethanol by fed-batch fermentation. Pertanika J. Sci. Technol. 17, 399 – 408.

69. Gómez JM, Cantero D. 1998 Kinetics of substrate consumption and product formation in closed acetic fermentation systems. Bioprocess Eng. 18, 439 – 444. (doi:10.1007/s004490050468)

70. Hines KE, Middendorf TR, Aldrich RW. 2014 Determination of parameter identifiability in nonlinear biophysical models: a Bayesian approach. J. General Physiol. 143, 401 – 416. (doi:10.1085/jgp.2013111162013111116[PML])

71. Hesseltine CW. 1992 Mixed-culture fermentations. In Applications of biotechnology to fermented foods: report of an Ad hoc panel of the Board on Science and Technology for International Development (ed. National Academies), ch. 6, pp. 52 – 57. Washington, DC: National Academies Press.

72. Price ND, Reed JL, Palsson BB. 2004 Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2, 886 – 897. (doi:10.1038/nrmicro1023)

73. Levy R, Borenstein E. 2013 Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc. Natl Acad. Sci. USA 110, 12 804 – 12 809. (doi:10.1073/pnas.1300926110)

74. Magnúsdóttir S et al. 2017 Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat. Biotechnol. 35, 81 – 89. (doi:10.1038/nbt.3703)

75. Ponomareva O et al. 2017 Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow. Cell Syst. 5, 345 – 357.e6. (doi:10.1016/j.cels.2017.09.002)

76. Moreno-Zambrano M, Grimbs S, Ullrich MS, Hütt M-F. 2018 Data from: A mathematical model of cocoa bean fermentation. Dryad Digital Repository. (doi:10.5061/dryad.321d33v)