JACOBS
UNIVERSITY

Mathematical Modeling of Cocoa Bean Fermentation

by

Mauricio Moreno-Zambrano

A Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Bioinformatics

Approved Dissertation Committee:

Prof. Dr. Marc-Thorsten Hütt (Chair)
Jacobs University Bremen

Prof. Dr. Matthias Ullrich (Internal)
Jacobs University Bremen

Dr. Sergio Grimbs (External)
Bayer AG

Date of Defense: 7 June 2021

Department of Life Sciences & Chemistry

JACOBS UNIVERSITY

Statutory Declaration

Family Name, Given/First Name	Moreno-Zambrano, Mauricio
Matriculation number	20331746
Kind of thesis	PhD Thesis

English: Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparison with future theses. This does not grant any rights of reproduction and usage, however.

This Thesis has been written independently and has not been submitted at any other university for the conferral of a PhD degree; neither has the thesis been previously published in full.

German: Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erstellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, sind als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der Thesis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbank eingegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auch verbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde in der vorliegenden Form weder einer anderen Prüfungsbehörde vorgelegt noch wurde das Gesamtdokument bisher veröffentlicht.

04.05.2021

Date, Signature

Abstract

The process of cocoa bean fermentation is the key step in chocolate manufacturing. During its conduction, aroma and flavor compounds are produced because of a complex series of reactions which will determine important characteristics of the final product. Historically, this process has been conducted in a spontaneous manner often dictated by local practices in each of the cocoa bean producing sites. Hence, there is a wide variety of techniques in its implementation, e.g., use of heaps, platforms and wooden boxes, that could vary even between locations within the same country. Added to different techniques employed, microbial diversity and use of several cocoa cultivars contribute to its complexity.

Regardless of its importance, cocoa bean fermentation has been scarcely studied beyond qualitative approaches, which currently make up the vast majority of sources for its understanding. Thus, what we know about its mechanistic functioning is based on experimental observations that had led to state several hypotheses of which few have been directly measured and proven. Among these, the sequential dominance of microbial groups during its conduction, i.e., yeast, lactic acid bacteria and acetic acid bacteria, represents the backbone of the process over which other regulatory microbial-metabolite interactions have been proposed and sometimes accepted indirectly.

In this thesis, a series of mathematical models are presented to bring a quantitative exploration of the mechanistic features of the process with the use of Ordinary Differential Equation Systems. In this sense, several sources of data have been used in order to fit these mathematical conceptualizations by a Bayesian framework to solve the parameter estimation problem. A baseline model is proposed, assessed and discussed in terms of its biological plausibility, and an interpretation of its got parameter estimates is introduced as an indirect indicative of differences between trials' features.

Therefore, I present a deeper analysis of model iterations based on the baseline with the purpose of accomplishing a wider exploration of five hypothesized mechanisms, e.g., over-oxidation of acetic acid and consumption of fructose by lactic acid bacteria. In that way, their likeliness of occurrence is determined by their overall success on fitting several data gathered from 23 different fermentation trials. Also, an analysis of obtained parameter estimates as classifying fermentation features is discussed.

Finally, the effect of temperature on kinetic modeling of cocoa bean fermentation is addressed by the use of Arrhenius terms and discussed in terms of gains in interpretation and biological plausibility of the parameter estimates and its effect on model accuracy.

The findings in this thesis provide new insights into the understanding of the complex process of cocoa bean fermentation by assessing candidate mechanisms, and interpreting parameter estimates from a biological point of view towards their use as an addition to chemical fingerprinting methods for classifying features.

Acknowledgements

This thesis summarizes countless hours of work that would not have been possible without the support of many people. First, I would like wholeheartedly thank my supervisor, Prof. Dr. Marc-Thorsten Hütt, for the opportunity to conduct my research under his advice and guidance. From our many conversations, I found not only ways to solve obstacles in this journey, but a deep passion for research aligned with ethics and respect for Science that surely will accompany me in further steps during my career development.

Second, I am also grateful to my co-supervisor, Prof. Dr. Matthias S. Ullrich, for all his patience and encouraging words about my work and being the head of the COMETA project; to which, under its umbrella, this work makes a part of. Hence, I also want to show my gratitude to Barry Callebaut for funding my research.

I would like to thank Dr. Sergio Grimbs, third member of my thesis committee, and Dr. Anne Grimbs, whom together were of great help during the first months of my Doctoral studies not only in an Academic aspect, but personal too. Special thanks go to Dr. Piotr Nyczka, Sarah Islam and Rossie Hossaim, I will never forget all our conversations around a glass of beer or a cup of coffee.

I am also beholden with friends, which regardless of any physical distance, they were there to cheer me up with their warm words, jokes and unconditional friendship. With no particular order, my sincerest gratitude to Ricardo Almeida, David Ron, Puki Martinez, Misha Alencastro, Elisabeth Deutschmann and Steffen Jaeckel.

Finally, I dedicate this work to the most important persons in my life which always have been there to lift me up when I was about to fall. To my Mom, for all courage, braveness, principles and humbleness I learned from her. Because she taught me to be kind, to persevere and act honestly. To my Dad's memory, which I will always keep alive with my correct actions and endeavors as he taught me before departing this world. And, to Silvana, the love of my life, for her selfless love, for her hand always ready to take mine and walk together, for being there to calm my tears, for being my partner, my friend and my soulmate. These times in my life would have meant nothing if you were not there.

Abbreviations

AAB	acetic acid bacteria.
Ac	acetic acid.
bulk-ESS	bulk effective sample size.
CFU	colony forming units.
CI	credible interval.
ESR	expected success rate.
ESS	effective sample size.
EtOH	ethanol.
FLAB	fructophilic lactic acid bacteria.
Fru	fructose.
Glc	glucose.
HMC	Hamiltonian Monte Carlo.
LA	lactic acid.
LAB	lactic acid bacteria.
M	monosaccharides.
M1	mechanism 1.
M2	mechanism 2.
M3	mechanism 3.
M4	mechanism 4.
M5	mechanism 5.
MCMC	Markov chain Monte Carlo.
MCSE	Monte Carlo standard error.
MI	model iteration.
NUTS	No-U-Turn sampler.
ODE	ordinary differential equation.
OSR	observed success rate.
PC1	principal component 1.
PC2	principal component 2.
PCA	principal component analysis.
PM1	pre-model 1.
PM2	pre-model 2.
PM3	pre-model 3.
PSIS-LOO	Pareto-smoothed importance sampling leave-one-out cross validation.
Suc	sucrose.
tail-ESS	tail effective sample size.
WAIC	widely applicable information criterion.
Y	yeast.

List of symbols

D_M	Malahanobis distance.
E_1	growth energy of activation.
E_2	mortality energy of activation.
$K_{\mathrm{Ac}}^{\mathrm{AAB}}$	substrate saturation constant of acetic acid bacteria growth on acetic acid.
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	substrate saturation constant of acetic acid bacteria growth on ethanol.
$K_{\mathrm{Fru}}^{\mathrm{LAB}}$	substrate saturation constant of lactic acid bacteria growth on fructose.
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	substrate saturation constant of yeast growth on fructose.
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	substrate saturation constant of lactic acid bacteria growth on glucose.
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	substrate saturation constant of yeast growth on glucose.
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	substrate saturation constant of acetic acid bacteria growth on lactic acid.
$K_{\mathrm{LA}}^{\mathrm{Y}}$	substrate saturation constant of yeast growth on lactic acid.
Q_L	heat loss.
T_e	environmental temperature.
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	acetic acid bacteria-to-acetic acid from ethanol yield coefficient.
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Fru}}$	lactic acid bacteria-to-acetic acid from fructose yield coefficient.
$Y_{\mathrm{Ac\|Y}}^{\mathrm{Fru}}$	yeast-to-acetic acid from fructose yield coefficient.
$Y_{\mathrm{EtOH\|LAB}}^{\mathrm{Fru}}$	lactic acid bacteria-to-ethanol from fructose yield coefficient.
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	yeast-to-ethanol from fructose yield coefficient.
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Fru}}$	lactic acid bacteria-to-lactic acid from fructose yield coefficient.
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	lactic acid bacteria-to-acetic acid from glucose yield coefficient.
$Y_{\mathrm{Ac\|Y}}^{\mathrm{Glc}}$	yeast-to-acetic acid from glucose yield coefficient.
$Y_{\mathrm{EtOH\|LAB}}^{\mathrm{Glc}}$	lactic acid bacteria-to-ethanol from glucose yield coefficient.
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	yeast-to-ethanol from glucose yield coefficient.
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	lactic acid bacteria-to-lactic acid from glucose yield coefficient.
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	acetic acid bacteria-to-acetic acid from lactic acid yield coefficient.
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{LA}}$	yeast-to-ethanol from lactic acid yield coefficient.
$Y_{Q\|\mathrm{EtOH}}$	ethanol-to-heat yield coefficient.
$Y_{Q\|\mathrm{Fru}}$	fructose-to-heat yield coefficient.
$Y_{Q\|\mathrm{Glc}}$	glucose-to-heat yield coefficient.

$Y_{Q\|\mathrm{LA}}$	lactic acid-to-heat yield coefficient.
$Y_{\mathrm{Ac}\|\mathrm{AAB}}$	acetic acid bacteria-to-acetic acid yield coefficient.
$Y_{\mathrm{EtOH}\|\mathrm{AAB}}$	acetic acid bacteria-to-ethanol yield coefficient.
$Y_{\mathrm{Fru}\|\mathrm{LAB}}$	lactic acid bacteria-to-fructose yield coefficient.
$Y_{\mathrm{Fru}\|\mathrm{Y}}$	yeast-to-fructose yield coefficient.
$Y_{\mathrm{Glc}\|\mathrm{LAB}}$	lactic acid bacteria-to-glucose yield coefficient.
$Y_{\mathrm{Glc}\|\mathrm{Y}}$	yeast-to-glucose yield coefficient.
$Y_{\mathrm{LA}\|\mathrm{AAB}}$	acetic acid bacteria-to-lactic acid yield coefficient.
$Y_{\mathrm{LA}\|\mathrm{Y}}$	yeast-to-lactic acid yield coefficient.
$\mu_{\max}^{\mathrm{AAB}_{\mathrm{Ac}}}$	maximum specific growth rate of acetic acid bacteria on acetic acid.
$\mu_{\max}^{\mathrm{AAB}_{\mathrm{EtOH}}}$	maximum specific growth rate of acetic acid bacteria on ethanol.
$\mu_{\max}^{\mathrm{AAB}_{\mathrm{LA}}}$	maximum specific growth rate of acetic acid bacteria on lactic acid.
$\mu_{\max}^{\mathrm{LAB}_{\mathrm{Fru}}}$	maximum specific growth rate of lactic acid bacteria on fructose.
$\mu_{\max}^{\mathrm{LAB}_{\mathrm{Glc}}}$	maximum specific growth rate of lactic acid bacteria on glucose.
$\mu_{\max}^{\mathrm{Y}_{\mathrm{Fru}}}$	maximum specific growth rate of yeast on fructose.
$\mu_{\max}^{\mathrm{Y}_{\mathrm{Glc}}}$	maximum specific growth rate of yeast on glucose.
$\mu_{\max}^{\mathrm{Y}_{\mathrm{LA}}}$	maximum specific growth rate of yeast on lactic acid.
σ	total standard deviation.
$\tilde{D}_M$	median of Malahanobis distances.
b_{Ac}	decay rate constant of acetic acid.
b_{EtOH}	decay rate constant of ethanol.
b_{LA}	decay rate constant of lactic acid.
d_1	decay rate of ethanol.
d_2	decay rate of lactic acid.
d_3	decay rate of acetic acid.
k_{AAB}	mortality rate constant of acetic acid bacteria.
k_{LAB}	mortality rate constant of lactic acid bacteria.
k_{Y}	mortality rate constant of yeast.
v_1	growth rate of yeast on glucose.
v_2	growth rate of yeast on fructose.
v_3	growth rate of lactic acid bacteria on glucose.
v_4	growth rate of acetic acid bacteria on ethanol.
v_5	growth rate of acetic acid bacteria on lactic acid.
v_6	mortality rate of yeast.
v_7	mortality rate of lactic acid bacteria.
v_8	mortality rate of acetic acid bacteria.
v_9	growth rate of lactic acid bateria on fructose.

v_{10} growth rate of yeast on lactic acid.

v_{11} growth rate of acetic acid bateria on acetic acid.

Table of Contents

Chapter 1

Introduction[1]

Beans from *Theobroma cacao* L. are the main raw material in chocolate production. Among the distinct steps in the manufacturing process of chocolate, cocoa bean fermentation is recognized as one of its main sub-processes. Besides facilitating the removal of mucilage, the microbial activity that occurs during the fermentation handles the death of the embryo. Furthermore, fermentation triggers a series of biochemical reactions which drives the development of characteristic chocolate aroma and flavor compounds [3, 4]. The typical procedure is a spontaneous fermentation in which the microorganisms are inoculated during the harvesting and fermentation itself.

Although the importance of its role in chocolate production, the fermentation procedure is highly non-standardized and varies across the countries of origin. In this sense, it has been reported that fermentation is performed in heaps (e.g., Ghana and Ivory Coast), boxes (e.g., Brazil, Indonesia and Malaysia), baskets (e.g., Nigeria and Ghana), trays (e.g.,Ghana), sacks and using platforms (e.g., Ecuador) [5, 6]. Apart from the variety of fermentation methods, one should consider the fact that different cocoa hybrids and species are used in each of the producing countries. This directly influences its duration, ranging from 2 to 3 days for Criollo varieties and 5 to 8 days for Forastero varieties [3].

Surprisingly, the microbial dynamics of cocoa bean fermentation processes under heterogeneous practices share some common aspects; such as the succession of dominant microorganisms, metabolite kinetics, temperature and pH profiles [4, 6]. On these foundations, quantitative approaches to describe the process beyond its general qualitative characteristics and affects on the final product could serve for a better understanding of its mechanistic properties towards improvements on cocoa processing methods.

Thus, in this first Chapter, basic concepts regarding physiology of *Theobroma cacao* L. are presented together with a comprehensive description of the fermentation process, mathematical basis for its modeling and an introduction of Bayesian inference on which parameter estimation of such models relies on.

1.1 The cocoa fruit

The *Theobroma cacao* L. is originally from South and Central America and has been cultivated for over 4000 years [7]. The cocoa fruits or "pods" vary in size and color depending on its variety and, thus, its origin. Concerning variety, cocoa can be classified as Criollo (from Central and South America) and Forastero (from the Amazon region). Criollo varieties are known for being the raw material for fine chocolate production; however, nowadays it is disappearing due

[1] Section 1.5.3 of this chapter is partially based on Sections 2.4.1 of the publications of Moreno-Zambrano *et al.* [1, 2].

to its low resistance to disease and low productivity. On the other hand, Forastero varieties have more resistance to diseases and higher productivity than Criollo [3].

For these reasons, several clones are used in each of the producing countries which are sometimes the product of crosses between both varieties in order to keep the quality characteristics of Criollo and the resistance to disease and rapid production of Forastero. These hybrids between Criollo and Forastero are known as Trinitario [3].

The cultivation of cocoa is restricted to regions where the temperature is constantly above 16 °C, with high humidity as well as even light-dark periods throughout the year [7]. Cocoa trees flourish seasonally and fertilized flowers bear fruit 170 days after pollination. During this period the fruit grows to maturity and changes color from green (or dark red-purple) to yellow, orange or red depending on the variety. Mature fruits have 30 to 40 beans, each of which is enclosed in a white mucilaginous pulp and attached to a placenta [8].

Only the beans are used in the production of chocolate; however, the pulp in which most microbial activity takes place during the fermentation process is necessary for developing the desired characteristics in terms of flavor and aroma of the final product. Both pulp and bean are rich mixes of compounds that will play important roles during the fermentation process. The former is mostly composed of water (82-87%), sugars (10-13%), pentosans (2-3%), citric acid (1-2%) and salts (8-10%). On the contrary, bean composition is more complex where water is the major constituent (32-39%) followed by fat (30-32%), proteins (8-10%), starch and pentosans (4-6%), polyphenols (5-6%), sucrose jointly with theobromine and cellulose (2-3%), and acids together with caffeine (1%) [3].

1.2 Fermentation of cocoa beans

Raw cocoa beans show bitter and astringent flavors. The bitterness is due to methylaxantines (e.g., theobromine and caffeine) in the cotyledons; while the astringency is caused by the presence of polyphenols and tannins (e.g., flavonoids, leucocyanidins, catechins and anthocyanidins) [9]. Consequently, raw beans are not used for manufacturing chocolate as they lack the compounds needed to develop the characteristics organoleptic properties of chocolate.

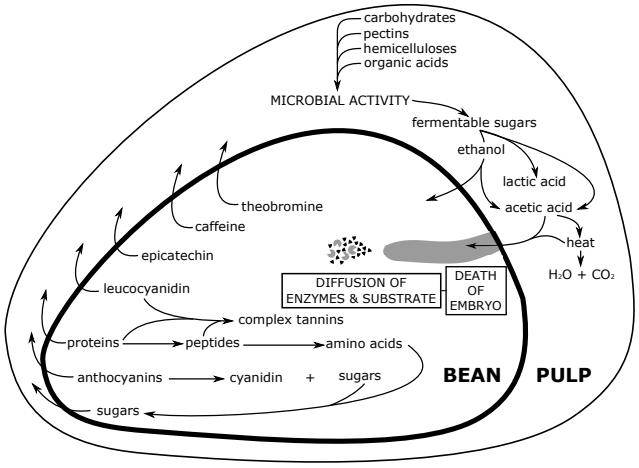


Figure 1.1: Overview of the biochemical modifications in the pulp and bean during fermentation of cocoa (taken from Lopez (1995) [3])

In this sense, a microbial driven fermentation process is needed in order to trigger the reactions that convert inner metabolites of raw cocoa beans in those that accomplish desirable chocolate characteristics (Figure 1.1). The interior of the pod is supposed to be micro-biologically

sterile [3]. However, it also has been reported that a few hundred microorganisms per gram were found within, particularly yeasts [10, 11]. Nevertheless, at latest when the pod is opened, it gets inoculated with microorganisms coming from the pod surface, workers' hands and tools [3].

The fermentation process in cocoa can be understood as an overlapping succession of microbial activities. In a first stage, microbial activity catabolizes carbohydrates (e.g., sucrose (Suc)), pectins, hemicelluloses and organic acids into fermentable sugars (e.g., glucose (Glc) and fructose (Fru)). These fermentable sugars are subsequently converted to ethanol (EtOH). During the second stage, lactic acid (LA) is produced from the remaining sugars. Finally, acetic acid (Ac) is produced from ethanol through exothermic reactions that jointly will produce the death of the embryo within the cotyledon approximately by the third day of the fermentation [8]. With the embryo's death, the diffusion of enzymes and substrates involved in ensuing biochemical processes diffuse into the bean. The details of each phase in the cocoa fermentation are provided in the following section.

1.3 Microbial succession and biochemistry during cocoa fermentation

The cocoa fermentation is characterized by a succession of microbial activities which refer to three major groups driving the whole process, being dominated in the early stage by yeast (Y), subsequently surpassed by lactic acid bacteria (LAB) and after the decline of these first two groups, acetic acid bacteria (AAB) takes over [12–15].

The diversity of microorganism species in each of these three major groups is large and differs, as reported in literature [12–15], across the countries of origin and the applied fermentation methods. However, it has been shown that in overview, for the case of yeast (Y), predominantly the genera *Saccharomyces*, *Hanseniaspora* and *Pichia* occur. Among the lactic acid bacteria (LAB) mainly species from the genus *Lactobacillus* have been reported, in particular *L. fermentum* and *L. plantarum*. For the third group of major microorganisms, the acetic acid bacteria (AAB), the dominating genus is *Acetobacter*, especially *A. pasteurianus* in most of the cases [4].

Under optimal conditions, this three-phased process lasts on average between 4 to 5 days producing brown-colored beans. In case of a short fermentation, the resulting beans are purple and characterized by its bitterness and astringency. When over-fermentation happens, the process results in black beans with a putrid or dull smell because of short-chain fatty acids produced by action of *Bacillus* spp. and filamentous fungi [4]. From the physico-chemical point of view, the fermentation process could be divided into two phases called *anaerobic hydrolytic phase* and *aerobic oxidative phase*, which occur concomitantly with the microbial overlapping phases previously mentioned[6]. For this research, only the phases involving microbial succession will be described in the following sub-sections.

1.3.1 Phase 1: Anaerobic growth of yeast

The anaerobic growth of yeast usually occurs during the first 24 to 48 hours of cocoa bean fermentation, where yeasts' growth ranges from initial 2-7 log colony forming units (CFU) to a maximum 6-9 log CFU per gram of cocoa. Saccharides (e.g., sucrose, glucose and fructose) are converted into ethanol and other sub-products (e.g., acetic and succinic acid [16]) under anaerobic conditions and pH below 4.0, with a preference for glucose. This production of ethanol induces an increase in temperature from 25-30 °C to 35-48 °C as producing the collapse of parenchyma cells that facilitate the entrance of air within the fermenting mass. This determines the initiation of a micro-aerobic environment that enhances the growth of LAB and AAB,

where ethanol concentration drops because of oxidization into acetic acid by AAB or other aerobically growing yeasts, diffusion into the bean cotyledons, and sweating and evaporation. Once concentrations of lactic acid and acetic acid have increased, and the temperature exceeds 45 °C due to the rising activity of LAB and AAB, yeasts communities start to disappear [6].

1.3.2 Phase 2: Microaerobic growth of lactic acid bacteria

The second phase of cocoa bean fermentation takes part between 24 to 42 hours and is dominated by lactic acid bacteria (LAB), which are present in low concentrations from the beginning and possibly interacting with yeasts [6]. The growth of LAB ranges from initial 3-5 log CFU to maximum 7-9 log CFU per gram of cocoa. LAB convert saccharides into lactic acid (mainly), acetic acid, ethanol and even some organic acids (e.g., citric and malic acid) into lactic acid or into acetic acid. The preferred monosaccharide by LAB is glucose. However, some LAB are fructophilic, preferring fructose over glucose [12, 17]. As a consequence of the assimilation of citric acid by LAB, pH increases in the fermenting mass from 3.5-4.0 to 4.2-5.0 allowing other bacteria to grow. Finally, towards the end of the fermentation process, the population of LAB declines slightly because of the depletion of carbon sources, high concentrations of ethanol and high temperatures [18].

1.3.3 Phase 3: Aerobic growth of acetic acid bacteria

The third phase of cocoa bean fermentation happens between 48 and 112 hours, which is characterized by aerobic AAB's growth and is when the embryo's death occurs. During this phase, AAB persist until the conditions for its growth are more favorable once pulp is metabolized and drained, leading to an aeration increase in the fermenting mass and the temperature rises above 37 °C. AAB grow from 3-5 log CFU to maximum 5-8 log CFU per gram of cocoa [11, 17]. The principal activity of AAB is the oxidation of ethanol, previously produced by yeasts, to acetic acid. This oxidation process is an exothermic reaction that produces an increase of the temperature of the fermenting mass up to 45-50 °C [6]. Once ethanol decreases, the concentration of lactic acid decreases as well. This decrease is happening because of a simultaneous oxidation of ethanol and lactic acid into acetic acid by AAB that occurs when the concentration of ethanol overpasses a certain threshold where its availability to AAB is reduced[19]. After acetic acid reaches its maximum peak, a decrease in its concentration is seen often, and can be due to the effect of evaporation because of the increase of the temperature of the fermenting mass [12, 17, 20] or even to the over-oxidation of acetic acid into carbon dioxide and water conducted by some AAB species [21]. Finally, the death of AAB results from the exhaustion of ethanol and an increase in the temperature of the fermenting mass.

1.4 Modeling cocoa bean fermentation

As I have introduced it in previous sections, the cocoa bean fermentation process comprises a complex series of interactions between microbial communities and metabolites confined within a physical environment. All these characteristics depict a glance at the complexity of this biological system which it has been scarcely subject to mathematical modeling. Before this current work, they have reported only two approaches in peer-reviewed journals.

On the one hand, Kresnowati *et al.* [22] proposed two models different between each other by the considered compartments in which the system takes place (i.e., pulp only and pulp-inner bean). In both cases, microbial dynamics could not be simulated properly and they fairly resembled metabolite kinetics. On the other hand, López-Pérez *et al.* [23] developed a more complex single compartment model and compared its implementation between either the use of

genetic algorithms or a Levenberg-Marquardt optimization routine. However, this work focuses more on comparing methods for accomplishing the solution of the parameter estimation problem rather than in their biological interpretation.

In this sense, this research represents a first approach where both, modeling and interpretation of obtained parameter estimates is performed focusing majorly on their biological plausibility (see Chapter 2) and their use in gaining more understanding of the mechanistic characteristics of the process (see Chapters 3 and 4).

1.5 Underlying concepts and methods in developing cocoa bean fermentation models

For a better understanding of the upcoming chapters, here I present some of the most important concepts and methods used during the whole conduction of the research: (1) description of the modeling process, (2) mathematical foundations to consider for conceptualizing cocoa bean fermentation, (3) formulation of the Bayesian framework used to obtain models' parameter estimates, (4) model comparison and evaluation, and (5) an example of fitting a toy model by Bayesian means. Methods only applied in each of the following chapters are described in their respective method sections.

1.5.1 The modeling process

A model can be defined as the simplified representation of a specific system by conceptual or quantitative means. Thus, developing a mathematical model requires an execution of a workflow fed with as much knowledge of the system is available [24, 25]. In general terms, such a workflow consists of three main groups of tasks, namely *formulation*, *calibration*, and *analysis and evaluation* of a candidate model (for a complete description of this workflow, see [25]). Within *formulation* of a model, important attention is put over the conceptualization of the system and the fundamental mathematical concepts behind the construction of its more complex depiction capable of embracing more interactions between the elements within it. Finally, a challenging task in any model construction is given by the technique to be used to estimate (within *calibration* of the model group of tasks) the parameters that make possible simulations of all regulatory mechanisms considered in previous steps.

1.5.2 Conceptualization

Sketching a conceptual model (usually as a network diagram [26]) is not a simple one. It requires special consideration as it represents the basis for the subsequent mathematical assumptions of the studied system, relying on knowledge about it. Consequently, it allows the modeler to change it as many times is needed in order to get a plausible mathematical representation.

As it will be described in Chapters 2 and 3, interplay of elements of the cocoa bean fermentation system is firstly conceptualized as network diagrams where microbial growth and mortality rates, jointly with metabolite kinetics, drive its temporal dynamics. Hence, in the following paragraphs, the bases used in this work to accomplish the translation of conceptual models to mathematical expressions are described.

Cocoa bean fermentation is a batch process

Cocoa bean fermentation can be considered a batch process, meaning that there is no continuous feeding of nutrients into the system aiming to keep the microbial populations in either their exponential or stationary phases. Moreover, as opposed to typical growth of microorganisms,

as depicted in Figure 1.2, where four phases are easily recognized (i.e., lag, exponential growth, stationary and mortality phases) experimental conduction of the cocoa bean fermentation process have shown a partial lack of lag phase (e.g., only for AAB a lag phase is noticeable) and an almost complete absence of a stationary phase for all microbial populations involved, or at least the time between each measurement does not have a sufficient short resolution evidencing them [1].

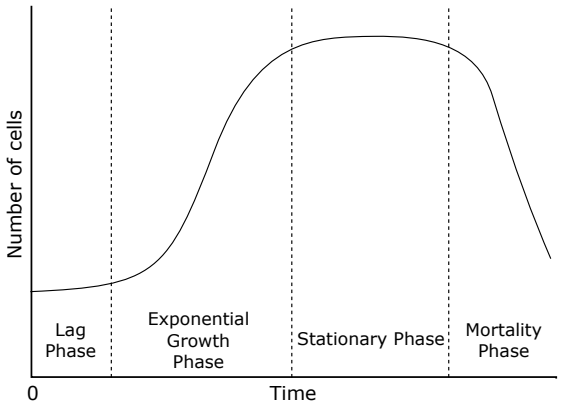


Figure 1.2: Microbial growth phases. The lag phase is a period of time where a microorganism adapts itself to the growing conditions. During the exponential growth phase, microbial cells start to divide continuously at a given rate constant. Stationary phase is characterized by a cells' net growth nearly equal to zero. Mortality phase describes the decline in the microbial population due to environmental factors, mainly lack of nutrients.

Thus, for this specific case, microbial growth can be expressed as a combination of their change on population over time between the exponential growth and mortality phases. In this sense, an ordinary differential equation (ODE) can describe these changes as

$$\frac{dN}{dt} = v_g - v_m \,, \tag{1.1}$$

where N is the population size of a given microbial community and v_g and v_m are equations describing the growth and mortality rates of N, respectively.

Specifically, the growth rate equation v_g for the exponential phase can be expressed as

$$v_g = \mu\, N \,, \tag{1.2}$$

where μ represents the specific growth rate of a certain microorganism N. Similarly, for the mortality phase, the mortality rate equation v_m can take the form

$$v_m = k\, N \,, \tag{1.3}$$

where k is the mortality rate constant of population N.

Microbial growth rate equation

An important aspect for modeling a batch process is to constraint microbial growth by the abundance of nutrients. Among different models that consider this, the Monod [27] equation is frequently used because it assumes the existence of a growth rate-limiting nutrient that, if it is present in saturating conditions, will not produce an increase in the maximal growth rate. In other words, the excess of a growth rate-limiting nutrient does not imply more biomass production [28]. Therefore, it will serve as a starting point for modeling microbial growth

in the process of cocoa bean fermentation. From the mathematical perspective, it resembles the Michaelis-Menten equation, with the difference that Monod equation is based on empirical observations rather than on theoretical considerations and is given by

$$\mu(s) = \mu_{\max} \frac{s}{s + K_s}\,, \tag{1.4}$$

where $\mu(s)$ is the specific microbial growth rate, $\mu_{\max}$ is the maximum specific growth rate, s is the substrate (rate-limiting nutrient) and K_s is the substrate saturation constant.

This formalization allows Eq. (1.2) to be expressed as

$$v_g = \mu_{\max} \frac{s}{s + K_s} N\,. \tag{1.5}$$

Furthermore, the Contois [29] equation is considered here for representing microbial growth of LA producing AAB species under the assumption that their growth is a function of their population size (see Chapter 2). Such an approach is implemented under the same scheme above described.

Microbial mortality rate equation

Contrary to microbial growth, the mortality rate equation does not have to depend on a substrate concentration. In principle, the constant mortality rate in Eq. (1.3) can take several forms (e.g., logistic, exponential or other algebraic function [30]). In the simplest case, the mortality rate equation can be given as a linear relationship of the microbial population with a constant mortality rate.

Such an approximation, within biochemical kinetics, can be got by an application of the law of mass action which states that the rate of a chemical reaction is proportional to the probability of a collision of the reactants [31]. Thus, the rate is proportional to the product of the reactants' concentration to the power of their molecularity. In this sense, consider the following chemical reaction:

$$\mathrm{S}_1 + \mathrm{S}_2 \underset{v_-}{\overset{v_+}{\rightleftharpoons}} 2\,\mathrm{P}\,, \tag{1.6}$$

where the reaction rate (v) can be expressed as the difference between the rate of the forward reaction (v_+) and the rate of the backward reaction (v_-) as

$$v = v_+ - v_-\,. \tag{1.7}$$

Then, applying the law of mass action, we get that the reaction rate is equal to

$$v = k_+ \mathrm{S}_1\,\mathrm{S}_2 - k_- \mathrm{P}^2\,, \tag{1.8}$$

where k_+ and k_- are the respective proportionality factors, or so-called kinetic constants.

With this in mind, the application of a mass action kinetics behavior to the mortality rate in cocoa bean fermentation could be formulated through the kinetics of a simple decay for microbial population x as

$$\mathrm{x} \longrightarrow \tag{1.9}$$

where its mortality rate equation, v_d, will be described as the product between a microbial population, N, with a mortality rate constant k for N as

$$v_m = k_N\, N \tag{1.10}$$

Under this same logic, further mortality rate equations in this research are planned under the use of the mass action law as interpreted by Chick-Watson equations [32] used in disinfection kinetics of microbial populations where non-linear decaying dynamics are produced by second- and third-order reactions between the former and chemical compounds in the medium (e.g., EtOH).

Metabolite kinetics

Another assumption that can be made on the Monod and Contois equations is that the consumed nutrients are immediately used to increase population. This assumption is helpful to represent the consumption of a substrate, s, for incrementing the microbial population or biomass, N, and is expressed by a linear proportionality of the specific growth rate μ (from Eq. (1.4)) as a substrate uptake rate, ρ, of the form

$$\rho\,(s) = -\frac{1}{Y_{N|s}}\,\mu\,(s)\ , \tag{1.11}$$

where $Y_{N|s}$ is the biomass-to-substrate yield coefficient, and it is understood as the relationship between the quantities of materials consumed and produced during a particular reaction [33].

In terms of microbial growth, it can be defined as the ratio between the change of biomass (ΔN) and the change of substrate concentration (Δs) as

$$Y_{N|s} = -\frac{\Delta N}{\Delta s}\,, \tag{1.12}$$

where its negative sign shows that an increase in biomass leads to a decline in substrate concentration.

In a similar fashion, a biomass-to-product yield coefficient, $Y_{p|N}$, can be defined as the ratio of the change of product (Δp) with the change of biomass (ΔN) as

$$Y_{p|N} = \frac{\Delta p}{\Delta N}\,, \tag{1.13}$$

where its positive sign indicates that an increase in biomass produces an increase in the product's concentration.

With this in mind, a metabolite production rate, $\gamma(s)$, can be expressed similarly to $\mu(s)$

$$\gamma\,(s) = Y_{p|N}\,\mu\,(s)\ . \tag{1.14}$$

These ratios, as shown in Eqs. (1.12) and (1.13), are commonly known as yield coefficients and serve as links between microbial dynamics and metabolite kinetics, as it will be briefly introduced in the following paragraphs.

Assembling a toy model

As an example of integrating above defined growth and mortality rate equations, consider the minimal model of Y's growth depicted in Figure 1.3.

In this small scenario, an ODE system of 3 state variables, namely monosaccharides (M), EtOH and Y, would describe the network. First, an ODE describing Y's dynamics will be simply represented similarly to Eq. (1.1) as

$$\frac{d[\mathrm{Y}]}{dt} = v_1 - v_2\,, \tag{1.15}$$

where v_1 and v_2 are growth and mortality rate equations as defined in Eqs. (1.5) and (1.10), respectively.

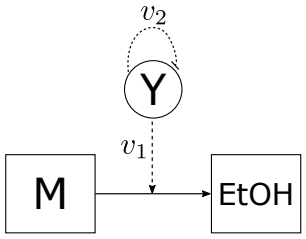


Figure 1.3: Yeast's growth network diagram. Yeast (Y) is represented as a circle. Metabolites monosaccharides (M) and ethanol (EtOH) are represented as squares. Yeast's growth rate equation v_1 is represented as a normal dashed arrow. Yeast's mortality rate equation v_2 is represented as a blunt dashed arrow. The solid arrow represents the direction in which the conversion of M occurs into EtOH.

Now, to describe the consumption of M by Y, we can make use of Eq. (1.12) expressing it as

$$Y_{\mathrm{Y|M}} = -\frac{\frac{d[\mathrm{Y}]}{dt}}{\frac{d[\mathrm{M}]}{dt}}\,, \tag{1.16}$$

where by replacing $\frac{d[\mathrm{Y}]}{dt}$, we get

$$\frac{d[\mathrm{M}]}{dt} = -\frac{1}{Y_{\mathrm{Y|M}}}\left(v_1 - v_2\right). \tag{1.17}$$

Since death cells cannot convert any substrate into products, v_2 in Eq. (1.17) simplifies to zero. Thus, the ODE describing M consumed by Y becomes

$$\frac{d[\mathrm{M}]}{dt} = -\frac{1}{Y_{\mathrm{Y|M}}}\,v_1 = -Y_{\mathrm{M|Y}}\,v_1\,, \tag{1.18}$$

where $Y_{\mathrm{M|Y}}$ is known as the yeast-to-monosaccharides yield coefficient.

In a similar fashion, an ODE for production of EtOH can be assessed using Eq. (1.13) resulting in

$$\frac{d[\mathrm{EtOH}]}{dt} = Y_{\mathrm{EtOH|Y}}\,v_1\,. \tag{1.19}$$

1.5.3 Parameter estimation under a Bayesian framework

Once a conceptual model has been translated into a mathematical system with help of aforementioned concepts, it is necessary to work on solving the estimation problem regarding values for unknown parameters (e.g., yield coefficients, maximum specific growth rates, mortality rate constants and substrate saturation constants). There are diverse ways to estimate these parameter values, including fixing them to previously reported ones, conducting controlled experiments only aimed to determine such values through numerical optimization routines for models fitted over experimental data.

In the particular case of cocoa bean fermentation, as mentioned before, literature regarding estimation of kinetic parameters is scarce [22, 23] and few reported works as kinetic studies have not covered their estimation with mathematical bases comparable to the conceptualization here presented, but descriptive approaches [14, 34]. Besides, as opposed to controlled experiments, sampling rates in cocoa bean fermentation trials are low and suffer from considerable variation within measured time series [2]. Data with these characteristics are commonly found in other biological research areas too; in which, Bayesian methodologies have gained quite popularity in

recent years because of their proven advantages in dealing with short and noisy measurements [35].

Besides such operational conveniences, another asset of Bayesian inference in biological investigations is its flexibility in incorporating available information about a phenomenon subject of study by either previous observations of its conduction, or experts' knowledge. This is a natural consequence of the formulation of the Bayes' rule on the basis of conditional probabilities. Said this, let us consider the Bayes' rule:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}\,, \tag{1.20}$$

where the (posterior) probability of an event A given an event B occurs is equal to the joint probability[2] of A and B divided by the total probability of B ($P(B)$). The term $P(A)$, also known as prior probability, provides a way to combine prior beliefs about A into the likelihood ($P(B|A)$) in Eq. (1.20) and its further updating in case of arrival of new relevant information. In cases where no prior information is available, $P(A)$ can represent aforesaid scenarios by using weakly informative terms as priors.

From its simple formulation, the application of Bayes' rule in mathematical modeling during the last two decades has significantly advanced along with further development of computational power and statistical tools capable of handling solution of ODE systems as here presented.

Bayesian framework formulation

First, let us consider any ODE system (e.g., the toy model depicted in Eqs. (1.15), (1.18) and (1.19)). Such a model can be represented by the general form:

$$\frac{dx_i}{dt} = f\left(x, \theta\right)\,, \tag{1.21}$$

where x represents a vector the state variables, x_i is its ith component and the function $f\left(x, \theta\right)$ summarizes the dependence of the right-hand side of the ODEs on x and all k model parameters $[\theta_1, \theta_2, ..., \theta_k]$ contained in vector θ.

If we assume that parameters θ are selected such that a set of data $\mathcal{Y}$ is described, a way to infer them is to compute the (posterior) probability of θ given $\mathcal{Y}$, $P\left(\theta \mid \mathcal{Y}\right)$, which by rewriting Eq. (1.20), it is equal to:

$$P\left(\theta \mid \mathcal{Y}\right) = \frac{P\left(\mathcal{Y} \mid \theta\right)\; P\left(\theta\right)}{P\left(\mathcal{Y}\right)}\,. \tag{1.22}$$

Here, since $P\left(\mathcal{Y}\right)$ is a normalizing constant allowing the posterior density to integrate to one, Eq. (1.22) can be written in terms of the likelihood of observing $\mathcal{Y}$ given θ, $P\left(\mathcal{Y} \mid \theta\right)$, and the prior distribution of vector θ, $P\left(\theta\right)$, as:

$$P\left(\theta \mid \mathcal{Y}\right) \propto P\left(\mathcal{Y} \mid \theta\right)\; P\left(\theta\right)\,. \tag{1.23}$$

If we consider that each component of $\mathcal{Y}$ contains T time steps, with N state variables being observed, Eq. (1.23) takes the form of a product over all series and each of their measured points as:

$$P\left(\theta \mid \mathcal{Y}\right) \propto \prod_{i=1}^{N} \prod_{j=1}^{T} P(\mathcal{Y}_{i,j} \mid \theta)\, P\left(\theta\right)\,. \tag{1.24}$$

[2]Note that the joint probability $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$.

Finally, as our purpose is to identify values of θ that lead to a best agreement between $\mathcal{Y}_{i,j}$ in Eq. (1.24) and $x_i(j)$ in Eq. (1.21), we can consider $\mathcal{Y}_{i,j}$ to be sampled from a normal distribution whose mean is equal to the model's prediction $f(x, \theta)$, with a standard deviation term, σ, (caused by noise of any kind) allowing us to reformulate the total posterior distribution as:

$$P(\theta \mid \mathcal{Y}) \propto \prod_{i=1}^{N} \prod_{j=1}^{T} \mathcal{N}(f(x_{i,j}, \theta), \sigma)\ P(\theta). \tag{1.25}$$

By applying the total posterior distribution in Eq. (1.25), an extra parameter corresponding to a total standard deviation, σ, is also estimated besides the rest of parameters of a model as a system of ODEs. Hence, the general framework here presented will serve as the basis for solving the parameter estimation problem, as it will be stated in following chapters.

Setting prior probabilities

In Eq. (1.25), the definition of a suitable prior probability (or distribution), $P(\theta)$, is a crucial preparation step in determining the posterior probabilities of θ. As briefly mentioned before, an advantage of Bayes' rule is to take into consideration prior beliefs regarding the phenomenon subject of modeling. However, this characteristic has also been the foundation of an ongoing debate questioning validity of Bayesian methods that turn around prejudices of giving a modeler the freedom to set seemingly arbitrary prior beliefs after seeing data and thus, biasing estimates in despite of what could be in reality learned from them [36].

A common practice to avoid such criticisms is the use of either non-informative, or weakly informative priors. Regarding complete non-informative priors, e.g., uniform distributions, its implementation has been largely discouraged due to their capability on dominating the likelihood into regions of non-realistic values for the posterior distribution [36–38]. In the contrary, weakly informative priors are less prone to dominate over the likelihood, reason they have been more accepted in recent years [37].

With this in mind, a useful technique to define weakly informative priors is to assume the parameters of interest to be on unit scale [37]. In our context, this approach results beneficial in three ways: (1) prevents numerical issues when fitting a model by decreasing the impact of differences between magnitudes of the involved time series, (2) allows the use of independent priors for each parameter around the unit scale, and (3) imposes soft constraints over the magnitude of the computed posteriors. In practice, this aim is accomplished by scaling the data between 0 and 1 (see Section 2.2.4 for more details).

Bayesian computation

Obtention of the posterior distribution, as defined in Eq. (1.25), will involve the solution of the integrals defined by the likelihood and chosen prior distributions for θ in its right-hand side. Solving mathematical models expressed with the Bayes' rule is not always analytically feasible for high dimensional and multi parametrized scenarios and instead, Markov chain Monte Carlo (MCMC) algorithms capable of sampling from it are necessary to approximate their solution [38].

Thus, since the 70s with the formulation of the Metropolis-Hastings algorithm [39], a diverse group of MCMC algorithms have been developed and used along statistical software specialized in Bayesian inference. For the specific conduction of this thesis, the computation of posterior distributions of models' parameters and predictions is performed by means of the No-U-Turn sampler (NUTS) [40], which is an adaptive form of the Hamiltonian Monte Carlo (HMC) algorithm [41].

In general, MCMC methods create samples from a target distribution through a random walk in the parameter space where each new step, also called iteration, will depend on the previous one. Such methods have improved enormously the solution of a wide variety of problems; however, they are less efficient when facing complicated models with high dimensionality and many parameters because of inefficiency of the random walk in those circumstances [42]. In contrast to other MCMC algorithms, HMC replaces random walk behavior by introducing auxiliary *momentum* variables; which transform the problem of sampling from the target distribution into simulating Hamiltonian dynamics [40, 43]. This improvement allows HMC gradients, from Hamilton's equations, to guide the iterations of Markov chains through regions of high probability admitting an efficient exploration of the parameter space of the target distribution while reducing problems of correlation among them as found in samplers purely based on the random walk paradigm [43].

Despite the benefits of HMC over classical MCMC methods, HMC has the drawback of using two input parameters defined by the user. These are a step size ϵ and a number of steps L necessary to run a simulated Hamiltonian system. Setting both these parameters is not a straightforward procedure. Instead, multiple runs of the system combined with expertise of the user are often needed to tune them up. In this sense, the recent implementation of the NUTS by Hoffman & Gelman [40] has facilitated users to take advantage of HMC characteristics without the need of going through the cumbersome process of hand-tuning parameters. In brief, the way NUTS accomplish this is by adapting L with each iteration of the Markov chain and ϵ during the warm-up phase (see [40] for more detail).

1.5.4 Diagnostics, evaluation and comparison of Bayesian models

As with any other statistical methods, Bayesian models also account with some statistics to check whether their assumptions are valid as well as ways to determine their predictive power and accuracy. In a first stage for all models here presented, convergence diagnostics were performed followed by their evaluation and comparisons when necessary. Hence, brief descriptions of such statistics are presented in the following paragraphs.

Convergence diagnostics

Two important questions that arise when fitting Bayesian models that rely on MCMC methods are (1) did Markov chains mix well to convergence onto the target distribution?, and (2) is the sample size large enough to get a stable estimate of uncertainty?. Since these methods get a sample from the posterior, convergence improves as the number of draws approaches infinite. Nevertheless, in reality there is no guarantee on how convergence would behave beyond the finite number of iterations chosen by a user [44]. Among methods to check whether the algorithm as reached convergence, visual representations of Markov chains have been widely used to inspect it (e.g., trace plots, scatter plots and posterior draws plots). However, this approach reflects a qualitative nature that depends on judgment of the user rather than on specific metrics developed for such a purpose. Within the latter, there are two important statistics to look at when examining convergence.

On the one hand, potential scale reduction factor $\widehat{R}$ [38, 45, 46] stands out for its use in controlling for chain mixing. $\widehat{R}$ is defined as the division between the standard deviation of each scalar quantity of interest from all chains together, and the root mean square of the separate within-chain standard deviations [44].

In this way, $\widehat{R}$ measures whether a set of simulations for any scalar in the model has mixed properly. The logic behind it is that when the variance of all chains together is lower than the variance of them individually, $\widehat{R}$ takes values near to 1. As a rule of thumb, $\widehat{R}$ values lower than

1.1 are considered a sign of good mixing of chains, while values greater than 1.1 are considered otherwise.

On the other hand, as a result of performing MCMC methods, iterations over each chain tend to be correlated (even thought HMC reduces this effect). This determines that if the sample size used for running chains is not large enough, there is no way to be sure if the uncertainty (measured as the Monte Carlo standard error (MCSE)) of an estimated parameter is stable towards infinite.

A manner to measure this effect is through the effective sample size (ESS). ESS can be roughly defined as the number of independent draws from a Markov chain that contains the same amount of information as the dependent sample got by the used MCMC method [38, 44]. Hence, the closest the ESS is to the number of iterations used after warm-up, the better. As a rule of thumb, regardless of the total number of transitions set for Markov chains, ESS greater than 400 are considered as sufficient to get stable MCSE values.

Recently, Vehtari *et al.* [44] have proposed improved versions for $\widehat{R}$ and ESS based on overcoming their limitations when facing posterior distributions with heavy tails or varying variance across Markov chains. They suggest the use of new rules of thumbs for $\widehat{R}$ and ESS as less than 1.05 and greater than 100, respectively. In this thesis, in Chapter 2 convergence diagnostics are employed as originally proposed in [38, 45, 46]; while in Chapter 3, these are performed as lately proposed by Vehtari *et al.* [44].

Evaluation and comparison of models

After successfully fitting a model in terms of convergence, one remaining task is to investigate its predictive accuracy, which could be for the sake of model selection, comparison or averaging. In Bayesian inference, there is a manifold of procedures to accomplish these. In the particular case of this thesis, model evaluation and comparison have been done using mainly two approaches: (1) widely applicable information criterion (WAIC) [47] (see Appendix A), and (2) Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO) [48] (see Appendix A and Chapter 2).

These methodologies are similar between each other in the sense that both estimate out-of-sample predictive accuracy using within-sample fits. For this purpose, a measure of predictive accuracy is defined over the likelihood of a model. This measure, known as expected log point-wise predictive density (ELPD), will then be approximated with either WAIC or PSIS-LOO allowing to assess predictive accuracy of future observations of stand-alone models, or their comparison between them (for a detailed description of these methodologies, see [48]).

1.5.5 Fitting a toy model under a Bayesian framework

The toy model defined in Eqs. (1.15), (1.17) and (1.19) represents a simple mathematical formulation using the Monod equation in modeling microbial dynamics. As an example of its application, let us consider data reported by Petrov [49] regarding growth of *Saccharomyces cerevisiae* in batch culture; where they recorded 12 observations along 12 hours of cultivation for biomass of yeast (Y), glucose (Glc) and ethanol (EtOH) (all in $\mathrm{g\,L^{-1}}$). Given the conditions under which they carried the experiment out, mortality of *S. cerevisiae* cannot be entirely distinguishable (measuring biomass combines dead and living cells). This allows to simplify the model even more as:

$$\frac{d[\mathrm{Y}]}{dt} = \frac{\mu_{\max}\,[\mathrm{Glc}]}{[\mathrm{Glc}] + K_s}\,[\mathrm{Y}]$$

$$\frac{d[\text{Glc}]}{dt} = -Y_{\text{Glc|Y}} \frac{\mu_{\max}\,[\text{Glc}]}{[\text{Glc}] + K_s}\,[\text{Y}]$$
$$\frac{d[\text{EtOH}]}{dt} = Y_{\text{EtOH|Y}} \frac{\mu_{\max}\,[\text{Glc}]}{[\text{Glc}] + K_s}\,[\text{Y}]$$

where $\mu_{\max}$, K_s, $Y_{\text{Glc|Y}}$ and $Y_{\text{EtOH|Y}}$ are the maximum specific growth rate of Y, substrate saturation constant, Y-to-Glc yield coefficient and Y-to-EtOH yield coefficient, respectively.

From here, the model can be formulated under a Bayesian framework as defined in Eq. (1.25). Consequently, definition of prior distributions to use for the vector of parameters θ and total standard deviation σ is highly important.

Priors' choice

As mentioned in Section 1.5.3, using a set of weakly informative priors on unit scale is a helpful choice. The most natural option to plan such priors is to scale the original data by dividing each time series per their corresponding maximum value. Hence, we can define a set of normal prior distributions for each kth element of θ around the unit of positive values only as

$$\theta_k \sim \mathcal{N}(0.5,\, 0.3), \quad \theta_k > 0\,.$$

Regarding total standard deviation, the same logic can be applied even by using another normal distribution[3] with same scale parameters as for θ:

$$\sigma \sim \mathcal{N}(0.5,\, 0.3), \quad \sigma > 0\,.$$

Implementation

Once the model has been defined, its Bayesian computation is done with Stan [50], a probabilistic language that offers the MCMC-NUTS [40] method for sampling from the posterior distribution. As any other usual MCMC method, NUTS can ran in multiple chains to better assess convergence on the target distribution. In this example, posterior distributions of parameters and predictions were got by running four Markov chains with 3000 iterations and 1000 of them as warm-up. Solution of the ODE system was performed by means of fourth- and fifth-order Runge-Kutta methods available in the solver *rk45* [51, 52], built-in Stan.

Diagnostics

Regarding diagnostics, here are shown firstly visual representations that help to qualitatively check whether convergence of the MCMC-NUTS method succeeded.

In Figure 1.4, trace plots of the 5 parameters of the toy model are depicted with the usual so-called caterpillar form that is an indicative of convergence. Other visualizations include representation of the posterior distributions (i.e., as density functions or histograms) and scatter plots between involved parameters. Such a representation is shown in Figure 1.5, where signs that convergence has been reached are noticeable. Regarding the diagonal of the plot, unimodal histograms of the parameters' posterior densities shown that independent chains have mixed well.

The off-diagonal scatter plots provide hints regarding identifiability of the parameters. Given that samples from the Markov chains are concentrated around well noticeable regions, leads to conclude that unique solutions for the parameters of interest exist. Furthermore, scatter plots

[3]For estimating standard deviation, prior distributions with heavier tails can cover for more variation on experimental data. See Section 2.2.4.

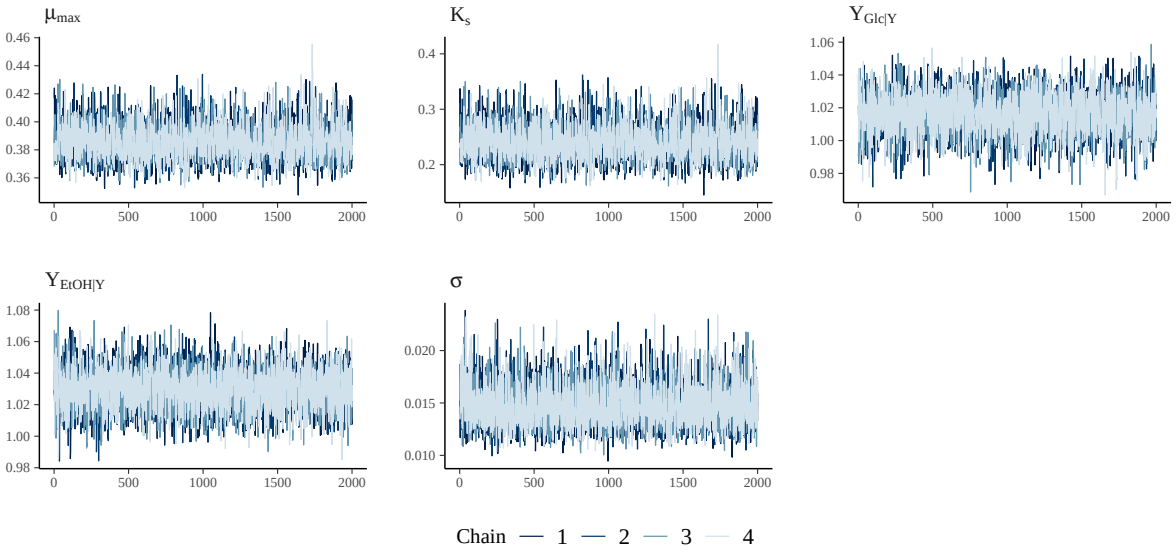


Figure 1.4: Trace plots for the 5 parameters of the toy model. Markov chains without warm-up are shown. Different colors represent independent chains.

also reflect the amount of correlation between the posteriors. Ideally, all parameters should be uncorrelated but, this is not the case for $\mu_{\max}$ vs. K_s as previously reported for specific maximum growth rates and substrate saturation constants by Rosenbaum *et al.* [53]. Ways to solve this issue are to assign more informative priors to either of these parameters or re-parameterize the model [38]. However, as it is seen in their corresponding scatter plot, high concentration of samples between both parameters is concentrated in a centered region of their parameter space reducing the likeliness of high variance that could be interpreted as evidence of practical non-identifiability.

As a final take away message, this kind of visual aids to check convergence are worth to deploy when the number of parameters in the model is relatively low. For this reason, where a model accounts with too many parameters, it makes more sense to check at the convergence statistics $\widehat{R}$ and ESS.

Hence, in Table 1.1 diagnostics and posterior moments and quantiles of parameter posteriors are shown for the toy model fitted on data reported by Petrov [49]. As mentioned in Section 1.5.3, convergence statistics accomplish their respective rules of thumbs with values less than 1.05, greater than 400 and greater than 100 for $\widehat{R}$, ESS, and bulk-/tail-ESS respectively.

Table 1.1: Diagnostics and scaled posterior moments and quantiles of parameter estimates for the toy model with data reported by Petrov [49]. $\widehat{R}$, bulk-ESS and tail-ESS are computed as reported in [48]. Mean, standard deviation (sd), Monte Carlo standard error (MCSE) and quantiles at 2.5%, 50%, 97.5% are also shown.

Parameter	$\widehat{R}$	ESS	bulk-ESS	tail-ESS	Mean	sd	MCSE	2.5%	50%	97.5%
$\mu_{\max}$	1.00	2785	2866	3098	0.39	0.01	0.00	0.36	0.39	0.41
K_s	1.00	2808	2891	3124	0.24	0.03	0.00	0.19	0.24	0.31
$Y_{\text{Glc\|Y}}$	1.00	3507	3535	4229	1.02	0.01	0.00	0.99	1.02	1.04
$Y_{\text{EtOH\|Y}}$	1.00	3969	3976	4125	1.03	0.01	0.00	1.01	1.03	1.05
σ	1.00	3804	3840	4372	0.01	0.00	0.00	0.01	0.01	0.02

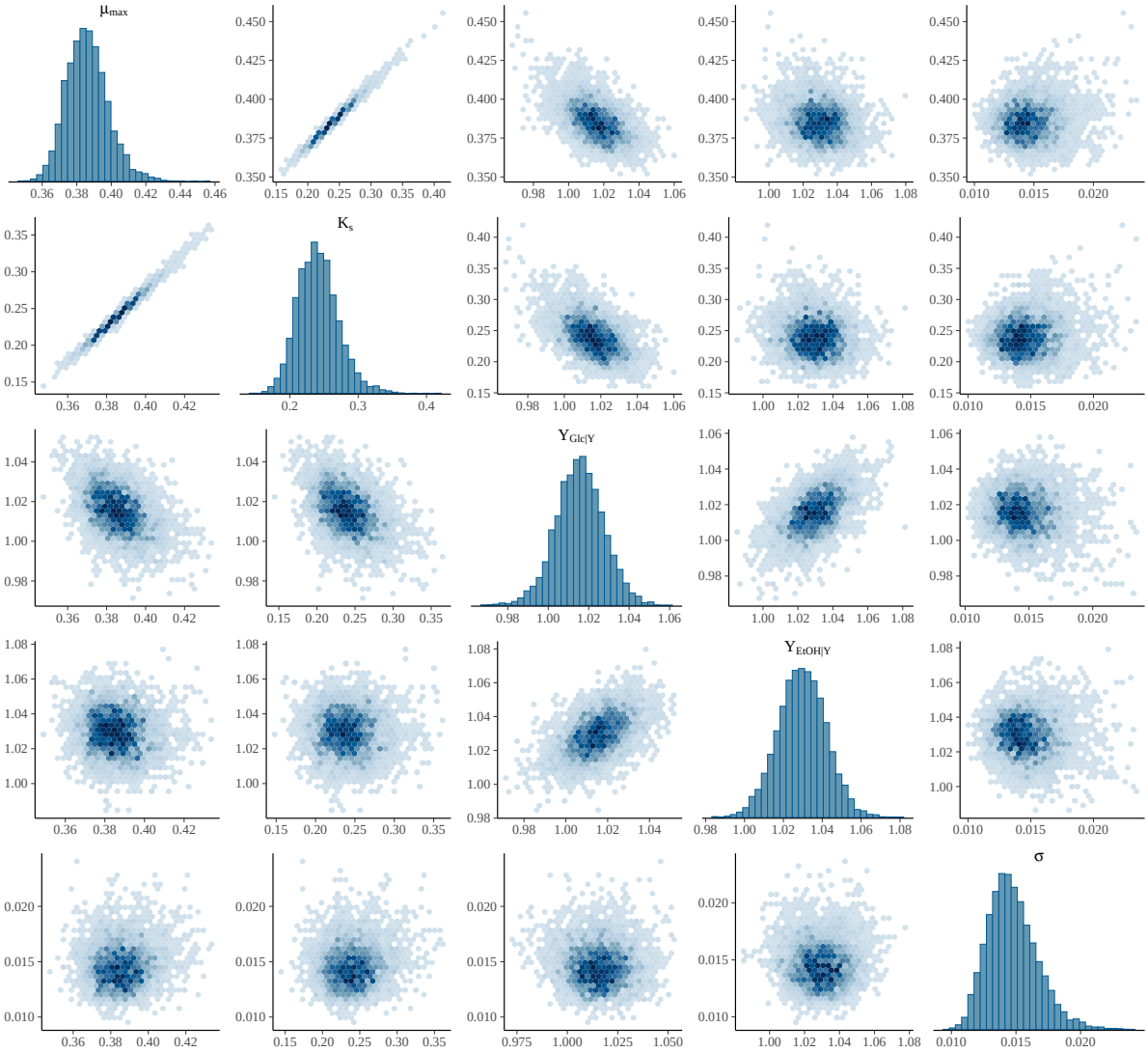


Figure 1.5: Markov chain Monte Carlo scatter matrix for the toy model. Panels in the diagonal show the histogram of the posterior distribution of each parameter. Off-diagonal panels show pair-wise parameters' posterior samples hex scatter plots. Note unimodality of the diagonal panels and highly concentrated regions in the scatter plots.

Posterior predictions and model evaluation

Once convergence of MCMC-NUTS has been accomplished, posterior predictions and model evaluation statistics can be presented. In Figure 1.6, posterior predictions of the toy model for data reported by Petrov [49] are shown.

After scaling the original data to values between 0 and 1 it is necessary to re-scale obtained parameters to their real units. This task can be summarized to a simple use of the maximum values within each time series as conversion factors of the obtained posterior samples (see Appendix B for further details). Hence, summarized parameter values for the model are depicted in Table 1.2.

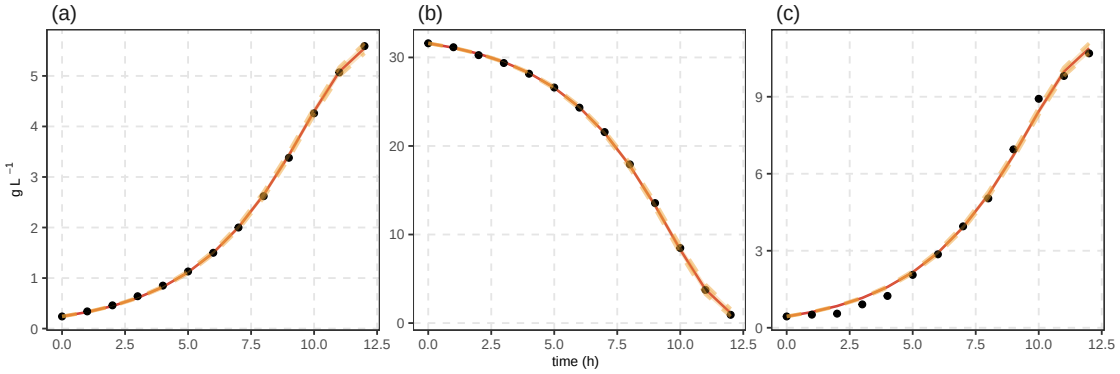


Figure 1.6: Posterior predictions for toy model fitted to data reported by Petrov [49]. (a) Biomass (yeasts), (b) glucose and (c) ethanol. Solid red lines represent posterior medians of the posterior predictions, solid black points denote experimental data and orange ribbon describe the 95% credible interval of posterior predictions.

Finally, WAIC and PSIS-LOO values computed using *loo* package developed in [48] are equal to -195.2 and -194.1, respectively. No further analysis based on these quantities is presented due to their primary use in models' comparison (see Appendix A and Chapter 3).

Table 1.2: Summary of re-scaled posterior parameter estimates for the toy model with data reported by Petrov [49]. Mean, standard deviation (sd), Monte Carlo standard error (MCSE) and 95% credible interval (CI) are shown.

Parameter	Units	Mean	sd	MCSE	95% CI
$\mu_{\max}$	h^{-1}	0.39	0.01	0.00	[0.36, 0.41]
K_s	$\mathrm{g\,L}^{-1}$	7.54	0.95	0.00	[6.00, 9.80]
$Y_{\mathrm{Glc\|Y}}$	$\mathrm{g\,L}^{-1}$	0.182	0.001	0.00	[0.175, 0.184]
$Y_{\mathrm{EtOH\|Y}}$	$\mathrm{g\,L}^{-1}$	0.534	0.005	0.00	[0.528, 0.549]
σ	–	0.01	0.00	0.00	[0.01, 0.02]

1.6 Organization of the thesis

The focus of this thesis is to implement kinetic modeling for the process of cocoa bean fermentation as primary steps in its quantitative mechanistic understanding, and to reveal how analyzes of got parameter estimates can address biological questions. In the following chapters, these questions are assessed through fitting a series of mathematical models over a rich set of data regarding cocoa bean fermentation trials previously reported in literature.

In Chapter 2, the first steps towards probing adequacy of basic regulatory mechanisms is introduced by developing a kinetic model of the process which will serve two main purposes: (1) work as a baseline model for the subsequent exploratory iterations of prevailing hypothesized interactions between microbial populations and metabolites, and (2) establish the use of obtained parameter estimation values as tools for inferring differences with respect to the conditions in which parallel trials were conducted.

In Chapter 3, a deeper exploration of hypothesized mechanisms previously backed up by qualitative descriptions is assessed by constructing a series of 31 candidates model fitted over 23 different datasets of fermentation trials. Because of this scheme, most plausible model variants are determined by including or discarding candidate mechanisms. Besides, resulting vectors

of parameters estimates are employed as classifiers for fermentation features, determining their conceivable use as an alternative to standard chemical fingerprinting in interpreting fermentation data.

As complement to Chapter 3, Chapter 4 explores the impact of introducing temperature as a dynamical variable over the baseline model used across previous chapters and how it does not show remarkable improvements in terms of biological interpretability of parameter estimates, nor predictive accuracy of the models.

Finally, Chapter 5 covers the discussion and conclusion of this research, as well as suggestions of a possible outline for further developments in modeling the process of cocoa bean fermentation.

Chapter 2

A mathematical model of cocoa bean fermentation[4]

Abstract

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 forms the precursors of the resulting chocolate's flavor 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: (1) Metabolite concentrations of glucose, fructose, ethanol, lactic acid and acetic acid, and (2) Population sizes of yeast, lactic acid bacteria and acetic acid bacteria. We demonstrate 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. Besides 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.

2.1 Introduction

The fermentation of cocoa beans is recognized as a key step in cocoa processing in terms of the development of chocolate's flavor and aroma [3, 4]. 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 2.1 (a)), 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 [12–15]. This so-called three-phase process, depending on the region and local farm practices, is expected to happen within a time frame of 2 to 10 days [4, 6, 18, 54, 55].

Because of the fermentation, a series of biochemical reactions is triggered upon the raw material, the qualitative characteristics of which have been exhaustively described in terms of the microbial groups involved and the associated metabolic alterations [54, 56, 57].

[4]This chapter is based on the publication of Moreno-Zambrano *et al.* [1], published by the Royal Society under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, provided the original author and source are credited.

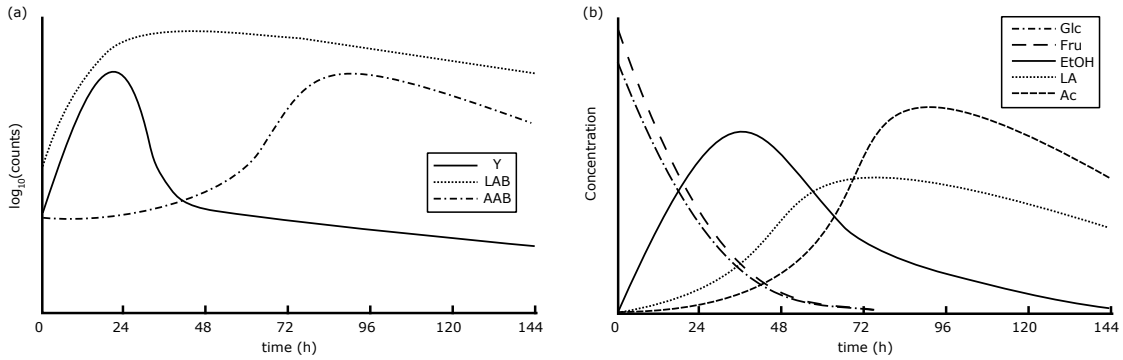


Figure 2.1: Typical time series for community dynamics and metabolite kinetics during cocoa bean fermentation. (a) Community dynamics; yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). (b) Metabolite kinetics; glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac). Both, counts and concentration are shown in arbitrary units. (After De Vuyst & Weckx [18]; use permitted under the Creative Commons Attribution License CC-BY 4.0 number 4354810766457).

Although this process is of high industrial relevance, there are hardly any attempts of constructing a mathematical model of cocoa bean fermentation. So far, the existing modeling attempts are focussing on specific post-fermentation steps such as drying kinetics [58, 59], restricted to the sequential interaction of microbial communities using metabolic flux analysis [19, 60] or kinetic approaches that "cannot explain the dynamics in microbial population" [22].

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 most other food fermentation processes [5] with a huge diversity of techniques and devices, e.g., heaps, boxes, baskets, trays, sacks and platforms [3, 5, 6]. Because of 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 strains [12, 14, 15, 18]. Hence, the complexity is precisely one of the biggest challenges to overcome since growth modeling of microorganisms has been classically applied under much more controlled conditions that involve single strain cultures where mortality phenomena have been scarcely considered [30]. Consequently, this needs to be considered for an approximation of the cocoa bean fermentation process.

Cocoa bean fermentation is a prototypical situation for the application of modeling using coupled non-linear ordinary differential equations: 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 [27] and Contois [29] equations, assuming that single strain kinetic modeling 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 [32].

While conceptionally the modeling approach presented here is rather in the tradition of theoretical biology, the way to analyze 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 [61], rather than asymptotic states as would be expected in continuous cultures.

With the model constructed along these lines, we could 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 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 plausible hypotheses on how these parameters are affected by slight changes within a particular region where the fermentation took place.

2.2 Material and Methods

2.2.1 Experimental data

The experimental data used in this study were reported in Camu *et al.* [12] and Papalexandratou *et al.* [62]. 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.* [12], the beans were fermented by the heaps method, while for Papalexandratou *et al.* [62], wooden boxes were used as fermenting devices. The data of Camu *et al.* [12] 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 (Suc), lactic acid (LA), acetic acid (Ac), ethanol (EtOH), mannitol, citric acid and succinic acid.

The data reported by Papalexandratou *et al.* [62] 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. 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, Suc, 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.* [12] and 14 time points for Papalexandratou *et al.* [62]. 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.

2.2.2 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.* [12] and Papalexandratou *et al.* [62]. 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 get 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 *S. cerevisiae* is equivalent to 15 picograms (pg), as assumed by Schwabe & Bruggeman [63]. For

LAB and AAB, since such conversion factors have not been reported yet, values were inferred by considering a geometric approximation based on the usual dimensions of the cells belonging to the genus of *Acetobacter* and to the species of *Lactobacillus plantarum* respectively, according to the Bergey's Manual of Systematic Bacteriology [64, 65] and assuming their shape given by a spherocylinder. Thus, using as reference a density value derived from the dry weight of a cell of *E. coli* of 0.28 pg [66] per micro cubic meter (μm^3) [67] 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 Appendix A.1).

2.2.3 Model development

Biochemical background

The fermentation of cocoa beans has been described in detail regarding its microbial dynamics and metabolite kinetics in both pulp and bean [6, 12, 14, 15, 18]. 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 2.1). Further processes such as the conversions of Glc into Ac by LAB and LA into Ac by AAB, have also been described [68]. The interpretation of these processes in a network diagram covering the pulp only, is shown in Figure 2.2, 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.

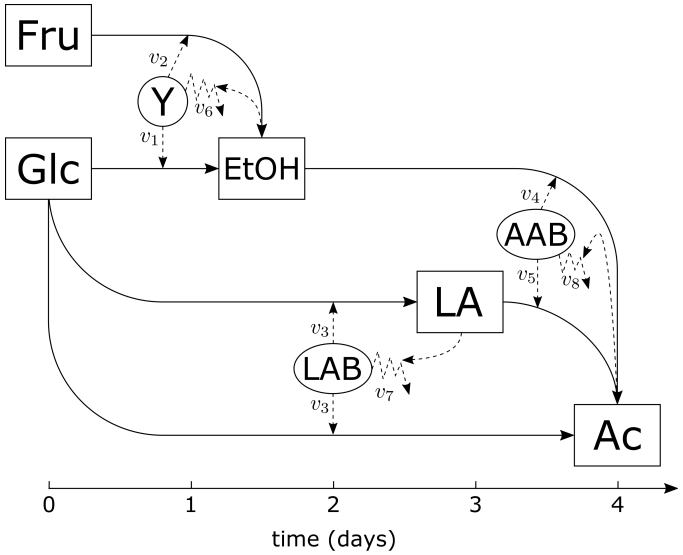


Figure 2.2: Network diagram of the cocoa bean fermentation model. Microbial groups: yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB) are represented as circles. Metabolites: Glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac) are represented as squares. 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 lactic 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 occur.

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.

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 [69]. 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 ordinary differential equation (ODE) for each of the state variables involved in such a way that the growth of microorganisms depends 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 modeled by different approaches. On the one hand, we use the classical Monod [27] and Contois [29] 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 EtOH, v_4, (shown in Figure 2.2) have the form of Monod equations, while the growth of AAB on LA, v_5, is a Contois equation. The use of a Contois term for v_5 was considered under the assumption that given that few specdies of AAB are capable of catabolize lactic acid [21, 60], the growth rate of these species is a function of their population size (see Appendix A.2). On the other hand, the mortality rates of all microbes, v_6, v_7 and v_8, are modeled as Chick-Watson equations [32] considering a non-linear decay of microbial populations produced by second and third order reactions of their corresponding metabolite products upon themselves as shown in Table 2.1. Together, all equations in Table 2.1 comprise 13 parameters: (1) five maximum specific growth rates, (2) five substrate saturation constants and (3) three mortality rate constants.

From the set of growth and mortality rate equations defined in Table 2.1, a system of ODEs can be established in order to mathematically express the network considering the 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 Eqs. (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 2.2.

$$\frac{d\,[\mathrm{Glc}]}{dt} = -Y_{\mathrm{Glc|Y}}\,v_1 - Y_{\mathrm{Glc|LAB}}\,v_3 \tag{2.1}$$

$$\frac{d\,[\mathrm{Fru}]}{dt} = -Y_{\mathrm{Fru|Y}}\,v_2 \tag{2.2}$$

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}\,v_2 - Y_{\mathrm{EtOH|AAB}}\,v_4 \tag{2.3}$$

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\,v_3 - Y_{\mathrm{LA|AAB}}\,v_5 \tag{2.4}$$

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\,v_3 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\,v_4 + Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}\,v_5 \tag{2.5}$$

$$\frac{d\,[\mathrm{Y}]}{dt} = v_1 + v_2 - v_6 \tag{2.6}$$

$$\frac{d\,[\mathrm{LAB}]}{dt} = v_3 - v_7 \tag{2.7}$$

$$\frac{d\,[\mathrm{AAB}]}{dt} = v_4 + v_5 - v_8 \tag{2.8}$$

The proposed model (as described in Eqs. (2.1) to (2.8)) relies on three simple general assumptions: (1) 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 [70], (2) 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 [71, 72], and (3) No impact of chemical and physical effects such as temperature and pH on the set of kinetic parameters.

Table 2.1: 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, both are shown within square brackets [ ]. Maximum specific growth rates $\mu_{\max}$ are shown of the form $\mu^{i_n}_{\max}$, where i can be either Y, LAB and AAB, and n refers whether μ 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
$v_1 = \dfrac{\mu^{\mathrm{Y_{Glc}}}_{\max}\,[\mathrm{Glc}]}{[\mathrm{Glc}] + K^{\mathrm{Y}}_{\mathrm{Glc}}}\,[\mathrm{Y}]$	$v_6 = k_{\mathrm{Y}}\,[\mathrm{Y}]\,[\mathrm{EtOH}]$
$v_2 = \dfrac{\mu^{\mathrm{Y_{Fru}}}_{\max}\,[\mathrm{Fru}]}{[\mathrm{Fru}] + K^{\mathrm{Y}}_{\mathrm{Fru}}}\,[\mathrm{Y}]$	
$v_3 = \dfrac{\mu^{\mathrm{LAB}}_{\max}\,[\mathrm{Glc}]}{[\mathrm{Glc}] + K^{\mathrm{LAB}}_{\mathrm{Glc}}}\,[\mathrm{LAB}]$	$v_7 = k_{\mathrm{LAB}}\,[\mathrm{LAB}]\,[\mathrm{LA}]$
$v_4 = \dfrac{\mu^{\mathrm{AAB_{EtOH}}}_{\max}\,[\mathrm{EtOH}]}{[\mathrm{EtOH}] + K^{\mathrm{AAB}}_{\mathrm{EtOH}}}\,[\mathrm{AAB}]$	$v_8 = k_{\mathrm{AAB}}\,[\mathrm{AAB}]\,[\mathrm{Ac}]^2$
$v_5 = \dfrac{\mu^{\mathrm{AAB_{LA}}}_{\max}\,[\mathrm{LA}]}{[\mathrm{LA}] + K^{\mathrm{AAB}}_{\mathrm{LA}}\,[\mathrm{AAB}]}\,[\mathrm{AAB}]$	

2.2.4 Parameter estimation

Bayesian framework

The parameter estimation was conducted using a Bayesian framework as described in Chapter 1, Section 1.5.3

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.* [12], the maximum concentration of AAB is approximately $0.0019\,\mathrm{mg\,g(pulp)}^{-1}$, while the maximum concentration of the main substrate of AAB, EtOH, is approximately $22.4920\,\mathrm{mg\,g(pulp)}^{-1}$.

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 Eqs. (2.1) to (2.8) (see Appendix A.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; here, by introducing a scale that needs a prior distribution to be sampled within values between 0 to 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 θ 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 σ, a Cauchy distribution $\mathcal{C}$ 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 θ, with the addition that the heavy tails of $\mathcal{C}$ 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

$$\begin{aligned} \theta_k &\sim \mathcal{N}(0.5,\, 0.3), \quad \theta_k > 0 \\ \sigma &\sim \mathcal{C}(0,\, 1), \qquad\quad \sigma > 0. \end{aligned} \tag{2.9}$$

Implementation

The Bayesian parameter estimation framework was performed with Stan [50], using the RStan interface package for R [73, 74]. The model was solved as an initial value problem, where the onset concentrations for the eight state variables in Eqs. (2.1) to (2.8) were provided as they were reported in the original works of Camu *et al.* [12] and Papalexandratou *et al.* [62]. Sampling for obtaining the posterior distributions of the unknown parameters as well as the model predictions, were conducted using full Bayesian inference through the Markov chain Monte Carlo (MCMC) No-U-Turn sampler (NUTS) [40]. The ODEs were specified and solved by the built-in mechanism of Stan *rk45*, which provides a fourth- and fifth-order Runge-Kutta method for solving non-stiff systems [51, 52]. All data sets 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 $\widehat{R}$ statistics computed by Stan.

2.2.5 Statistical analyses

Once the parameters of Eqs. (2.1) to (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.* [62] were compared between each other. For doing so, an effect size statistic was

Table 2.2: 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
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	h^{-1}	Maximum specific growth rate of Y on Glc
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	h^{-1}	Maximum specific growth rate of Y on Fru
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	h^{-1}	Maximum specific growth rate of LAB on Glc
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	h^{-1}	Maximum specific growth rate of AAB on EtOH
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	h^{-1}	Maximum specific growth rate of AAB on LA
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	$\mathrm{mg(Glc)\,g(pulp)}^{-1}$	Substrate saturation constant of Y growth on Glc
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	$\mathrm{mg(Fru)\,g(pulp)}^{-1}$	Substrate saturation constant of Y growth on Fru
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	$\mathrm{mg(Glc)\,g(pulp)}^{-1}$	Substrate saturation constant of LAB growth on Glc
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	$\mathrm{mg(EtOH)\,g(pulp)}^{-1}$	Substrate saturation constant of AAB growth on EtOH
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	$\mathrm{mg(LA)\,g(pulp)}^{-1}$	Substrate saturation constant of AAB growth on LA
k_{Y}	$\mathrm{mg(EtOH)}^{-1}\,\mathrm{h}^{-1}$	Mortality rate constant of Y
k_{LAB}	$\mathrm{mg(LA)}^{-1}\,\mathrm{h}^{-1}$	Mortality rate constant of LAB
k_{AAB}	$\mathrm{mg(Ac)}^{-2}\,\mathrm{h}^{-1}$	Mortality rate constant of AAB
$Y_{\mathrm{Glc\|Y}}$	$\mathrm{mg(Glc)\,mg(Y)}^{-1}$	Y-to-Glc yield coefficient
$Y_{\mathrm{Glc\|LAB}}$	$\mathrm{mg(Glc)\,mg(LAB)}^{-1}$	LAB-to-Glc yield coefficient
$Y_{\mathrm{Fru\|Y}}$	$\mathrm{mg(Fru)\,mg(Y)}^{-1}$	Y-to-Fru yield coefficient
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	$\mathrm{mg(EtOH)\,mg(Y)}^{-1}$	Y-to-EtOH from Glc yield coefficient
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	$\mathrm{mg(EtOH)\,mg(Y)}^{-1}$	Y-to-EtOH from Fru yield coefficient
$Y_{\mathrm{EtOH\|AAB}}$	$\mathrm{mg(EtOH)\,mg(AAB)}^{-1}$	AAB-to-EtOH yield coefficient
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	$\mathrm{mg(LA)\,mg(LAB)}^{-1}$	LAB-to-LA yield coefficient
$Y_{\mathrm{LA\|AAB}}$	$\mathrm{mg(LA)\,mg(AAB)}^{-1}$	AAB-to-LA yield coefficient
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	$\mathrm{mg(Ac)\,mg(LAB)}^{-1}$	LAB-to-Ac yield coefficient
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	$\mathrm{mg(Ac)\,mg(AAB)}^{-1}$	AAB-to-Ac from EtOH yield coefficient
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	$\mathrm{mg(Ac)\,mg(AAB)}^{-1}$	AAB-to-Ac from LA yield coefficient

used as proposed by Cohen [75], 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

$$d = \frac{\overline{\theta}_{k,1} - \overline{\theta}_{k,2}}{S_{w_k}}, \tag{2.10}$$

where $\overline{\theta}_{k,1}$ and $\overline{\theta}_{k,2}$ are the sampled means corresponding to the parameter k obtained from box 1 and box 2, respectively, and S_{w_k} 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

$$S_{w_k} = \sqrt{\frac{S_{k,1}^2 + S_{k,2}^2}{2}}, \tag{2.11}$$

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 [76].

2.3 Results

2.3.1 Model's diagnostics

In all three datasets, the proposed model was fitted without major issues. The calculated $\widehat{R}$ statistic was 1 for all three cases (see Tables A.4, A.5 and A.6), showing that convergence of the MCMC sampler was accomplished. Such a behavior is also noticeable in the obtained trace plots (see Figures A.5, A.6 and A.7), 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 Appendix A.5).

2.3.2 Metabolite and microbial population dynamics

The proposed model, as described in Eqs. (2.1) to (2.8), fits each of the collections of data reported previously by Camu *et al.* [12] (Figure 2.3) and Papalexandratou *et al.* [62] (Figures 2.4 and 2.5) remarkably well.

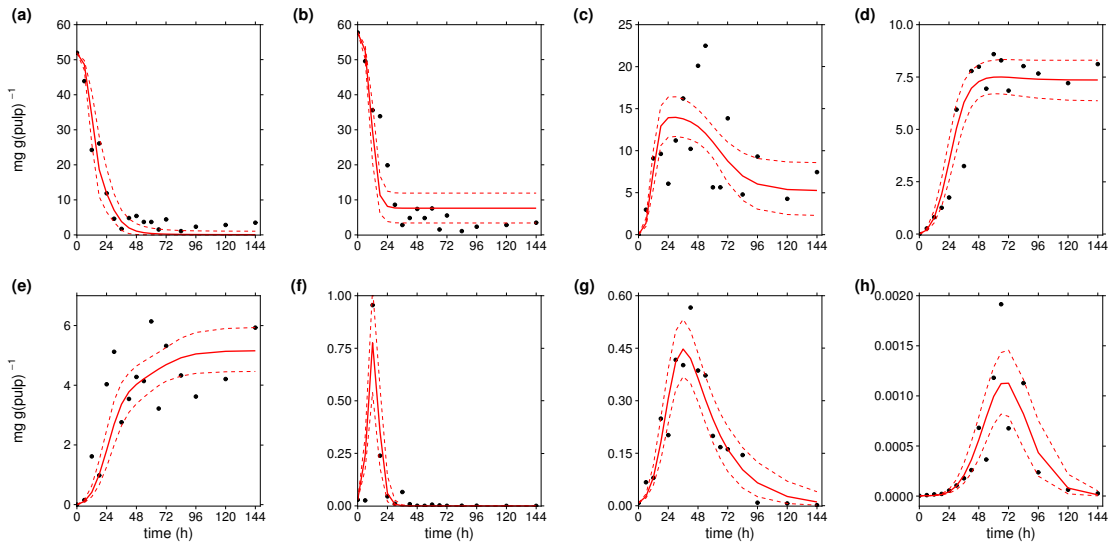


Figure 2.3: Simulation results of the cocoa bean fermentation model for the data reported by Camu *et al.* [12]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid, and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines show to the simulations of the model, while black points denote the experimental data of Camu *et al.* [12]. The red dashed lines represent the 95% credible interval of the model predictions.

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.* [12] and Papalexandratou *et al.* [62] fall within the computed 95% credible intervals of the simulations showing that the model can predict 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.* [12], where its concentrations seem to stay steady after 48 hours of fermentation, which would contradict its consumption by AAB as reported by Pereira *et al.* [68].

2.3.3 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 literature. In this way, from 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 ten and 5 (11.90%) are out of the referenced range by more than the estimate divided or multiplied by ten (Table 2.3). In the following paragraphs our results will be structured in accordance to their type as (1) Maximum specific growth rates, (2) substrate saturation constants, (3) Mortality rates and (4) Yield coefficients.

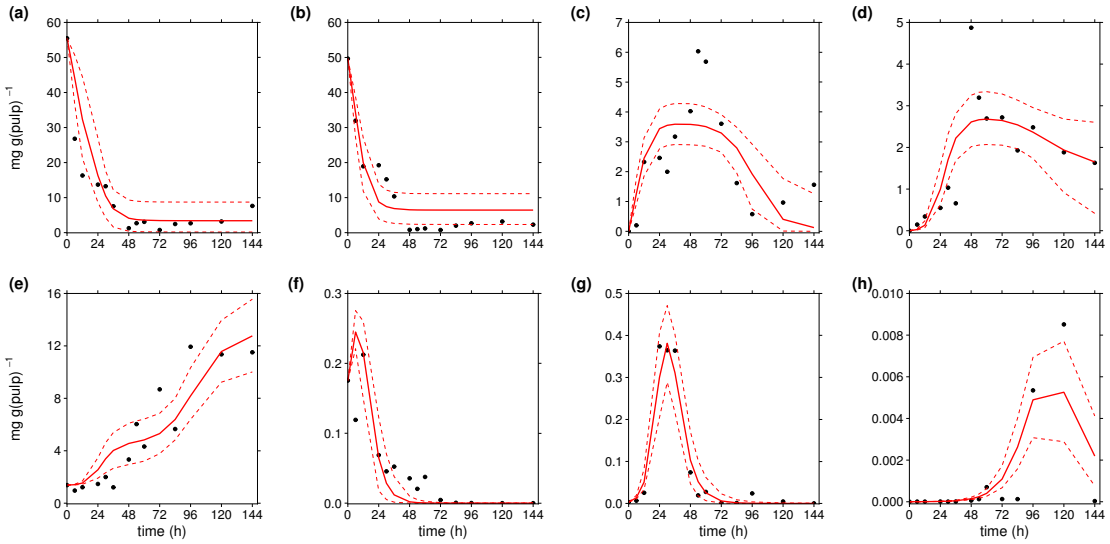


Figure 2.4: Simulation results of the cocoa bean fermentation model for the data reported for box 1 by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid, and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines show to the simulations of the model, while black points denote the experimental data for box 1 of Papalexandratou *et al.* [62]. The red dashed lines represent the 95% credible interval of the model predictions.

Maximum specific growth rates

With respect to the maximum specific growth rates, the estimated values mostly fall within their reported values in literature, with the exception of three estimates belonging to particular

trials. Specifically, for the maximum specific growth rate of Y on Glc ($\mu_{\max}^{\mathrm{Y_{Glc}}}$), 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} [77–84] for Camu *et al.* [12] and Box 2 [62] (green cells in Table 2.3), while the estimate for Box 1 [62] is far from the range in less than the estimate divided by ten (orange cell in Table 2.3). In a similar fashion, the estimated parameters for the maximum growth rate of LAB ($\mu_{\max}^{\mathrm{LAB_{Glc}}}$) 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}[85–90] (all parameters are within green cells in Table 2.3). In contrast, the maximum specific growth rate of Y on Fru ($\mu_{\max}^{\mathrm{Y_{Fru}}}$) and AAB on EtOH ($\mu_{\max}^{\mathrm{AAB_{EtOH}}}$), do not agree completely with reported values in literature. In the first case, only the estimate obtained from box 1 conducted by Papalexandratou *et al.* [62] falls within the range of 0.01 to 0.166 h^{-1} [78, 83, 91, 92] (green cell in Table 2.3). In the latter case, the only estimate that does not fall within the range of 0.0106 to 0.25 h^{-1} [93–97] is the one obtained from the data reported by Camu *et al.* [12] (orange cell in Table 2.3). For the maximum specific growth rate of AAB on LA ($\mu_{\max}^{\mathrm{AAB_{LA}}}$), no values were reported in literature for this microbial group on LA as carbon source. The estimated values for this parameter were between 0.01 to 0.02 h^{-1}.

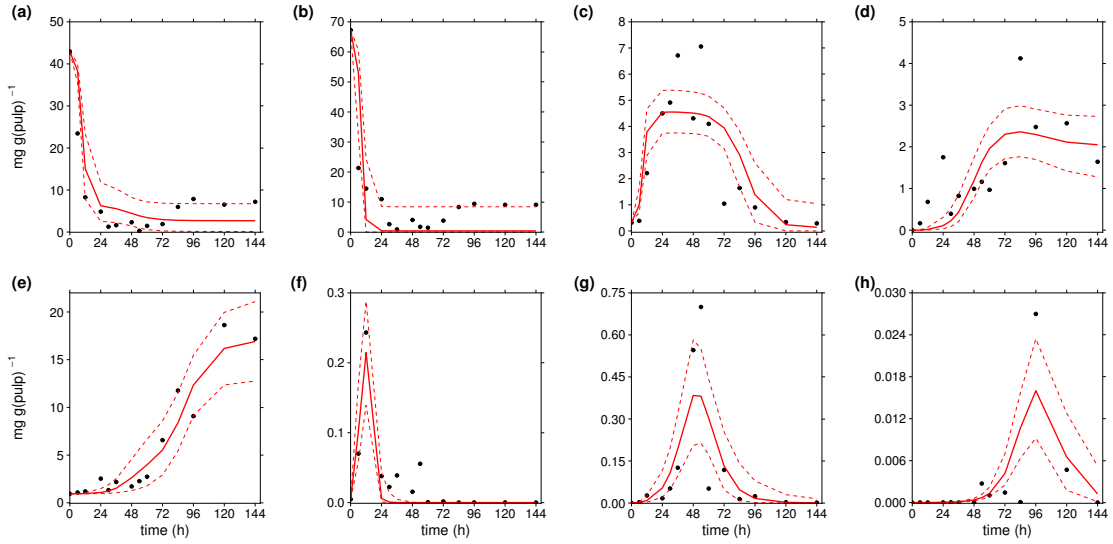


Figure 2.5: Simulation results of the cocoa bean fermentation model for the data reported for box 2 by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid, and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines show to the simulations of the model, while black points denote the experimental data for box 2 of Papalexandratou *et al.* [62]. The red dashed lines represent the 95% credible interval of the model predictions.

Substrate saturation constants

About the substrate saturation constants, here denoted by K, their estimated values for the three datasets agree with reported ones in 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 milliliter of medium (mg(substrate) mL^{-1}), and assuming that one gram of pulp is equivalent to one milliliter of medium, since our estimates are given in mg(substrate) $mg(pulp)^{-1}$.

On the one hand, values reported in literature for these parameters were found for the substrate saturation constants of Y on Glc ($K_{\mathrm{Glc}}^{\mathrm{Y}}$) [82, 84], Y on Fru ($K_{\mathrm{Fru}}^{\mathrm{Y}}$) [92] and LAB on

Glc ($K_{\text{Glc}}^{\text{LAB}}$) [88–90]. The obtained estimated values ranged from 30.01 to 35.32, 25.02 to 41.39 and 19.27 to 37.97 mg(substrate) g(pulp)$^{-1}$ respectively. Estimated values for $K_{\text{Glc}}^{\text{Y}}$ are far from the reported ranges by less than the estimate multiplied by ten (orange cells in Table 2.3); while for $K_{\text{Fru}}^{\text{Y}}$, the estimates were farer from the referenced range than the estimate multiplied by ten (red cells in Table 2.3). From these estimates, the ones corresponding to $K_{\text{Glc}}^{\text{LAB}}$ were those which fall within the reported range of 0.79 to 178.0 mg(Glc) mL^{-1} [88–90] (green cells in Table 2.3).

On the other hand, no values of substrate saturation constants were reported either for AAB on EtOH ($K_{\text{EtOH}}^{\text{AAB}}$) or AAB on LA ($K_{\text{LA}}^{\text{AAB}}$). 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 $K_{\text{LA}}^{\text{AAB}}$, this considerably higher value for the upper limit of the range was obtained for the Camu *et al.* [12] data. This inflation of values was observed for other parameters, i.e., AAB-to-EtOH yield coefficient ($Y_{\text{EtOH|AAB}}$), AAB-to-LA yield coefficient ($Y_{\text{LA|AAB}}$) and AAB-to-Ac from LA yield coefficient ($Y_{\text{Ac|AAB}}^{\text{LA}}$), of this dataset as well.

Mortality rate constants

For the mortality rate constants, k, no values were reported in literature. Here, their estimated values for Y (k_{Y}), LAB (k_{LAB}) and AAB (k_{AAB}) were in the ranges of 0.033 to 0.092 mg(EtOH)$^{-1}$ h^{-1}, 0.0054 to 0.067 mg(LA)$^{-1}$ h^{-1} and 0.0004 to 0.0069 mg(Ac)$^{-2}$ h^{-1}, respectively. These estimates varied considerably between the data of Camu *et al.* [12] and Papalexandratou *et al.* [62]. In the latter, the only estimate that differed much between box 1 and box 2 was the one corresponding to k_{Y}.

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.* [12] and Papalexandratou *et al.* [62], as well as the two trials (box 1 and 2) of the latter. In more detail, the estimated yield coefficient of Y-to-EtOH from Glc ($Y_{\text{EtOH|Y}}^{\text{Glc}}$) was the only one that agree in all three datasets with those reported in literature (green cells in Table 2.3). Their estimated values ranged between 7.44 to 11.94 mg(EtOH) mg(Y)$^{-1}$, falling in the referenced range of 1.39 to 21.49 mg(EtOH) mg(Y)$^{-1}$ [81–83, 98]. For the yield coefficient of LAB-to-Glc ($Y_{\text{Glc|LAB}}$), the fits corresponding to the data of Camu *et al.* [12] and Box 1 [62] were contained in the referenced range of 1.56 to 66.67 mg(Glc) mg(LAB)$^{-1}$ [90, 99] (green cells in Table 2.3) with values of 29.23 and 20.22 mg(Glc) mg(LAB)$^{-1}$ respectively. The remaining estimate $Y_{\text{Glc|LAB}}$ for Box 2 was far from the reported range in less than the estimate multiplied by ten, with a estimated mean of 3.32 mg(Glc) mg(LAB)$^{-1}$ (orange cell in Table 2.3).

In 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 2.3) for the yield coefficient of Y-to-Glc ($Y_{\text{Glc|Y}}$), in the dataset of Camu *et al.* [12] only, agree with the ranges of 1.56 to 66.67 mg(Glc) mg(Y)$^{-1}$ [78–82, 100, 101]. Their counterparts from Boxes 1 and 2 reported by Papalexandratou *et al.* [62], are away from the reported range in less than the estimate multiplied by ten (orange cells in Table 2.3) with values of 240.93 and 119.67 mg(Glc) mg(Y)$^{-1}$ respectively. For the yield coefficient of Y-to-Fru ($Y_{\text{Fru|Y}}$), all estimated parameters are far from the referenced range of 43.48 to 200mg(Fru) mg(Y)$^{-1}$ [78, 91] in less than the estimate divided by ten (orange cells in Table 2.3) with values between 41.11 to 244.15 mg(Fru) mg(Y)$^{-1}$. A similar situation is observed for coefficients Y-to-EtOH from Fru ($Y_{\text{EtOH|Y}}^{\text{Fru}}$) and AAB-to-EtOH ($Y_{\text{EtOH|AAB}}$) (orange cells in Table 2.3), with estimated parameters between 5.927 to 11.195 mg(EtOH) mg(Y)$^{-1}$ and a single referenced value of 5.7878 mg(EtOH) mg(Y)$^{-1}$

Table 2.3: Parameter estimates of the cocoa bean fermentation model using a Bayesian estimation framework. Means and standard deviations (in parenthesis) of the posterior distributions for the data of Camu *et al.* [12] and the fermentation boxes 1 and 2 of Papalexandratou *et al.* [62] are shown in columns Camu, P. Box 1 and P. Box 2, respectively. Additionally, the absolute value of the standardized mean difference (effect size $|d|$) between the estimates of P. Box 1 and P. Box 2 is shown. Effect sizes marked with (*) are significant ($|d| > 1.2$). Reported values in the literature have same units as shown in Table 2.2, except for the substrate saturation constants, which are given in milligrams of substrate per millilitre of medium ($\mathrm{mg(substrate)\,mL^{-1}}$). Green, orange and red colored cells correspond to parameter estimates that lie within the referenced range, are out of the range by less than the estimate divided or multiplied by ten and are out of the range by more than the estimate divided or multiplied by ten respectively.

Parameter	Camu	P. Box 1	P. Box 2	Effect size $\|d\|$	Reported values	Reference
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	0.253 (0.094)	0.063 (0.025)	0.368 (0.131)	3.23*	0.0781 – 0.53	[77–84]
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	0.359 (0.106)	0.083 (0.03)	0.572 (0.151)	4.49*	0.01 – 0.166	[78, 83, 91, 92]
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	0.358 (0.067)	0.414 (0.096)	0.499 (0.152)	0.67	0.0072 – 1.41	[85–90]
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	0.380 (0.092)	0.150 (0.051)	0.168 (0.052)	0.35	0.0106 – 0.25	[93–97]
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	0.008 (0.012)	0.025 (0.017)	0.022 (0.016)	0.18	NA	NA
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	35.322 (13.826)	34.366 (14.964)	30.015 (10.887)	0.33	9.73×10^{-4} – 0.5	[82, 84]
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	35.492 (15.253)	25.015 (14.187)	41.386 (17.259)	1.04	582 – 772	[92]
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	37.966 (12.37)	31.664 (13.882)	19.272 (10.599)	1.00	0.790 – 178.0	[88–90]
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	16.056 (5.646)	3.818 (1.637)	4.056 (1.98)	0.13	NA	NA
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	2509.622 (1234.114)	312.051 (153.064)	81.981 (41.365)	2.05*	NA	NA
k_{Y}	0.0333 (0.0051)	0.0517 (0.0119)	0.092 (0.0248)	2.07*	NA	NA
k_{LAB}	0.0054 (0.0014)	0.0637 (0.0183)	0.0686 (0.025)	0.22	NA	NA
k_{AAB}	0.0069 (0.0014)	0.0004 (0.0002)	0.0004 (0.0002)	0.00	NA	NA
$Y_{\mathrm{Glc\|Y}}$	33.400 (11.255)	240.926 (59.899)	119.672 (35.155)	2.47*	1.56 – 66.67	[78–82, 100, 101]
$Y_{\mathrm{Glc\|LAB}}$	29.259 (10.852)	20.217 (13.24)	3.323 (3.417)	1.75*	4.50 – 200	[90, 99]
$Y_{\mathrm{Fru\|Y}}$	41.105 (11.215)	244.153 (46.219)	232.383 (62.958)	0.21	43.48 – 200	[78, 91]
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	7.436 (4.526)	11.941 (6.061)	8.201 (4.503)	0.70	1.39 – 21.49	[81–83, 98]
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	5.927 (3.351)	11.195 (5.005)	6.008 (3.882)	1.16	5.7878	[83]
$Y_{\mathrm{EtOH\|AAB}}$	1298.070 (637.461)	378.452 (152.674)	170.44 (56.123)	1.81*	8.06 – 166.67	[94, 96, 102]
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	10.617 (1.54)	2.785 (0.664)	2.138 (0.799)	0.88	NA	NA
$Y_{\mathrm{LA\|AAB}}$	1928.619 (1216.393)	287.511 (141.381)	62.9 (38.531)	2.17*	7.94	[96]
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	5.612 (0.919)	3.279 (1.125)	2.899 (1.898)	0.24	NA	NA
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	104.056 (98.355)	576.471 (271.434)	321.858 (132.72)	1.19	NA	NA
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	1427.225 (865.061)	666.903 (357.131)	385.172 (177.474)	1.00	NA	NA
σ	0.149 (0.010)	0.168 (0.014)	0.168 (0.013)	0.00	NA	NA

[83] for $Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}$ and estimated parameters between 170.44 to 1298.070 mg(EtOH) mg(AAB)$^{-1}$ with a referenced range of 8.06 to 166.67 mg(EtOH) mg(AAB)$^{-1}$[94, 96, 102] for $Y_{\mathrm{EtOH|AAB}}$.

Moreover, the estimated yield coefficients for AAB-to-LA ($Y_{\mathrm{LA|AAB}}$) do not agree in any of the data collections with the reference value of 7.94 mg(LA) mg(AAB)$^{-1}$[96] with values far from the reference over ten times the parameter for the datasets of Camu *et al.* [12] and Box 1 of Papalexandratou *et al.* [62] (red cells in Table 2.3) and one value far from the reference in less than the estimate divided by ten for Box 2 (orange cell in Table 2.3).

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

Special attention needs to be given to the values of $Y_{\mathrm{EtOH|AAB}}$, $Y_{\mathrm{LA|AAB}}$ and $Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}$ for the data of Camu *et al.* [12], which showed inflated values that are not biologically plausible.

The effect of the measurement errors is discussed in Appendix A.6.

2.3.4 Statistical comparison of fermentation trials

The statistical comparison of the parameter estimates between the two fermentation trials conducted by Papalexandratou *et al.* [62] showed that significant differences exist among them (Table 2.3), 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 $\mu^{\mathrm{Y_{Glc}}}_{\mathrm{max}}$, $\mu^{\mathrm{Y_{Fru}}}_{\mathrm{max}}$, $K^{\mathrm{AAB}}_{\mathrm{LA}}$, k_{Y}, $Y_{\mathrm{Glc|Y}}$, $Y_{\mathrm{Glc|LAB}}$, $Y_{\mathrm{EtOH|AAB}}$ and $Y_{\mathrm{LA|AAB}}$. 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 [76]. 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.

2.4 Discussion

2.4.1 Model fitting

As shown in the aforementioned results, our current model for cocoa bean fermentation can reproduce each of the datasets with high accuracy. This means that the mechanistic assumptions made here are under 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 dynamically, as opposed to current metabolic flux analyses [19, 60] that assume steady-state metabolic conditions. Moreover, it represents a fully working kinetic model opposed to a previous attempt which can simulate metabolites and products time-courses only [22].

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 setups behind the datasets.

2.4.2 Regulatory assumptions

By analyzing 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 particularity of the fitted model is the presence of strongly elevated estimates for the data of Camu *et al.* [12]. Specifically, the parameters showing such values were: (1) The substrate saturation constant $K_{\mathrm{LA}}^{\mathrm{AAB}}$ and (2) The yield coefficients $Y_{\mathrm{EtOH|AAB}}$, $Y_{\mathrm{LA|AAB}}$ and $Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}}$. Looking at the standard deviations (Table 2.3, Figures A.8, A.9 and A.10) of their corresponding posterior distributions, it can be noticed that there is a huge uncertainty in their values (large errors). These uncertainties reveal 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.* [12]. 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 [103]. 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., $K_{\mathrm{LA}}^{\mathrm{AAB}}$, $Y_{\mathrm{LA|AAB}}$ and $Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}}$, can be a result of noise in experimental data. Thus, as revealed by visual inspection of Figure 2.3 panel (d) where the real data does not reflect a stressed decrease in the concentration of LA as opposed to the data reported by Papalexandratou *et al.* [62] (Figures 2.4 and 2.5, panel (d)), where such decrease exists after 72 hours of the fermentation process. For the remaining elevated parameter which is related to the consumption of EtOH by AAB ($Y_{\mathrm{EtOH|AAB}}$), a straight interpretation of its value of ≈ 1300 mg(Glc) mg(Y)$^{-1}$ would imply that 1300 mg of EtOH are consumed by 1 mg of AAB; or that for generating 1 mg of AAB, it is required 1300 mg of EtOH. 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 $Y_{\mathrm{EtOH|AAB}}$ 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.* [12] shows higher temperatures of the fermentation mass much earlier in the process compared to the data reported by Papalexandratou *et al.* [62]. In the first case, a temperature above 35^oC was reached right after 30 hours and its maximum of $\approx 45^o$C at 70 hours of the fermentation process. In the latter, similar temperatures were reached at 40 and 80 hours of the process. This disparity might explain why the estimated parameter values for $Y_{\mathrm{EtOH|AAB}}$ in the Papalexandratou *et al.* [62] data collection are between 3 to 5 times less compared to the value obtained for the data of Camu *et al.* [12]. Finally, correlation between the estimated parameters might play 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 Appendix A.7), 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.

2.4.3 Parameter conformance to values in literature

The parameter ranges indicated as *reported values* in Table 2.3 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 literature as pointed out in 2.3.3. For the substrate saturation constants $K^{\mathrm{Y}}_{\mathrm{Glc}}$ and $K^{\mathrm{Y}}_{\mathrm{Fru}}$, such a difference can be explained because 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 its values reported in literature for single species confirms the high growth rates and yields coefficients that mixtures of microorganisms might show in fermentation processes [104].

2.4.4 Comparison of parameter estimates

As mentioned in section 2.2.1, the study performed by Papalexandratou *et al.* [62], 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 (see Table 2.3), in the following paragraphs, three main differences were taken into account in order to formulate hypotheses in 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: (1) Initial concentration of microorganisms, (2) Concentration ratios of initial substrates, and (3) Evaporation rates.

Initial concentration of microorganisms

The initial concentrations of microorganisms between boxes 1 and 2 differed little, except for Y. For Y, the initial concentrations in box 1 and 2 were equal to 0.18 and 0.005 $\mathrm{mg\,g^{-1}}$, respectively. This difference as well as the microbial diversity that has been seen along different fermentation trials [4] determined that the estimates of these growth rates differ between each fermentation trial; with higher values of the growth rates $\mu^{\mathrm{Y_{Glc}}}_{\mathrm{max}}$ and $\mu^{\mathrm{Y_{Fru}}}_{\mathrm{max}}$ for box 2 than for box 1. A similar effect is evident in the yield coefficient related to the growth of Y on Glc ($Y_{\mathrm{Glc|Y}}$), 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.

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 $\mathrm{mg\,g^{-1}}$ respectively; while for box 2, these are 42.936 and 67.249 $\mathrm{mg\,g^{-1}}$ respectively. This ratio would explain the significant difference

in the yield coefficient of the growth of LAB on Glc ($Y_{\mathrm{Glc|LAB}}$) 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 [71, 72], which determined a less successful $Y_{\mathrm{Glc|LAB}}$ for LAB under a lower initial concentration of its main substrate Glc in box 2.

Evaporation rates

Among the subtle differences between box 1 and box 2 reported by Papalexandratou *et al.* [62], 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 ($Y_{\mathrm{EtOH|AAB}}$) 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.* [12] dataset. A similarly higher evaporation rate for LA and Ac in box 1 could explain the differences among the remaining parameter estimates, i.e., the Contois substrate saturation constant for the growth of AAB on LA ($K_{\mathrm{LA}}^{\mathrm{AAB}}$), the yield coefficient of consumption of LA by AAB ($Y_{\mathrm{LA|AAB}}$) as well as the large variances of the yield coefficients of the production of Ac from EtOH and LA by AAB ($Y_{\mathrm{Ac|AAB}}^{\mathrm{EtOH}}$ and $Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}}$ 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 losses 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.

2.5 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: (1) Resource-type competition between Y and LAB, (2) Microbial death is determined in a good degree by direct action of fermentation products upon their respective producing microorganisms, (3) 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 to 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 incorporating 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 [19, 60]. Should high-quality genome-scale metabolic models be availabe, flux balance analysis [105] may provide a suitable theoretical framework for such an approach on three levels: (1) A metabolic pathway analysis of synergies and competitions (e.g. using the methodology from Levy & Borenstein [106]) may point to additional modes of interaction among the species involved. (2) 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. (3) 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 [107], we expect this avenue of research to become feasible in the very near future.

The recent finding [108] 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.

Ethics. No ethical considerations apply.
Data, code and materials. The datasets and codes supporting this article are deposited at Dryad: `https://doi:10.5061/dryad.321d33v`
Competing interests. The authors declare no competing interests.
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 analyzes. M.U. contributed and endorsed the biological hypotheses derived from the model. All contributed to writing and editing of the manuscript.
Acknowledgements. We gratefully acknowledge Dr. Anne Grimbs for her excellent and valuable advise 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.
Funding. This work was funded by Barry-Callebaut through the Cocoa Metabolomics (COMETA) project, driven by Jacobs University Bremen.

Chapter 3

Exploring cocoa bean fermentation mechanisms by kinetic modeling[5]

Abstract

Compared to other fermentation processes in food industry, cocoa bean fermentation is uncontrolled and not standardized. A detailed mechanistic understanding can therefore be relevant for cocoa bean quality control. Starting from an existing mathematical model of cocoa bean fermentation, we analyze five additional biochemical mechanisms derived from the literature. These mechanisms, when added to the baseline model either in isolation or in combination, were evaluated in terms of their capacity to describe experimental data. In total, we evaluated 32 model variants on 23 fermentation datasets. We interpret the results from two perspectives: (1) success of the potential mechanism, (2) discrimination of fermentation protocols based on estimated parameters. The former provides insight into the fermentation process itself. The latter opens an avenue towards reverse-engineering empirical conditions from model parameters. We find support for two mechanisms debated in the literature: consumption of fructose by lactic acid bacteria and production of acetic acid by yeast. Furthermore, we provide evidence that model parameters are sensitive to differences in the cultivar, temperature control, and usage of steel tanks compared to wooden boxes. Our results show mathematical modeling can provide an alternative to standard chemical fingerprinting in interpreting fermentation data.

3.1 Introduction

Cocoa beans from *Theobroma cacao* L. are the raw material of chocolate. Their fermentation plays a fundamental role as being responsible for eliminating undesired properties from freshly harvested beans, e.g., astringency and bitterness, besides of yielding chocolate-related flavor and aroma precursor compounds [3, 4]. In contrast to the highly controlled fermentation processes known from other food products, this process is conducted *in situ* at each of the producing farms in a spontaneous form varying in both methodology, e.g., wooden boxes, heaps and platforms [4, 18, 109], and observed microbial diversity [110].

This heterogeneity due to different fermentation methods and indigenous microbiota, leads to a plethora of studies that have qualitatively described the process, e.g. [109, 110]. Among all these, sequentiality of microbial populations thriving on the beans' enclosing pulp constitutes the process dynamics with greatest acceptance [3, 4, 109].

In further detail, regardless of the wide range of factors that could differentiate fermentation trials, sequential succession of microbial groups during their execution can be understood as a three-phased process, where a microbial group dominates each phase during a distinct time

[5]This chapter is based on the publication of Moreno-Zambrano *et al.* [2].

period. In a first stage, anaerobic conditions due to the packed nature of the pulp favor the growth of yeasts that bloom as a consequence of a carbohydrate-rich environment producing mainly ethanol. Through their pectinolytic action yeasts drive to a liquefaction of the pulp. As a consequence, a drainage of pulp permits air to enter into the fermenting mass contributing to the decline of yeast population. Under these conditions, a second stage is dominated by the growth of microaerophilic lactic acid bacteria that at the onset of the process were reproducing at a lower rate than yeasts. Therefore, by depletion of remaining sugars from the first stage, lactic acid bacteria yield mainly lactic and acetic acids. At this point, after considerable drainage of pulp, a fully aerobic phase is reached. This third and final stage is characterized by an almost complete dominance of aerophilic acetic acid bacteria that oxidize lactic acid into acetoin, and ethanol into acetic acid [3, 18, 54, 109].

As a consequence, microbial sequentiality during the fermentation has served as the basis for formulating a few mathematical approaches for its quantitative description [1, 22, 23, 111]. Among these, the model proposed in Eqs. (2.1) to (2.8) served us here as baseline. The main sub-processes implemented in this study focus on the activity of major microbial groups, namely yeasts (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). As a result, we developed a successful model based on well known regulatory assumptions: In a first instance, Y come into play by converting glucose (Glc) and fructose (Fru) into ethanol (EtOH). Concomitantly, LAB consumes Glc leaving as products lactic acid (LA) and acetic acid (Ac). Finally, AAB takes over the last phase of fermentation by oxidizing EtOH and LA into Ac [1].

Beyond these key components, more regulatory mechanisms have been mentioned across experimental studies that could bring more insight into the dynamics of cocoa bean fermentation. Among these, we here put special emphasis in five phenomena (see detailed references in Materials and Methods, below): (1) decrease of product metabolites by physical causes, (2) consumption of Fru by LAB, (3) production of Ac by Y, (4) consumption of LA by Y, and (5) over-oxidation of Ac by AAB.

Along these lines, we were able to assess the plausibility of stand-alone and simultaneous occurrence of these mechanisms when added to our baseline model and to identify systematic differences of fermentation features by applying classification methods over their resulting vectors of parameter estimates. Our key questions are: (1) Which model variants describe the experimental data better than the baseline model? (2) For which model can parameter differences be related to differences in the fermentation process?

3.2 Materials and Methods

3.2.1 Identification and processing of experimental data

A literature survey concerning cocoa bean fermentation trials was performed with the purpose of gathering experimental data. Reported trials considered in this study were papers published between 2000 to 2019. As inclusion criteria, only English-written works with time series of minimum 5 observations for metabolites Glc, Fru, EtOh, LA and Ac, besides total population counts of Y, LAB and AAB were included.

In all cases, population growth of Y, LAB and AAB were transformed from log base 10 of colony forming units (log_{10}(CFU)) to milligrams of microbial group (MG) per gram of pulp (mg(MG) g(pulp)$^{-1}$). Moreover, all time series were scaled by dividing each observation by its own maximum value. These data preparation steps have been performed, in order to get kinetic parameter values comparable with previously reported ones and facilitating their estimation by avoiding numerical issues during model calibration [1].

For this current research, distinct trials were given a code name based on country of origin and fermentation method. A complete detail of data included in this research where at least

one model variation successfully fit it (see following sections for their explanation) is shown in Table 3.1. For a comprehensive list of all data initially considered see Appendix B.1.

Table 3.1: Considered data sources. Only fermentation trials that were successfully described by at least one model iteration (MI) are listed. Author, year of publication, cocoa country of origin, cocoa cultivar, used methodology, code name given in the original trial, re-coded given name in this research, turning of the fermenting mass and controlled temperature are shown.

Reference	Year	Country	Cultivar	Method	Trial	Code	Turning	Ctrl. Temp.
Camu *et al.* [12]	2007	Ghana	Criollo/Forastero	heap	heap 5	ghhp1	✗	✗
Lagunes Gálvez *et al.* [112]	2007	Dominican Republic	Trinitario	wooden box	NA	dowb1	✓	✗
Camu *et al.* [17]	2008	Ghana	NA*	heap	heap 10	ghhp2	✓	✗
					heap 11	ghhp3	✗	✗
					heap 12	ghhp4	✓	✗
					heap 13	ghhp5	✗	✗
Papalexandratou *et al.* [62]	2011	Brazil	Criollo/Forastero	wooden box	box 1	brwb1	✓	✗
					box 2	brwb2	✓	✗
Papalexandratou *et al.* [113]	2011	Ecuador	Nacional/Trinitario	platform	P1	ecpt1	✗	✗
					P2	ecpt2	✗	✗
				wooden box	B1	ecwb1	✓	✗
					B2	ecwb2	✓	✗
Pereira *et al.* [68]	2012	Brazil	NA*	plastic box	PC	brpb1	✓	✓
				stainless tank	ST	brst1	✓	✓
Pereira *et al.* [14]	2013	Brazil	Mixed hybrids*	wooden box	WB1	brwb3	✓	✗
					WB2	brwb4	✓	✗
				stainless tank	SST	brst2	✓	✗
Moreira *et al.* [15]	2013	Brazil	PH16	wooden box	PH16	brwb7	NA	✗
Papalexandratou *et al.* [114]	2013	Malaysia	Mixed hybrids	wooden box	box 2	mywb3	✓	✗
Romanens *et al.* [115]	2018	Honduras	IMC-67, UF-29, UF-668	wooden box	OF-F	hnwb1	✓	✗
†Lee *et al.* [116]	2019	Ecuador	Criollo	plastic box	NA	ecpb1	NA	✓
Papalexandratou *et al.* [117]	2019	Nicaragua	Nugu/O'payo	wooden box	NUGU	niwb1	✓	✗
					O'PAYO	niwb2	✓	✗

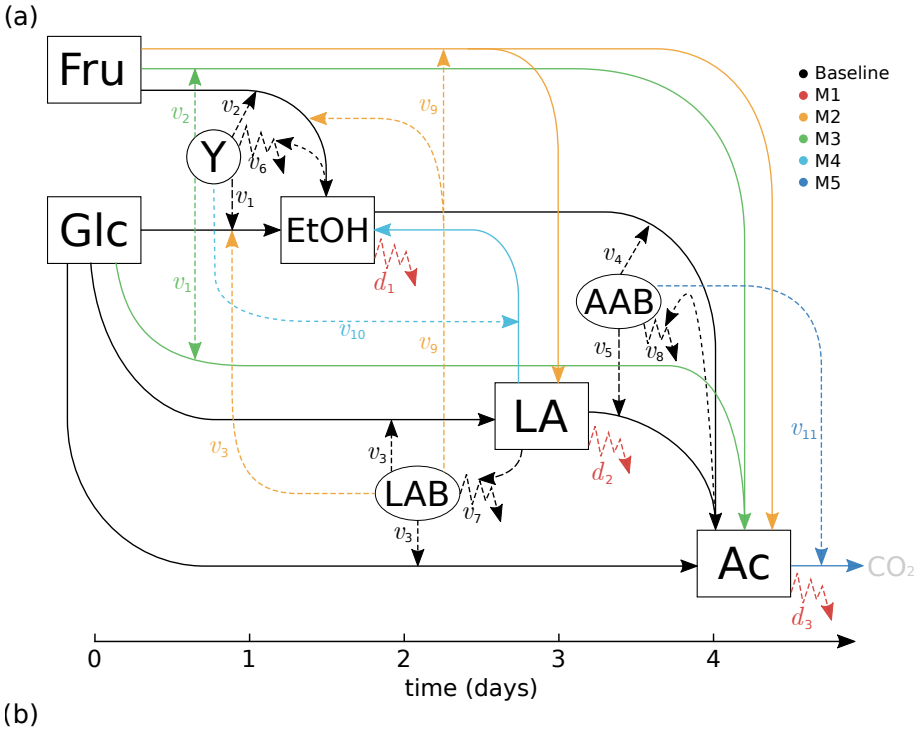
* Unidentified cultivars used by Camu *et al.* [17], Pereira *et al.* [68] and Pereira *et al.* [14] were coded as *un1*, *un2* and *un3*, respectively for further PCA.
† Simulated fermentation.

3.2.2 Formulation of candidate models

Starting from the baseline model, we implemented five regulatory mechanisms that have been reported or hypothesized in multiple studies. In the following paragraphs, the baseline model will be described and proposed mechanisms reasoning will be presented conforming what we considered their likeliness of occurrence as (1) decay of fermentation's products, (2) consumption of Fru by LAB, (3) production of Ac by Y, (4) consumption of LA by Y, and (5) over-oxidation of Ac by AAB.

Baseline model

As baseline, we used the model proposed in Eqs. (2.1) to (2.8) that consist of 8 ODEs describing the dynamics of metabolites: Glc, Fru, EtOH, LA and Ac, besides microbial groups: Y, LAB and AAB. Both, metabolites and microbial groups are interdependent in the dynamic process by means of growth and mortality rates of the latter (Figure 3.1 (a)). Monod [27] and Contois [29] type equations were employed to describe the growth rates of microbial groups. Growth



(b)

$$\frac{d\,[\mathrm{Glc}]}{dt} = -Y_{\mathrm{Glc|Y}}\,v_1 - Y_{\mathrm{Glc|LAB}}\,v_3$$

$$\frac{d\,[\mathrm{Fru}]}{dt} = -Y_{\mathrm{Fru|Y}}\,v_2 - Y_{\mathrm{Fru|LAB}}\,v_9$$

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}\,v_2 + Y^{\mathrm{Glc}}_{\mathrm{EtOH|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|LAB}}\,v_9 + Y^{\mathrm{LA}}_{\mathrm{EtOH|Y}}\,v_{10} - Y_{\mathrm{EtOH|AAB}}\,v_4 - d_1$$

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{LA|LAB}}\,v_9 - Y_{\mathrm{LA|AAB}}\,v_5 - Y_{\mathrm{LA|Y}}\,v_{10} - d_2$$

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{Ac|LAB}}\,v_9 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\,v_4 + Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}\,v_5 + Y^{\mathrm{Glc}}_{\mathrm{Ac|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{Ac|Y}}\,v_2 - Y_{\mathrm{Ac|AAB}}\,v_{11} - d_3$$

$$\frac{d\,[\mathrm{Y}]}{dt} = v_1 + v_2 + v_{10} - v_6$$

$$\frac{d\,[\mathrm{LAB}]}{dt} = v_3 + v_9 - v_7$$

$$\frac{d\,[\mathrm{AAB}]}{dt} = v_4 + v_5 + v_{11} - v_8$$

Figure 3.1: Summary of models iterations. (a) Network diagram of mechanisms over baseline 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 Y on Glc (v_1), Fru (v_2) and LA (v_{10}), of LAB on Glc (v_3) and Fru (v_9), and of AAB on EtOH (v_4), LA (v_5) and Ac (v_{11}) are represented as straight dashed arrows. The mortality rates of Y (v_6), LAB (v_7) and AAB (v_8) are represented as zigzag dashed arrows as the decay rates of EtOH (d_1), LA (d_2) and Ac (d_3). 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 occur. Baseline model comprehends mechanisms depicted in black (•). (b) Representation of full model with mechanisms M1, M2, M3, M4 and M5 together. M1 (•), encompasses loses of EtOH, LA and Ac. M2 (•), involves conversion of Glc into EtOH, and Fru into EtOH, LA and Ac by LAB. M3 (•), comprises conversion of Glc and Fru into Ac by Y. M4 (•), refers to conversion of LA into EtOH by Y. M5 (•), represents over-oxidation of Ac by AAB.

rates v_1 and v_2 of Y on Glc and Fru respectively, as well as growth rates v_3 of LAB on Glc and v_4 of AAB on EtOH, correspond to Monod equations, while the growth of AAB on LA, v_5, corresponds to a Contois term. Mortality rates of Y, LAB and AAB were modeled as Chick-Watson equations [32] by considering second- and third-order death kinetics, as shown in Table 3.2.

The model contains 24 parameters: five maximum specific growth rates, five substrate

saturation constants, three mortality rate constants and eleven yield coefficients as depicted in Eqs. (2.1) to (2.8).

Table 3.2: Growth, mortality and decay rates for cocoa bean fermentation models. 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). Microbial groups and metabolites are expressed as concentrations, both within square brackets [ ]. Maximum specific growth rates $\mu_{\max}^{i_n}$, correspond to the maximum growth rate of microbial group i, growing on substrate n. Substrate saturation constants K_m^i, correspond to the substrate saturation constant of microbial group i, growing on substrate m. Constant mortality rates k_i, correspond to mortality of microbial group i. Decay rates d_j, correspond to decay rate of metabolite j. All rates with the exception of d_1, d_2, d_3, v_9, v_{10} and v_{11}, are part of the baseline model.

Growth rate equation	Mortality rate equation	Decay rate equation
$v_1 = \dfrac{\mu_{\max}^{\mathrm{Y_{Glc}}}\ [\mathrm{Glc}]}{[\mathrm{Glc}] + K_{\mathrm{Glc}}^{\mathrm{Y}}}\ [\mathrm{Y}]$ $v_2 = \dfrac{\mu_{\max}^{\mathrm{Y_{Fru}}}\ [\mathrm{Fru}]}{[\mathrm{Fru}] + K_{\mathrm{Fru}}^{\mathrm{Y}}}\ [\mathrm{Y}]$ $v_{10} = \dfrac{\mu_{\max}^{\mathrm{Y_{LA}}}\ [\mathrm{LA}]}{[\mathrm{LA}] + K_{\mathrm{LA}}^{\mathrm{Y}}}\ [\mathrm{Y}]$	$v_6 = k_{\mathrm{Y}}\ [\mathrm{Y}]\ [\mathrm{EtOH}]$	$d_1 = b_{\mathrm{EtOH}}\ [\mathrm{EtOH}]$
$v_3 = \dfrac{\mu_{\max}^{\mathrm{LAB_{Glc}}}\ [\mathrm{Glc}]}{[\mathrm{Glc}] + K_{\mathrm{Glc}}^{\mathrm{LAB}}}\ [\mathrm{LAB}]$ $v_9 = \dfrac{\mu_{\max}^{\mathrm{LAB_{Fru}}}\ [\mathrm{Fru}]}{[\mathrm{Fru}] + K_{\mathrm{Fru}}^{\mathrm{LAB}}}\ [\mathrm{LAB}]$	$v_7 = k_{\mathrm{LAB}}\ [\mathrm{LAB}]\ [\mathrm{LA}]$	$d_2 = b_{\mathrm{LA}}\ [\mathrm{LA}]$
$v_4 = \dfrac{\mu_{\max}^{\mathrm{AAB_{EtOH}}}\ [\mathrm{EtOH}]}{[\mathrm{EtOH}] + K_{\mathrm{EtOH}}^{\mathrm{AAB}}}\ [\mathrm{AAB}]$ $v_5 = \dfrac{\mu_{\max}^{\mathrm{AAB_{LA}}}\ [\mathrm{LA}]}{[\mathrm{LA}] + K_{\mathrm{LA}}^{\mathrm{AAB}}\ [\mathrm{AAB}]}\ [\mathrm{AAB}]$ $v_{11} = \dfrac{\mu_{\max}^{\mathrm{AAB_{Ac}}}\ [\mathrm{Ac}]}{[\mathrm{Ac}] + K_{\mathrm{Ac}}^{\mathrm{AAB}}}\ [\mathrm{AAB}]$	$v_8 = k_{\mathrm{AAB}}\ [\mathrm{AAB}]\ [\mathrm{Ac}]^2$	$d_3 = b_{\mathrm{Ac}}\ [\mathrm{Ac}]$

In regard to our proposed mechanisms, their inclusion into the baseline model is conducted by adding extra growth and mortality rates, as well as linear terms when needed (see Table 3.2). For a deeper look into their mathematical formulation see Appendix B.2.

Mechanism 1 (M1): Decay of fermentation products

This mechanism is based on concentration decline of product metabolites at later stages of fermentation that has been hypothesized as a consequence of both, physical and biological con-

straints. Here, we will take into account the first group only. Among these, volatile compounds (e.g., EtOH and Ac) might decrease as a result of evaporation and leakage of fermentation sweating [12, 17, 18, 54]. Regarding non-volatile compounds (e.g., LA), the widely described diffusion process of metabolites from the pulp into the cocoa bean, might also play an important role in their reduction [54, 109, 118].

Mechanism 2 (M2): Consumption of Fru by LAB

Opposed to our original approach of modeling LAB growth exclusively based on Glc uptake [1], mechanism 2 takes into account obligatory and facultatively heterofermentative species which are capable of using Glc and Fru as carbon sources (e.g., *L. fermentum* and *L. plantarum*, respectively) with an accompanying production of EtOH besides LA and Ac [18, 34, 109, 112, 117, 119].

Mechanism 3 (M3): Production of Ac by Y

Mechanism 3 is based on the evidence of among fermentation products that Y generate (e.g., ethanol, glycerol and carbon dioxide), Ac can be created through pyruvate metabolism and tricarboxylic acid cycle [18, 68, 109]. Besides, under controlled conditions, production of Ac by Y could explain concentrations of Ac that do not correspond to AAB's population sizes [14].

Mechanism 4 (M4): Consumption of LA by Y

During the first stage of fermentation, the Y population prevails due to the anaerobic conditions in the pulp. However, under an aerobic environment as during the third stage, yeasts such as *S. cerevisiae* are capable of oxidize LA to produce pyruvate [21, 120]. Additionally, other species of yeast (e.g., *Pichia fermentans* and *Candida krusei*) can assimilate LA and produce EtOH .

Mechanism 5 (M5): Over-oxidation of Ac by AAB

During the last stage of fermentation, AAB dominates microbial population by taking advantage of a fully aerobic environment, while consuming EtOH and LA previously produced by Y and LAB, respectively. Once EtOH has mostly diminished, it has been argued that AAB starts over-oxidizing Ac into carbon dioxide, which would lead to halt the cocoa fermentation process due to an increase of temperature that results in the declining of Y, LAB and AAB [11, 18, 54, 109].

A comprehensive graphical representation of all proposed mechanisms is shown in Figure 3.1, panel (a), full model including all mechanisms here proposed is shown in Figure 3.1, panel (b) and a detailed interpretation of all model parameters is shown in Table 3.3.

3.2.3 Models iterations

To check the plausibility of different mechanisms working together, a series of model variants with combinations of M1, M2, M3, M4 and M5 were created, starting from the baseline model. Hence, 31 model iterations (MIs) plus the baseline model were object of being fitted to experimental data under a Bayesian parameter estimation framework. Each model iteration (MI) is labeled according to the mechanisms involved. For example, the full model containing all 5 proposed regulatory schemes is labeled MI(1,2,3,4,5), while the baseline is labeled MI(0).

Table 3.3: Parameters of the cocoa bean fermentation baseline model and proposed mechanisms. 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). B, M1, M2, M3, M4 and M5 refer to baseline model and mechanisms 1 to 5, respectively.

Parameter	Mechanism	Units	Interpretation
$\mu_{max}^{Y_{Glc}}$	B	h^{-1}	Maximum specific growth rate of Y on Glc
$\mu_{max}^{Y_{Fru}}$	B	h^{-1}	Maximum specific growth rate of Y on Fru
$\mu_{max}^{Y_{LA}}$	M4	h^{-1}	Maximum specific growth rate of Y on LA
$\mu_{max}^{LAB_{Glc}}$	B	h^{-1}	Maximum specific growth rate of LAB on Glc
$\mu_{max}^{LAB_{Fru}}$	M2	h^{-1}	Maximum specific growth rate of LAB on Fru
$\mu_{max}^{AAB_{EtOH}}$	B	h^{-1}	Maximum specific growth rate of AAB on EtOH
$\mu_{max}^{AAB_{LA}}$	B	h^{-1}	Maximum specific growth rate of AAB on LA
$\mu_{max}^{AAB_{Ac}}$	M5	h^{-1}	Maximum specific growth rate of AAB on Ac
K_{Glc}^{Y}	B	$mg(Glc)\,g(pulp)^{-1}$	Substrate saturation constant of Y growth on Glc
K_{Fru}^{Y}	B	$mg(Fru)\,g(pulp)^{-1}$	Substrate saturation constant of Y growth on Fru
K_{LA}^{Y}	M4	$mg(Fru)\,g(pulp)^{-1}$	Substrate saturation constant of Y growth on LA
K_{Glc}^{LAB}	B	$mg(Glc)\,g(pulp)^{-1}$	Substrate saturation constant of LAB growth on Glc
K_{Fru}^{LAB}	M2	$mg(Fru)\,g(pulp)^{-1}$	Substrate saturation constant of LAB growth on Fru
K_{EtOH}^{AAB}	B	$mg(EtOH)\,g(pulp)^{-1}$	Substrate saturation constant of AAB growth on EtOH
K_{LA}^{AAB}	B	$mg(LA)\,g(pulp)^{-1}$	Substrate saturation constant of AAB growth on LA
K_{Ac}^{AAB}	M5	$mg(Ac)\,g(pulp)^{-1}$	Substrate saturation constant of AAB growth on Ac
k_{Y}	B	$mg(EtOH)^{-1}\,h^{-1}$	Mortality rate constant of Y
k_{LAB}	B	$mg(LA)^{-1}\,h^{-1}$	Mortality rate constant of LAB
k_{AAB}	B	$mg(Ac)^{-2}\,h^{-1}$	Mortality rate constant of AAB
$Y_{Glc\|Y}$	B	$mg(Glc)\,mg(Y)^{-1}$	Y-to-Glc yield coefficient
$Y_{Glc\|LAB}$	B	$mg(Glc)\,mg(LAB)^{-1}$	LAB-to-Glc yield coefficient
$Y_{Fru\|Y}$	B	$mg(Fru)\,mg(Y)^{-1}$	Y-to-Fru yield coefficient
$Y_{Fru\|LAB}$	M2	$mg(Fru)\,mg(LAB)^{-1}$	LAB-to-Fru yield coefficient
$Y_{EtOH\|Y}^{Glc}$	B	$mg(EtOH)\,mg(Y)^{-1}$	Y-to-EtOH from Glc yield coefficient
$Y_{EtOH\|Y}^{Fru}$	B	$mg(EtOH)\,mg(Y)^{-1}$	Y-to-EtOH from Fru yield coefficient
$Y_{EtOH\|Y}^{LA}$	M4	$mg(EtOH)\,mg(Y)^{-1}$	Y-to-EtOH from LA yield coefficient
$Y_{EtOH\|LAB}^{Glc}$	M2	$mg(EtOH)\,mg(LAB)^{-1}$	LAB-to-EtOH from Glc yield coefficient
$Y_{EtOH\|LAB}^{Fru}$	M2	$mg(EtOH)\,mg(LAB)^{-1}$	LAB-to-EtOH from Fru yield coefficient
$Y_{EtOH\|AAB}$	B	$mg(EtOH)\,mg(AAB)^{-1}$	AAB-to-EtOH yield coefficient
$Y_{LA\|LAB}^{Glc}$	B	$mg(LA)\,mg(LAB)^{-1}$	LAB-to-LA from Glc yield coefficient
$Y_{LA\|LAB}^{Fru}$	M2	$mg(LA)\,mg(LAB)^{-1}$	LAB-to-LA from Fru yield coefficient
$Y_{LA\|AAB}$	B	$mg(LA)\,mg(AAB)^{-1}$	AAB-to-LA yield coefficient
$Y_{LA\|Y}$	M4	$mg(LA)\,mg(Y)^{-1}$	Y-to-LA yield coefficient
$Y_{Ac\|LAB}^{Glc}$	B	$mg(Ac)\,mg(LAB)^{-1}$	LAB-to-Ac from Glc yield coefficient
$Y_{Ac\|LAB}^{Fru}$	M2	$mg(Ac)\,mg(LAB)^{-1}$	LAB-to-Ac from Fru yield coefficient
$Y_{Ac\|AAB}^{EtOH}$	B	$mg(Ac)\,mg(AAB)^{-1}$	AAB-to-Ac from EtOH yield coefficient
$Y_{Ac\|AAB}^{LA}$	B	$mg(Ac)\,mg(AAB)^{-1}$	AAB-to-Ac from LA yield coefficient
$Y_{Ac\|Y}^{Glc}$	M3	$mg(Ac)\,mg(Y)^{-1}$	Y-to-Ac from Glc yield coefficient
$Y_{Ac\|Y}^{Fru}$	M3	$mg(Ac)\,mg(Y)^{-1}$	Y-to-Ac from Fru yield coefficient
$Y_{Ac\|AAB}$	M5	$mg(Ac)\,mg(AAB)^{-1}$	AAB-to-Ac yield coefficient
b_{EtOH}	M1	h^{-1}	Decay rate of EtOH
b_{LA}	M1	h^{-1}	Decay rate of LA
b_{Ac}	M1	h^{-1}	Decay rate of Ac

3.2.4 Kinetic parameter estimation

The number of parameters among MIs constructed over combination of mechanisms ranges from 24 in the baseline model, to 43 in the full model, including all mechanisms. In each case, the same general Bayesian framework as presented in Chapter 1 was used to sample their posterior distributions, where means were taken as point estimates with their corresponding 95% credible interval (CI) [1].

Choice of priors

Posterior distributions of θ and σ were computed using weakly informative priors, namely a normal distribution with mean 0.5 and standard deviation of 0.3 for each element of the parameter vector θ and a Cauchy distribution with location 0 and scale of 1 for σ. With the purpose of avoiding estimates with negative values, both priors were truncated to the positive set of real numbers, as in Eq. (2.9).

The regularization procedure of the data described in Section 3.2.1 allows us to use the same prior distributions for all parameters. The motivation of this choice of priors is discussed in more detail in Section 2.2.4. For a detailed description of prior distributions re-scaled to the parameters' original units see Table B.2.

Implementation

The fit of MIs to experimental data was performed with Stan [50] via RStan package in R [121, 122]. Posterior distributions of θ, σ and $f\left(x_{i,j}, \theta\right)$, were obtained by Markov chain Monte Carlo (MCMC) No-U-Turn sampler (NUTS) method [40]. Each model was treated as an initial value problem, where ODEs were solved by the built-in Stan numerical solver *rk45* for non-stiff systems by means of fourth- and fifth-order Runge–Kutta method [51, 52]. All MIs were fitted to data by running four parallel Markov Chains of 3000 iterations each, with 1000 of them used for warm-up. Sampling convergence was assessed by examining $\widehat{R}$ statistic, bulk effective sample size (bulk-ESS) and tail effective sample size (tail-ESS) as described by Vehtari *et al.* [44]. In cases where either bulk-ESS or tail-ESS were rejected at first, calibration routine was re-run doubling iterations (2000 for warm-up, 6000 in total) before reporting non-convergence.

3.2.5 Model assessment

The quality of the models was assessed from two perspectives: (1) success of each MI across all data, and (2) predictive accuracy comparison of all MIs for each dataset.

In more detail, in the first case, an observed success rate (OSR) and expected success rate (ESR) were determined on the basis of times where the model was satisfactorily fit to a given dataset. OSR is then defined by the number of successful fits over the total of datasets that were fitted at least one time by any MI. In order to properly compare the success rates of models with only a single additional mechanism with those models containing a combination of mechanisms, we compute an expected success rate (ESR) as the product of the OSRs of the elementary MIs. For model variant MI(1,2,5), for example, the expected success rate is then the product of the observed success rates of the elementary models MI(1), MI(2) and MI(5).

All MIs that were suitably fitted to each dataset were compared by means of Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO) with the aim of checking on their predictive accuracy in case a certain MI could perform outstandingly better than its counterparts [48].

3.2.6 Principal components analysis

Six main features of fermentation trials were taken into account as groups and analysed via principal component analysis (PCA) over parameter estimates. Of these, three consist of multiple classes: (1) country of origin, (2) cacao cultivar, and (3) fermentation method. The other three features are binary: (4) use of starter culture, (5) turning of fermenting mass, and (6) controlled temperature during fermentation. Experiments with features reported as unknown or missing were not considered for any PCA. Only the medians of posterior distributions resulted from each chain in the MCMC-NUTS runs were taken into account to perform PCA. Hence, 4 parameter vector estimates were considered as representative per successful MI. The assessment of groups within the PCA results was possible only for features with more than 1 successful fit.

Moreover, PCAs were also performed over subgroups of parameter defined by the type of parameter and its association with a certain microbial group. Considered subgroups comprised: (1) all MI parameters, (2) maximum specific growth rates, (3) mortality rates, (4) yield coefficients, (5) Y-related parameters, (6) LAB-related parameters, and (7) AAB-related parameters. No PCA was run over substrate saturation constants due to their known correlation with maximum growth rates [53]. All PCAs used mean-centered data with no scaling given that solutions of MIs were determined over scaled time series.

Finally, pair-wise squared Malahanobis distances (D_M), [123] were computed between grouping classes of each feature to quantify the magnitude of their separation. To achieve this, centroids of PCA scores from principal component 1 (PC1) and principal component 2 (PC2) were computed for each j grouping class within an i feature and used to determine D_M as:

$$D_M\ (\mathrm{PC1}_{i,j}, \mathrm{PC2}_{i,j}) = (\overline{x}_1 - \overline{x}_2)^\intercal\ \mathbf{S}^{-1}\ (\overline{x}_1 - \overline{x}_2)\ , \tag{3.1}$$

where $\overline{x}_1$ and $\overline{x}_2$ are the centroid values of the scores of $\mathrm{PC1}_{i,j}$ and $\mathrm{PC2}_{i,j}$ respectively; and $\mathbf{S}^{-1}$ is the inverse of the covariance matrix between groups classes [123, 124].

Both, PCA and D_M were implemented in R using functions *'prcomp'* [121] and *'pairwise.mahalanobis'* [125], respectively.

3.3 Results

3.3.1 First assessment of the models

First, we want to understand how well the different models – the baseline model and the MIs containing one or more of the additional mechanisms – perform. In order to identify differences in the success rate of the model variants, we apply every MI to every fermentation dataset. Table 3.4 summarizes the result.

In general terms, MIs summed up to 1024 runs over 32 available datasets; of which, 207 resulted in successful fits with values of $\widehat{R}$ below 1.05, bulk-ESS and tail-ESS higher than 100 indicating that convergence of the MCMC-NUTS was accomplished (see Tables B.3 to B.5). A number of 9 datasets reported by Lefeber *et al.* [13], Moreira *et al.* [15], Lefeber *et al.* [126], Bastos *et al.* [127] and Racine *et al.* [128] were not possible to fit with any MI at all. The remaining 23 fermentation datasets constitute the scope of our further investigation (see Table 3.1). As an example, Figure 3.2 shows one MI, MI(2,3) describing the time series of one of the datasets (*mywb3* from Papalexandratou *et al.* [114]).

A striking observation is that the vast majority of MIs involving M5 were not able to produce successful fits to experimental data. Among these, exceptions are datasets described by Papalexandratou *et al.* [62], *brwb1* and *brwb2*. Both were well fitted by MI(5), MI(1,5) and MI(1,2,3,5); while MI(2,5) and MI(3,5) fitted *brwb1* and MI(1,4,5) fitted *brwb2* only. Thus, MIs

(4,5), (2,3,5), (2,4,5), (3,4,5), (1,2,4,5), (1,3,4,5), (2,3,4,5) and (1,2,3,4,5) could not describe any dataset at all (see Table 3.4).

Table 3.4: Summary of successful fits across 31 models iterations (MIs) and baseline. Light green-colored cells indicate successful fits. Light-red colored cells indicate non-successful fits. Columns "MI( )", "#", "OSR" and "ESR" refer to combination of mechanisms deployed in MI, number of parameters, observed success rate (OSR) and expected success rate (ESR), respectively.

glhp1	dowb1	glhp2	glhp3	glhp4	glhp5	brwb1	brwb2	ecpt1	ecpt2	ecwb1	ecwb2	brpb1	brst1	brwb3	brwb4	brst2	brwb7	mywb3	hnwb1	ecpb1	niwb1	niwb2	MI( )	#	OSR	ESR
																							0	24	0.78	
																							1	27	0.52	
																							2	31	0.61	
																							3	26	0.83	
																							4	28	0.65	
																							5	27	0.09	
																							1,2	34	0.39	0.32
																							1,3	29	0.48	0.43
																							1,4	31	0.39	0.34
																							1,5	30	0.09	0.05
																							2,3	33	0.74	0.50
																							2,4	35	0.43	0.40
																							2,5	34	0.04	0.05
																							3,4	30	0.57	0.54
																							3,5	29	0.04	0.07
																							4,5	31	0.00	0.06
																							1,2,3	36	0.30	0.26
																							1,2,4	38	0.43	0.21
																							1,2,5	37	0.09	0.03
																							1,3,4	33	0.39	0.28
																							1,3,5	32	0.09	0.04
																							1,4,5	34	0.04	0.03
																							2,3,4	37	0.52	0.33
																							2,3,5	36	0.00	0.04
																							2,4,5	38	0.00	0.03
																							3,4,5	33	0.00	0.05
																							1,2,3,4	40	0.39	0.17
																							1,2,3,5	39	0.09	0.02
																							1,2,4,5	41	0.00	0.02
																							1,3,4,5	36	0.00	0.02
																							2,3,4,5	40	0.00	0.03
																							1,2,3,4,5	43	0.00	0.01

3.3.2 Model success

Computation of OSR resulted in a value of 0.78 for the baseline model (MI(0)). For MIs containing single mechanisms, MI(1), MI(2), MI(3), MI(4) and MI(5), OSRs were 0.52, 0.61, 0.83, 0.65 and 0.09, respectively. Among more complex combinations of mechanisms, OSRs ranged between 0.00 to a maximum of 0.74 reached by the combination of M2 and M3 (MI(2,3)), as listed in Table 3.4. Note that in general we expect a decrease of OSR with an increasing number of parameters in the model, due to the higher complexity of the model. Values for the single-mechanism MIs are therefore not directly comparable to the one of the baseline model.

For even larger MIs (composite mechanisms) we have the ESR to partially correct for this.

Pertaining ESRs, leaving out non-successful MIs, 2 out of 18 MIs showed higher values than their corresponding OSRs. These two MIs with higher ESRs correspond to iterations including M5 (i.e., MI(2,5) and MI(3,5)). Leaving aside MI(2,5) and MI(3,5) due to be the only exceptions of M5 ending up in successful fits, in overall combinations of mechanisms seems to lead to increases of their OSR over ESR on describing different datasets despite not over-passing the OSR of the baseline model (see Table 3.4).

3.3.3 Posterior predictions

Next, we resort to the distributions of posterior probabilities, in order to assess differences in the fit's quality for the different MIs. Among the 23 datasets that were fitted by at least one MI, posterior predictions describe their dynamics remarkably well. In each data collection, despite highly influential observations and sampling rates ranging from 6 to 17 data points, time courses are simulated to an acceptable level. Again we refer to the example shown in Figure 3.2, showing the fit of MI(2,3) to the dataset *mywb3* (see Figures B.2 – B.17 for posterior predictions made by MI(2,3) and Figures B.18 – B.23 for MIs, where M(2,3) was not suitable).

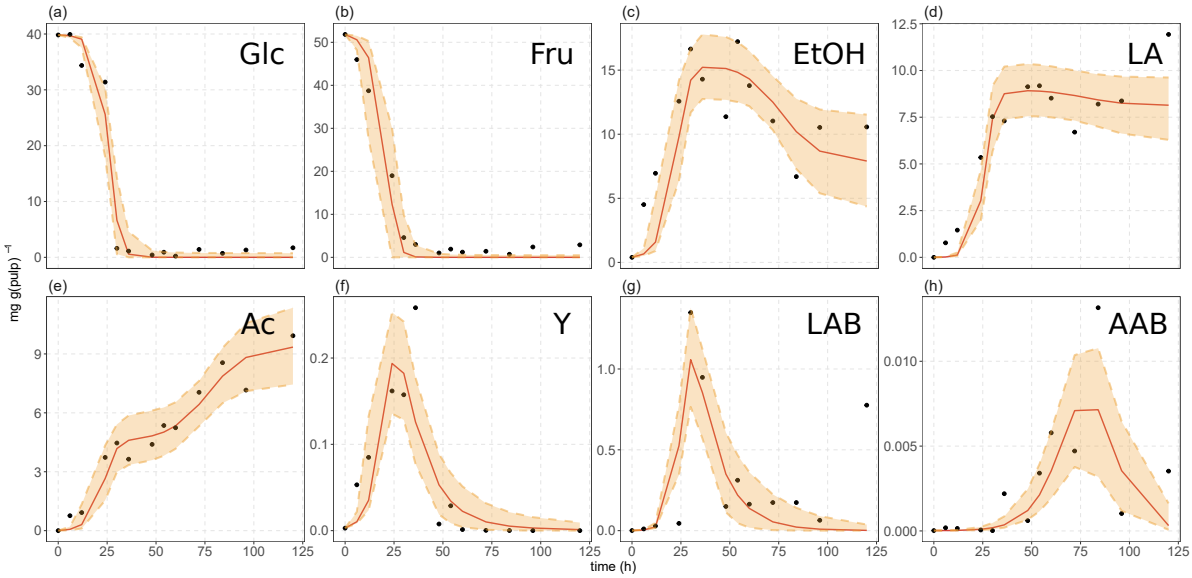


Figure 3.2: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *mywb3* reported by Papalexandratou *et al.* [114]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent posterior medians of the posterior predictions, solid black points denote experimental data and orange ribbon describe the 95% credible interval of posterior predictions.

In terms of predictive accuracy among MIs fitted on each dataset, there were no outstanding differences on basis of obtained PSIS-LOO deviance values that had overlapping standard errors between each other. Thus, is not surprising that for these cases posterior predictions resulted to be extremely similar (see Figure B.24 for an example). Nevertheless, slight distinct PSIS-LOO were observed towards favoring MIs involving M1 in for datasets *brpb1*, *brwb4*, *brst2*, and *niwb2* (see Figure B.1).

These subtle differences provided visually better fits by MI(1) with respect to MI(2,3) for datasets *brpb1* (see Figure B.25). However, the same does not seem to be clear for *niwb2* where predictions made by M(1) and M(2,3) overlay each other with no clear improvements for either both MIs (see Figure B.26).

3.3.4 Fermentation features

We now turn to the second question raised, namely whether the model parameters obtained by describing the datasets with all MIs are informative of the fermentation features behind the datasets. Via principal component analysis performed on the full parameter vectors or biologically meaningful subsets of the parameters, we want to assess whether distinct clusters emerge in agreement with differences in fermentation setups.

After dropping the use of a starter culture as a feature (see Appendix B.1), 490 PCAs were performed from the remaining 5 features and 7 parameters subsets. Note that MI(1,4) did not converge for datasets representing more than one used fermentation method, and MI(1), MI(1,2), MI(1,3), MI(1,4), MI(2,4), MI(1,2,3), MI(1,2,4), MI(1,3,4), MI(2,3,4) and MI(1,2,3,4) were not capable of describing datasets with more than one class of controlled temperature (see Figure B.27).

In terms of group separation measured by D_M for cases with more than one pair-wise comparison, medians of Malahanobis distances ($\tilde{D}_M$) were computed to visualize the magnitude of separation as single values. From this analysis, it can be described in general terms that cultivar, temperature and fermentation method showed the highest $\tilde{D}_M$ values; while, origin countries and turning of fermenting mass showed almost no separation between groups (with the exception of a few cases). Details are provided in the following subsections (see also Figure B.27).

Grouping of fermentation trials according to cultivar

PCAs with cultivar as the feature of interest showed a consistent pattern of high values for $\tilde{D}_M$ with special emphasis on the subgroup of all parameters. With regard to MIs, clearer separations were the product of mostly complex MIs involving combinations of M2, M3 and M4 (see Figure B.27, panel (b)).

Figure 3.3 shows a PCA plot for MI(2,3). Among the four cultivar varieties three showed a clear separation, namely Criollo/Forastero, Nacional/Trinitario and *un2* with explained variances of 29.83% by the first PCA component (PC1) and 18.53% by the second component (PC2). From its loading plot (Figure 3.3, panel (b)), parameters with negative loadings in PC1, mainly $\mu_{\max}^{\mathrm{Y_{Glc}}}$, $\mu_{\max}^{\mathrm{Y_{Fru}}}$, $\mu_{\max}^{\mathrm{LAB_{Glc}}}$, $\mu_{\max}^{\mathrm{LAB_{Fru}}}$, $\mu_{\max}^{\mathrm{AAB_{LA}}}$, $\mu_{\max}^{\mathrm{AAB_{EtOH}}}$, $Y_{\mathrm{LA|LAB}}^{\mathrm{Glc}}$, $Y_{\mathrm{LA|LAB}}^{\mathrm{Fru}}$, $Y_{\mathrm{EtOH|LAB}}^{\mathrm{Glc}}$, $Y_{\mathrm{EtOH|LAB}}^{\mathrm{Fru}}$, $Y_{\mathrm{EtOH|Y}}^{\mathrm{Glc}}$, $Y_{\mathrm{EtOH|Y}}^{\mathrm{Fru}}$, $Y_{\mathrm{Ac|LAB}}^{\mathrm{Glc}}$, $Y_{\mathrm{Ac|AAB}}^{\mathrm{EtOH}}$, $Y_{\mathrm{Ac|Y}}^{\mathrm{Glc}}$, $Y_{\mathrm{Ac|Y}}^{\mathrm{Fru}}$, k_{Y}, and k_{AAB} determine the classes separation.

Grouping of fermentation trials according to temperature control

Temperature control showed its highest $\tilde{D}_M$ values with MIs involving M2 (see Figure B.27, panel (e)). The PCA of the set of all parameters for MI(2,3) resulted in a clear separation of groups determined by the use of controlled and non-controlled temperature with PC1 and PC2 scores explaining 24.97% and 16.14% of variance, respectively (Figure 3.4). Likewise, the same set of parameters as in Section 3.3.4, showed negative loadings in PC1, indicating that these are defining the observed separation.

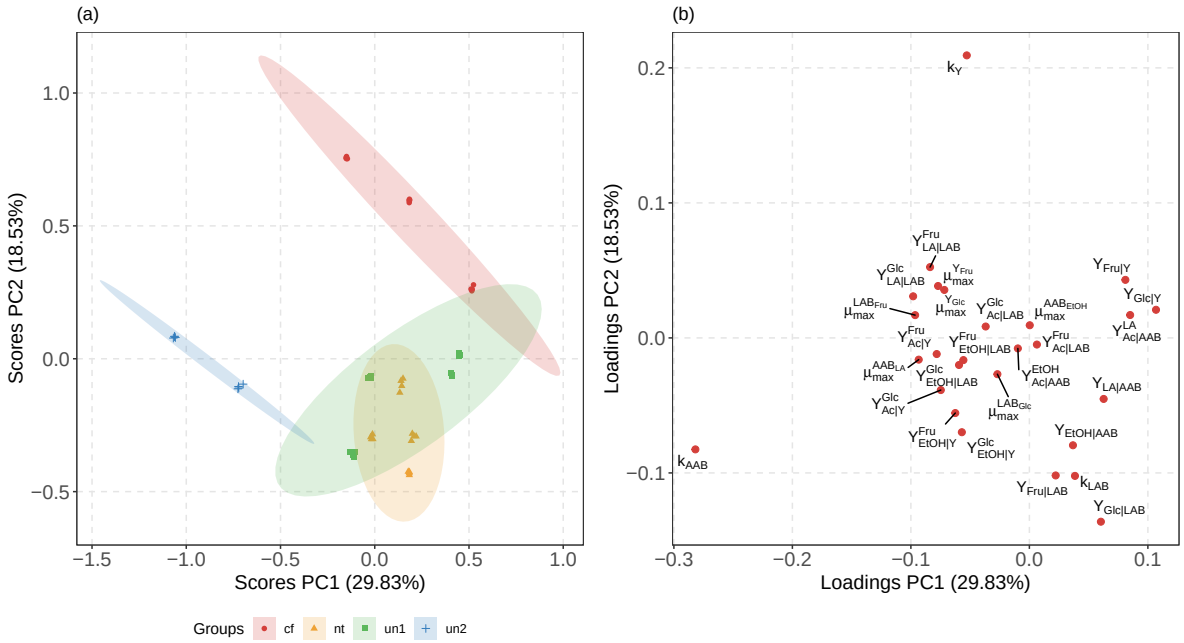


Figure 3.3: PCA score (a) and loading plot (b) from all parameters of model iteration MI(2,3), feature cultivar. Criollo/Forastero (cf), Nacional/Trinitario (nt), unknown cultivar used by Camu *et al.* [17] (un1) and unknown cultivar used by Pereira *et al.* [68] (un2) are shown. Parameters located on the left and right with respect to 0 in PC1 loading plot determine differentiation between cf, nt and un2.

Grouping of fermentation trials according to method

Fermentation method as classification feature resulted in less clear separation patterns. Highest values of $\tilde{D}_M$ were observed for MI(2) (see Figure B.27, panel (c)). In this case a clear separation between fermentations carried in stainless-steels tanks and the rest of fermentation methods can be seen with total explained variance of 33.79% by PC1 and 17.44% by PC2 (see Supplementary Figure B.28). Furthermore, parameters with negative loadings in PC1, namely $\mu_{\max}^{Y_{Fru}}$, $\mu_{\max}^{Y_{Fru}}$, $\mu_{\max}^{LAB_{Glc}}$, $\mu_{\max}^{LAB_{Fru}}$, $\mu_{\max}^{AAB_{LA}}$, $\mu_{\max}^{AAB_{EtOH}}$, $Y_{LA|LAB}^{Glc}$, $Y_{LA|LAB}^{Fru}$, $Y_{EtOH|Y}^{Glc}$, $Y_{EtOH|Y}^{Fru}$, $Y_{EtOH|LAB}^{Glc}$, $Y_{EtOH|LAB}^{Fru}$, $Y_{Ac|AAB}^{EtOH}$, and k_{AAB} seem to influence the separation between stainless-steel tank and the rest of methods (see Figure B.28, panel (b)).

Grouping of fermentation trials according to country of origin and turning of fermenting mass

In contrast to previous features, country of origin and turning of fermenting mass have lower $\tilde{D}_M$ values for single mechanisms and less cases of notorious high distances. Instead, combinations of mechanisms and other subgroups of parameters rather than the whole set, led to better classes separations. For country of origin, PCAs over LAB-related parameters on MIs involving combinations of M1, M2 and M4 resulted in higher $\tilde{D}_M$. For turning of fermenting mass, there were no sizeable differences (see Figure B.27, panels (a) and (d)). For this case PCA from MI(1,3,4) explains a total variance of 60.36% for PC1 and 29.46% for PC2. Furthermore, a clear separation among countries, namely Brazil, Ecuador and Ghana, is defined over PC1. From its loadings, parameters on the left with respect to 0 in PC1 ($\mu_{\max}^{LAB_{Glc}}$, $Y_{LA|LAB}$ and k_{LAB}) are clearly denoting the separation between classes (see Figure B.29). Lastly, turning of fermenting mass did not show any separation either for MI nor any subset of parameters.

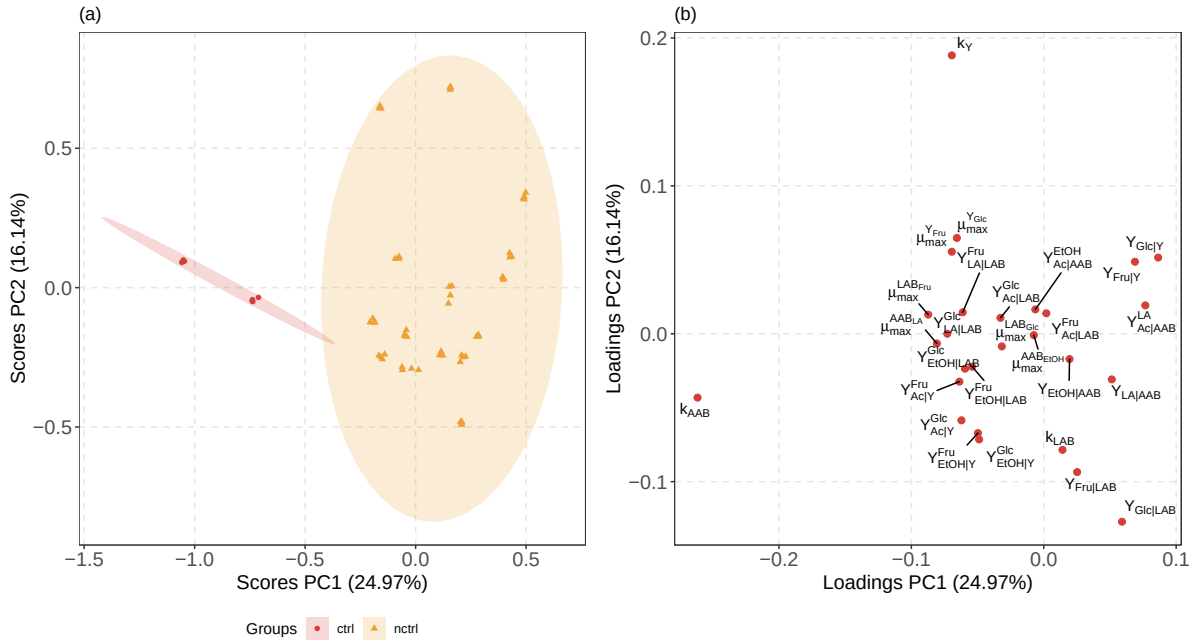


Figure 3.4: PCA score (a) and loading plot (b) from all parameters of model iteration MI(2,3), feature temperature. Controlled temperature (ctrl) and non-controlled temperature (nctrl) are shown. Parameters located on the left and right with respect to 0 in PC1 loading plot determine differentiation between ctrl and nctrl.

3.4 Discussion

3.4.1 Model plausibility and convergence

We have assessed a series of mathematical model variants (or MIs) for cocoa bean fermentation in terms of two levels of plausibility: convergence of each run of a MI over a given dataset, and success of each MI to adequately describe the whole range of fermentation datasets. In this sense, non-convergence might be a consequence of practical non-identifiability of the parameters caused by weakly informative observations or priors, and mis-specification of the model [129]. While success of a model, quantitatively represented as OSR per MI, highly depends on its capability to describe as many datasets as possible. Correspondingly, non-convergence becomes an indirect diagnostic tool that serve us to argue whether hypothesized regulatory interactions of the cocoa bean fermentation process are actually likely to be an influencing factor in the observed fermentation time series.

In general, the mechanisms discussed here have shown in most of cases that their stand-alone and concomitant inclusion in the baseline model lead to convincing OSRs values, in particular for M2, M3 and M4. We can consider a lack of success in some runs involving these mechanisms is the product of numerically conflicting combinations of mechanisms rather than possible misspecification of the whole MI itself.

On the other hand, wide non-convergence patterns in runs involving M5 (see Table 3.4) and to a lesser extent also for M1, leads to the conclusion that these two mechanisms are not such significant influencing factors in the fermentation data studied here.

In the following, we will elaborate on the implications of non-convergence and plausibility of each mechanism considered in descending order relative to their OSR of the stand-alone inclusion in the baseline model.

Starting with M3, production of Ac by Y's metabolism has not been hypothesized from

direct experimental measurements as is the case for other mechanisms. Instead, indirect kinetic studies of isolated strains have shown such an ability of some species of Y [68]. This property has also been argued to be a possible explanation of high Ac production yields where populations of AAB seemed incapable of producing such amounts, as proposed by Pereira *et al.* [14]. Hence, given the OSRs obtained from its inclusion in the series of MIs presented here, M3 can be considered as quantitative evidence backing up this role of Y during fermentation. We strongly believe that cases where these MIs failed to converge, are the consequence of weakly informative priors incapable of being sampled properly.

With respect to M2, in light of the recent characterization of fructophilic lactic acid bacteria (FLAB) in cocoa bean fermentation processes [119], the level of OSR achievement of MIs involving M2 is not surprising. The existence of this bacterial group can then explain an apparently discrepancy in the amounts of Glc and Fru consumed during the process. Despite both substrates being depleted in parallel, Glc is usually consumed first as Y populations reach their end. Thus, uptake of remaining Fru might be consequence of FLAB activity. In contrast to M3, non-convergence of M2 might be a result of weakly informative observations in some datasets with no time lag between Glc and Fru consumption.

Similarly to M3, the reasoning behind M4 relies on indirect characterization of Y strains capable of metabolizing LA into EtOH [112] via their known metabolic pathways producing pyruvate from LA [21, 120] for further production of EtOH [109]. However, in contrast to M3, M4 obtained an appreciable OSR for its stand-alone iteration rather than for its occurrence jointly with either M2 and M3. As an example, consider MI(2,3) compared to MI(2,4) and MI(3,4). While MI(2,3) stands out as the MI with more than one mechanism with largest OSR (equal to 0.74), its counterparts involving M4 perform poorly with OSRs equal to 0.43 for MI(2,4) and 0.57 for MI(3,4). In our opinion, this counter-intuitive performance of M4 when combined with M2 and M3 can be the result of numerical issues due to conflicting interactions of these mechanisms preventing the Bayesian optimization to converge, rather than biological causes against M4.

MIs employing M1 and M5 resulted in the lowest OSRs among all, both stand-alone and combined with other mechanisms. However, important distinctions need to be made between these two. First, M1 is formulated on the basis of several experimental studies that have brought evidence of metabolites diffusing into the bean [12, 14, 17, 62, 68], e.g., EtOH, LA and Ac, besides evaporation and degradation processes not directly measured, but highly likely. In this sense, low success of MIs accounting M1 can be due to a lack of the models to describe dynamics of metabolites diffusing into the inner bean. In particular, the pure degradation mechanism included in our MI does not fully account for all possible sinks (e.g., due to diffusion and evaporation) of these substances. Cases in which M1 and its iterations resulted in convergence, are those where clear decreases of EtOH, LA and Ac are visible in the time series (see Figures B.18, B.19 and B.25).

Second, regardless of the widely accepted mechanism of AAB consuming Ac once EtOH concentration has reached minimum levels in the fermenting mass [11, 18, 54, 109], we have found solid quantitative evidence through assessing M5 that such a phenomenon has a tiny impact on the process dynamics and thus is unlikely. By the end of fermentation, the AAB populations have been highly diminished and, in most of the datasets considered here, drops in Ac concentration were seldom reported. In other words, if M5 actually has an impact on the whole process, it would be necessary that AAB counts remain viable up to its completion in order to deplete Ac. This observation can also explain the few exceptions, in which M5 led to successful fits. In total, the vast majority of datasets showed minimal counts of AAB even before the penultimate day of fermentation, with the exception of *brwb1* and *brwb2* reported by Papalexandratou *et al.* [62], where drops of AAB counts are quite abrupt by its last day, limiting the capability of these MIs to simulate a complete diminution of its population (see

Figures B.8 and B.9).

Finally, after reviewing all above-mentioned causes of non-convergence and how they could have affected success ratios of each MI, it is plausible to assume that harsher combinations of such causes are responsible for failing to fit any model to 9 datasets (see Section 3.3.1) as well for some others that were scarcely described, e.g., *dowb1* and *niwb2*.

3.4.2 Interpretation of the posterior distributions of model parameters

Here we would like to discuss general aspects of the posterior distributions and how they allow us to further assess the different MIs. We will focus on the agreement of the obtained posteriors with values in the literature as well as practical non-identifiability of parameters.

For parameter agreement, let us consider the set of posterior distributions for MI(2,3) fitted over several datasets (see Table B.6). As stated in [1], among all estimated posteriors, few did not agree with reported estimates in the literature. Those which culminated in values far away from reported ranges (more than ten-fold) as well as others, whose biological plausibility is unlikely suggest evidence of practical non-identifiability, as evidenced by posteriors with wide credible intervals. This might be due to weakly informative observations that do not capture entirely the dynamics of the included mechanisms [1, 129].

This assumption seems to be supported by parameters posteriors under the scope of different MIs. By the inclusion of extra terms acting upon the dynamics, in which non-identifiable parameters are suspected to exist, their wide ranges should be visibly reduced. In fact, focussing on examples already mentioned in [1] (particularly $Y_{\mathrm{LA|AAB}}$, $Y_{\mathrm{EtOH|AAB}}$, and $Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}$), we can see this reduction across different MIs describing datasets *ghhp1*, *ecwb1*, *ecwb2* and *brpb1* (see Figure B.30).

Nevertheless, inclusion of extra parameters does not entirely eliminate non-identifiability, which suggests the need of more informative priors for further developments of cocoa bean fermentation modeling.

3.4.3 Grouping of fermentation features

We have seen that fermentation features can be distinguishable with respect of all parameter estimates derived from their posterior distributions, especially those from MI(2) and MI(2,3). From our perspective, this is a clear illustration, how ODE-based modeling, rather than the usual methods of chemical fingerprinting [130–132], can show differences in features of the process.

An example is the clear differences in fermentation features identified through kinetic parameter estimates of MI(2,3) for cultivar, controlled temperature, and fermentation methodology. Regarding cultivar, similar findings have been reported for biochemical characterization studies where same cultivars used in different countries have shown to be part of similar classes within PCA [130]. This would then explain why country of origin used as a feature, did not result in clear separation patterns. Instead, only few subsets of parameters for certain MIs ended up in clear group separations in that case, as it happens for the LAB-related parameters in MI(1,3,4) (see Section 3.3.4). This could be an indicator that indeed dynamics of different LAB populations are linked to the location where fermentation took place [110].

In a similar fashion, grouping of fermentation trials dependent on whether they were performed under controlled temperature settings, reflects how kinetic parameters of MI(2,3) might be influenced by this feature. In this regard, explicit inclusion of temperature in these models would be a natural option. From our point of view, we firmly consider that incorporation of temperature in this modeling scheme would be beneficial in case remarkable improvements on the assessing statistics presented here were seen. However, explicitly incorporating temperature

as a dynamical variable did not dramatically change either PSIS-LOO, posterior predictions or parameter ranges towards more biological plausibility (see Chapter 5).

Lastly, classification of trials with respect to methodology also tends to lead to a clear separation with a special emphasis on trials performed in stainless-steel tanks fitted with MI(2). This finding suggests that the use of stainless-steel tanks affects kinetic parameters, making them distinct from other methods. Besides, it becomes an indication that inclusion of M2 seems to drive this difference, as well as other feature discriminations. The latter observation is based on the loading plots that for all these PCAs are determined by parameters related to M2, such as $\mu_{\max}^{\mathrm{LAB_{Fru}}}$, $Y_{\mathrm{EtOH|LAB}}^{\mathrm{Fru}}$ and $Y_{\mathrm{LA|LAB}}^{\mathrm{Fru}}$.

3.5 Conclusion

The series of MIs presented here constitute a first kinetic exploration of the plausibility of regulatory dynamics of cocoa bean fermentation not considered in our previous modeling [1], but long reported and hypothesized in the literature. Thus, it allows us to evaluate the plausibility of various mechanisms in a stand-alone and concomitant manner. Among the five mechanisms discussed here, M2 (consumption of fructose by lactic acid bacteria) and M3 (production of acetic acid by yeast) have gathered the strongest support in our investigation.

Our scheme also allows us to conclude that loss of metabolites by physical phenomena (M1) is quite minimal, relative to their consumption and formation rates emphasizing the importance of microbial biochemical processes. Furthermore, it also offers quantitative evidence that a widely hypothesized mechanism, M5 (over-oxidation of acetic acid by acetic acid bacteria), does not agree with experimental data.

With reference to fermentation features, the rich set of parameter estimation results grants for interpretation on three levels: (1) We find that the parametrised time courses separate different fermentation features with different quality. Across all models, origin countries seem to only have a small influence on systematic time course differences. In contrast, temperature and cultivar seem to have a strong effect on fermentation dynamics (and hence on systematic differences in the resulting parameter vectors). (2) Orthogonal to this view, we can assess, which model versions lead to better discrimination fermentation features, compared to the basic baseline model. This complements our assessment of parameter estimation convergences, which is summarized in Table 3.4. (3) By splitting parameters into groups we can assess the involvement of certain microorganisms in the systematic differences between fermentation features.

Lastly, this work commends that in a pure sense of describing fermentation dynamics in the pulp, inclusion of temperature as a dynamical variable does not add improvements to fits obtained under the proposed scheme. However, future advancements in cocoa bean fermentation modeling might find necessary to take it into account for a more refined description of more detailed experimental data and to capture its reported importance in mediating important dynamics, for instance its role in diffusion processes of acids into the bean [133].

Ethics. No ethical considerations apply.
Data, code and materials. The datasets and codes supporting this article are deposited at: `https://github.com/mmorenozam/ecbfm`
Competing interests. The authors declare no competing interests.
Authors' contributions. M.H. and M.M. conceived the original idea. M.M. conducted simulations and analyses. M.U. contributed and endorsed the biological hypotheses derived from the research. All contributed to writing and editing of the manuscript.
Acknowledgements. We gratefully acknowledge Dr. Ben Bales, Dr. Bob Carpenter and the

rest of the Stan Development team for their advice and fruitful discussions in the Stan Forums (`https://discourse.mc-stan.org/`). M.M. acknowledges Dr. Silvana Beltrán for her support through the conduction of this research.
Funding. This work was funded by Barry Callebaut through the Cocoa Metabolomics (COMETA) project at Jacobs University Bremen.

Chapter 4

Temperature in cocoa bean fermentation kinetic modeling[6]

4.1 Introduction

From a bibliographical point of view, several authors have widely discussed the effect of temperature upon fermentation of cocoa beans. It is well known that increase in temperature within the fermenting mass results from microbial activities, mainly yeast (Y) and acetic acid bacteria (AAB) by the production of ethanol (EtOH) and acetic acid (Ac), respectively [18]. Its importance lies in several aspects, such as its role in the seed embryo death process [109], its possible involvement in AAB inhibition at final stages of the fermentation [54], and evaporation of volatile metabolites (i.e., EtOH and Ac) [17].

In regard to include temperature in a quantitative approach, López-Pérez *et al.* [23] stands as the only attempt among the few mathematical models of cocoa bean fermentation that has implemented it as another state variable using the cardinal temperature model as proposed by Rosso *et al.* [134]. Opposed to this approach, here we present the use of the Arrhenius equation applied to maximum specific growth and mortality rates coupled with the mathematical expression accounting for temperature produced by microbial dynamics as stated by Boulton [135].

4.2 Formulation

4.2.1 Heat transfer

According to Boulton [135], the temperature rate change in a fermentation process can be considered as an energy balance between the rates of heat generation of the fermenting mass and its surroundings. Thus, this rate is:

$$\frac{dT}{dt} = \frac{\Delta Q}{\rho\, C_{\mathrm{p}}\, V}\,\frac{dS}{dt} - \frac{U A}{\rho\, C_{\mathrm{p}}\, V}\,(T - T_e)\ , \tag{4.1}$$

where, for our case, ΔQ is the heat released during the fermentation per mg of substrate ($\frac{kJ}{\mathrm{mg}_{\mathrm{substrate}}}$), ρ is the density of cocoa beans ($\frac{\mathrm{kg}_{\mathrm{bean}}}{m^3}$), C_{p} is the cocoa beans heat capacity ($\frac{kJ}{\mathrm{mg}_{\mathrm{bean}} K}$), V is the volume of the fermenting mass (m^3), T_e is the environmental temperature (K), U is the overall heat transfer coefficient ($\frac{kJ}{m^2 K}$), A is the contact area of the fermenting mass with the environment (m^2), and $\frac{dS}{dt}$ is the substrate rate change. With the purpose of

[6]This chapter is based on the Supplementary material of Moreno-Zambrano *et al.* [2], Section 4.

facilitating any optimization routine, $\frac{\Delta Q}{\rho\, C_{\mathrm{p}}\, V}$ and $\frac{UA}{\rho\, C_{\mathrm{p}}\, V}$ can be comprised into single parameters each. Then, let's define a substrate-to-heat yield coefficient ($Y_{Q|\mathrm{substrate}}$) such as:

$$Y_{Q|\mathrm{substrate}} = \frac{\Delta Q}{\rho\, C_{\mathrm{p}}\, V}\,, \tag{4.2}$$

and a heat-loss (Q_L) parameter as:

$$Q_L = \frac{UA}{\rho\, C_{\mathrm{p}}\, V}\,; \tag{4.3}$$

hence, Eq. (4.1) can be re-written as:

$$\frac{dT}{dt} = Y_{Q|\mathrm{substrate}}\, \frac{dS}{dt} - Q_L\, (T - T_e)\,. \tag{4.4}$$

The inclusion of Eq. (4.4) into the baseline model proposed in Eqs. (2.1) to (2.8) is then reduced to appropriately replace $\frac{dS}{dt}$. For this, we account the total release of heat product of the growth of all microbial groups involved. With this in mind, Eq. (4.4) can take the form:

$$\begin{aligned}\frac{dT}{dt} = {} & Y_{Q|\mathrm{Glc}}\, \left(Y_{\mathrm{Glc|Y}}\, v_1 + Y_{\mathrm{Glc|LAB}}\, v_3\right) + Y_{Q|\mathrm{Fru}}\, \left(Y_{\mathrm{Fru|Y}}\, v_2\right) + Y_{Q|\mathrm{EtOH}}\, \left(Y_{\mathrm{EtOH|AAB}}\, v_4\right) \\ & + Y_{Q|\mathrm{LA}}\, \left(Y_{\mathrm{LA|AAB}}\, v_5\right) - Q_L\, (T - T_e)\,,\end{aligned} \tag{4.5}$$

where $\frac{dS}{dt}$ is replaced by the necessary growth rates and yield coefficients explained in Tables 2.1 and 2.2, $Y_{Q|\mathrm{Glc}}$, $Y_{Q|\mathrm{Fru}}$, $Y_{Q|\mathrm{EtOH}}$ and $Y_{Q|\mathrm{LA}}$, are the glucose-to-heat, fructose-to-heat, ethanol-to-heat and lactic acid-to-heat yield coefficients, respectively.

4.2.2 Temperature effect on kinetic parameters

Among several ways to include temperature in bacterial growth models, the Arrhenius equation is one of the most widely utilized. Regardless its limitations in interpretation and recommended application only in sub-optimal growth conditions [136], here we introduce its use as a first approximation to look at how temperature might affect kinetic parameters of our baseline model.

Thereby, considering that Eq. (4.5) adds 6 parameters to baseline model, a simple manner to relate kinetic parameters to temperature would be by besetting Arrhenius terms to as few as possible and thus reducing risks of over-parametrization.

Then, original Arrhenius equation is expressed as:

$$k(T) = A\, e^{-\frac{Ea}{RT}} \tag{4.6}$$

as in Eqs. (4.2) and (4.3), energy of activation (Ea) and gas constant (R) can be comprised into a single term:

$$k(T) = A\, e^{-\frac{E}{T}} \tag{4.7}$$

assuming that all growth and mortality rates share common energies of activation E_1 and E_2, respectively, these can be then rewritten in function of temperature as:

$$\mu_{\mathrm{max}}^{\mathrm{Y_{Glc}}} = \mu_1\, e^{-\frac{E_1}{T}}$$
$$\mu_{\mathrm{max}}^{\mathrm{Y_{Fru}}} = \mu_2\, e^{-\frac{E_1}{T}}$$

$$
\begin{aligned}
\mu_{\max}^{\mathrm{LAB_{Glc}}} &= \mu_3\, e^{-\frac{E_1}{T}} \\
\mu_{\max}^{\mathrm{AAB_{EtOH}}} &= \mu_4\, e^{-\frac{E_1}{T}} \\
\mu_{\max}^{\mathrm{AAB_{LA}}} &= \mu_5\, e^{-\frac{E_1}{T}} \\
k_{\mathrm{Y}} &= k_1\, e^{-\frac{E_2}{T}} \\
k_{\mathrm{LAB}} &= k_2\, e^{-\frac{E_2}{T}} \\
k_{\mathrm{AAB}} &= k_3\, e^{-\frac{E_2}{T}}\ ,
\end{aligned}
$$

where μ_i and k_i represent the growth rates of corresponding microorganisms and mortality rates at exponential phases.

Finally, above expressions are incorporated into baseline model, Eqs. (2.1) to (2.8).

4.3 Model's simulations

Parameters' posterior distributions for a model as described above can be computed using the same Bayesian framework and priors as proposed in Sections 1.5.3 and 2.2.4, respectively. The only addition to this scheme is the need of using an extra parameter not considered before. We refer to term T_e in Eqs. (4.1) and (4.5), which represents environmental temperature. For this, we treated T_e as another free parameter with a more informative prior. Assuming that temperature during the fermentation process had a mean value of 26°C, we can add random noise by assigning a prior of the form:

$$T_e \sim \mathcal{N}(26,\, 0.01) \tag{4.8}$$

which, after time series scaling by their maximum value becomes:

$$T_e \sim \mathcal{N}\left(\frac{26}{T_{\max}},\, \frac{0.01}{T_{\max}}\right)\ , \tag{4.9}$$

where $T_{\max}$ is the maximum value observed in each temperature time series.

4.4 Results

As mentioned in the main manuscript, addition of temperature to the model does not improve fits in terms of their PSIS-LOO, posterior predictions and adjustment of problematic parameters' ranges to more plausible values. Here, we present each of these claims on basis of calibrating aforementioned model to datasets *ghhp1* [12] (know for accounting non-identifiable parameters), *brwb1* and *brwb2* [62].

4.4.1 PSIS-LOO deviance values

As depicted in figure 4.1, PSIS-LOO deviance values between baseline model (MI(0)) and baseline model added temperature as state variable (MI(0+t)), there are no significant improvements in prediction adequacy between them both over datasets *ghhp1* [12], *brwb1* and *brwb2* [62].

4.4.2 Posterior predictions

Besides MI(0+t) being capable of fitting remarkably well temperature, no outstanding enhancement of posterior predictions for rest of state variable is perceptible with its counterparts with

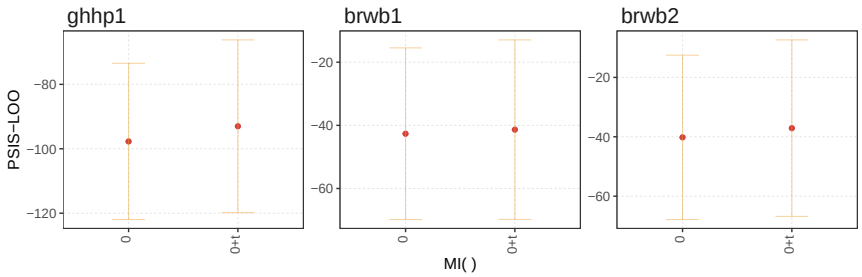


Figure 4.1: PSIS-LOO deviance values of baseline model (MI(0)) and baseline model added temperature (MI(0+t)) over datasets *ghhp1* [12], *brwb1* and *brwb2* [62].

no temperature interactions. See figures 4.2, 4.3 and 4.4 for posterior predictions of *ghhp1*, *brwb1* and *brwb2*, respectively.

4.4.3 Parameters' ranges

Finally, regarding ranges for parameters' likely affected by non-identifiability, as it can be seen in Table 4.1, do not change in a great extend nor fall within reported ranges in literature.

4.5 Conclusion

From this rather simple approach of including temperature in our baseline model, no significant improvements are clearly seen on the basis of statistics used for assessing the models used in this work.

Table 4.1: Parameter estimates of baseline model (MI(0)) and baseline model added temperature (MI(0+t)) calibrated over datasets *ghhp1* [12], *brwb1* and *brwb2* [62]. Means and standard deviations (within parenthesis) of the posterior parameters where temperature was not explicitly included are shown. All reported values in literature have same units as depicted in table 2.MM with the exception of substrate saturation constants (mg(substrate) mL^{-1}). Green colored cells indicate parameter estimates that lie within reported ranges. Orange colored cells indicate parameter estimates that are far from reported ranges by less than the estimate either divided or multiplied by 10. Parameters fitted with the temperature model are indicated with (t) next to dataset coding.

Parameter	ghhp1	ghhp1 (t)	brwb1	brwb1 (t)	brwb2	brwb2 (t)	Reported values	Reference
K^{Y}_{Glc}	35.178 (13.809)	36.343 (13.275)	34.607 (14.961)	36.069 (15.006)	30.097 (10.79)	30.704 (10.528)	9.73×10^{-4} - 0.5	[82, 84]
K^{Y}_{Fru}	35.682 (15.148)	36.708 (13.806)	25.172 (13.874)	27.683 (13.632)	41.026 (17.479)	27.152 (16.461)	5.82 - 161.45	[92, 137, 138]
K^{LAB}_{Glc}	37.949 (12.291)	35.876 (11.578)	31.821 (14.1)	21.171 (10.839)	18.936 (10.595)	15.779 (9.424)	0.790 - 178.0	[88–90]
K^{AAB}_{EtOH}	16.201 (5.559)	15.063 (5.355)	3.796 (1.666)	3.832 (1.585)	4.104 (1.941)	4.197 (1.965)		
K^{AAB}_{LA}	2468.808 (1246.01)	2504.518 (1230.219)	314.451 (153.855)	330.123 (154.135)	80.936 (41.545)	81.353 (41.802)		
$Y_{Glc\|Y}$	33.833 (11.286)	35.245 (10.925)	241.469 (57.414)	247.991 (56.39)	120.043 (35.606)	128.349 (34.287)	1.56 - 66.67	[78–82, 100, 101]
$Y_{Glc\|LAB}$	29.094 (10.903)	26.525 (10.147)	19.96 (12.78)	18.558 (12.482)	3.289 (3.391)	3.604 (3.381)	4.50 - 200	[90, 99]
$Y_{Fru\|Y}$	40.913 (11.183)	43.431 (10.847)	243.705 (47.332)	248.639 (46.601)	231.885 (63.675)	274.426 (58.585)	43.48 - 200	[78, 91]
$Y^{Glc}_{EtOH\|Y}$	7.575 (4.505)	7.351 (4.386)	11.924 (6.043)	11.921 (5.984)	8.333 (4.507)	9.46 (4.559)	1.39 - 21.49	[81–83, 98]
$Y^{Fru}_{EtOH\|Y}$	5.87 (3.301)	6.707 (3.516)	11.102 (4.928)	11.575 (5.07)	5.886 (3.897)	6.6 (4.034)	5.679 - 5.896	[83]
$Y_{EtOH\|AAB}$	1330.24 (734.365)	1441.682 (886.113)	380.088 (151.185)	425.236 (162.987)	171.249 (55.576)	178.163 (55.176)	8.06 - 166.67	[94, 96, 102]
$Y^{Glc}_{LA\|LAB}$	10.573 (1.518)	10.651 (1.507)	2.781 (0.665)	2.853 (0.666)	2.128 (0.783)	2.169 (0.751)		
$Y_{LA\|AAB}$	1923.213 (1182.916)	1862.392 (1191.561)	286.443 (141.585)	287.649 (137.542)	62.138 (38.388)	61.628 (37.531)	7.51 - 8.34	[96]
$Y^{Glc}_{Ac\|LAB}$	5.586 (0.907)	5.65 (0.902)	3.249 (1.094)	3.344 (1.094)	2.879 (1.856)	2.973 (1.83)		
$Y^{EtOH}_{Ac\|AAB}$	107.838 (113.797)	100.5 (132.648)	579.784 (271.612)	550.655 (276.068)	321.488 (136.809)	312.075 (133.57)		
$Y^{LA}_{Ac\|AAB}$	1418.047 (868.763)	1390.92 (860.409)	665.158 (348.199)	634.929 (339.7)	385.605 (182.323)	383.586 (179.92)		
σ	0.149 (0.01)	0.142 (0.009)	0.168 (0.014)	0.161 (0.013)	0.167 (0.013)	0.162 (0.012)		

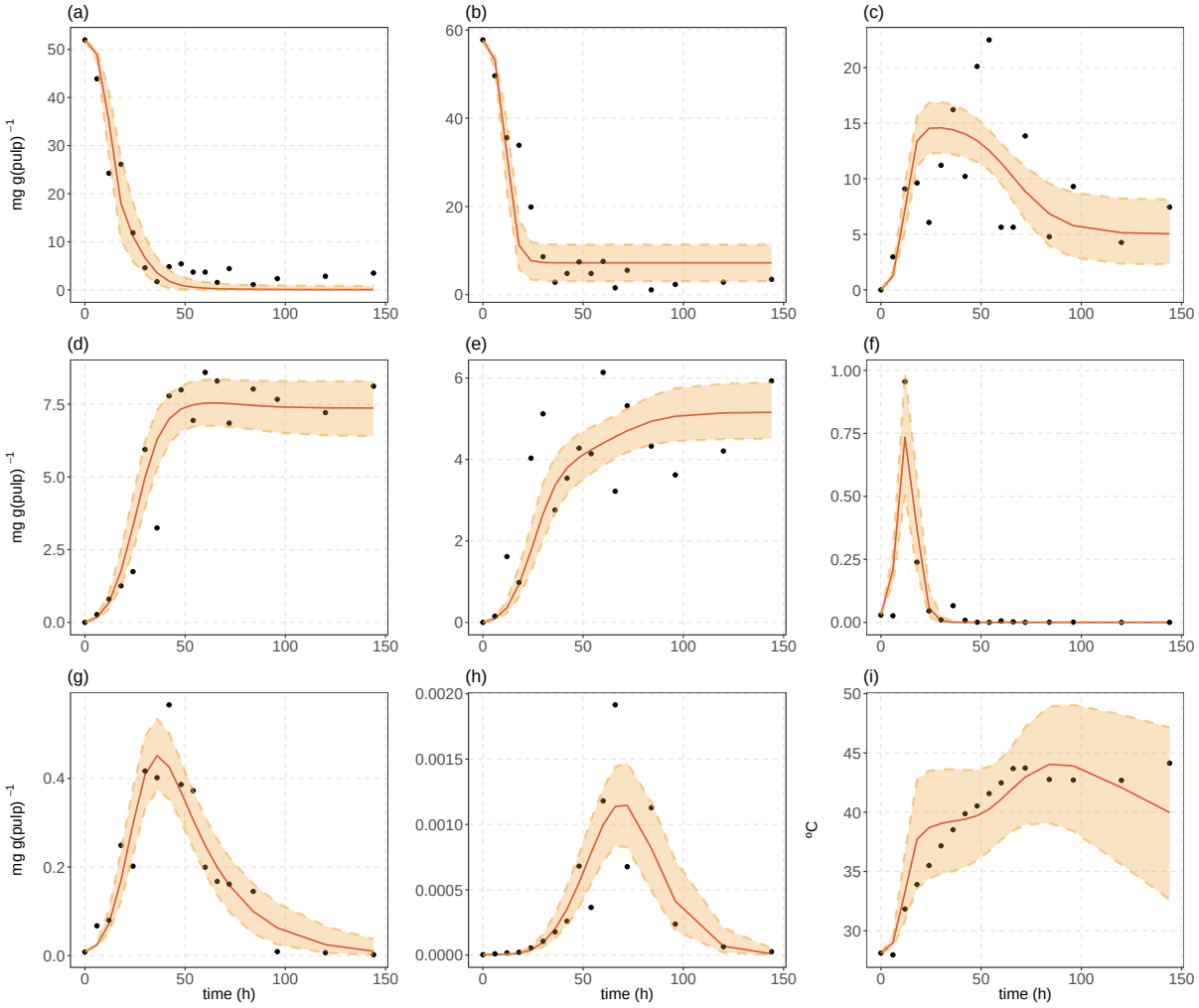


Figure 4.2: Posterior predictions of MI(0+t) over dataset *ghhp1* reported by Camu *et al.* [12]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. (i) Temperature. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

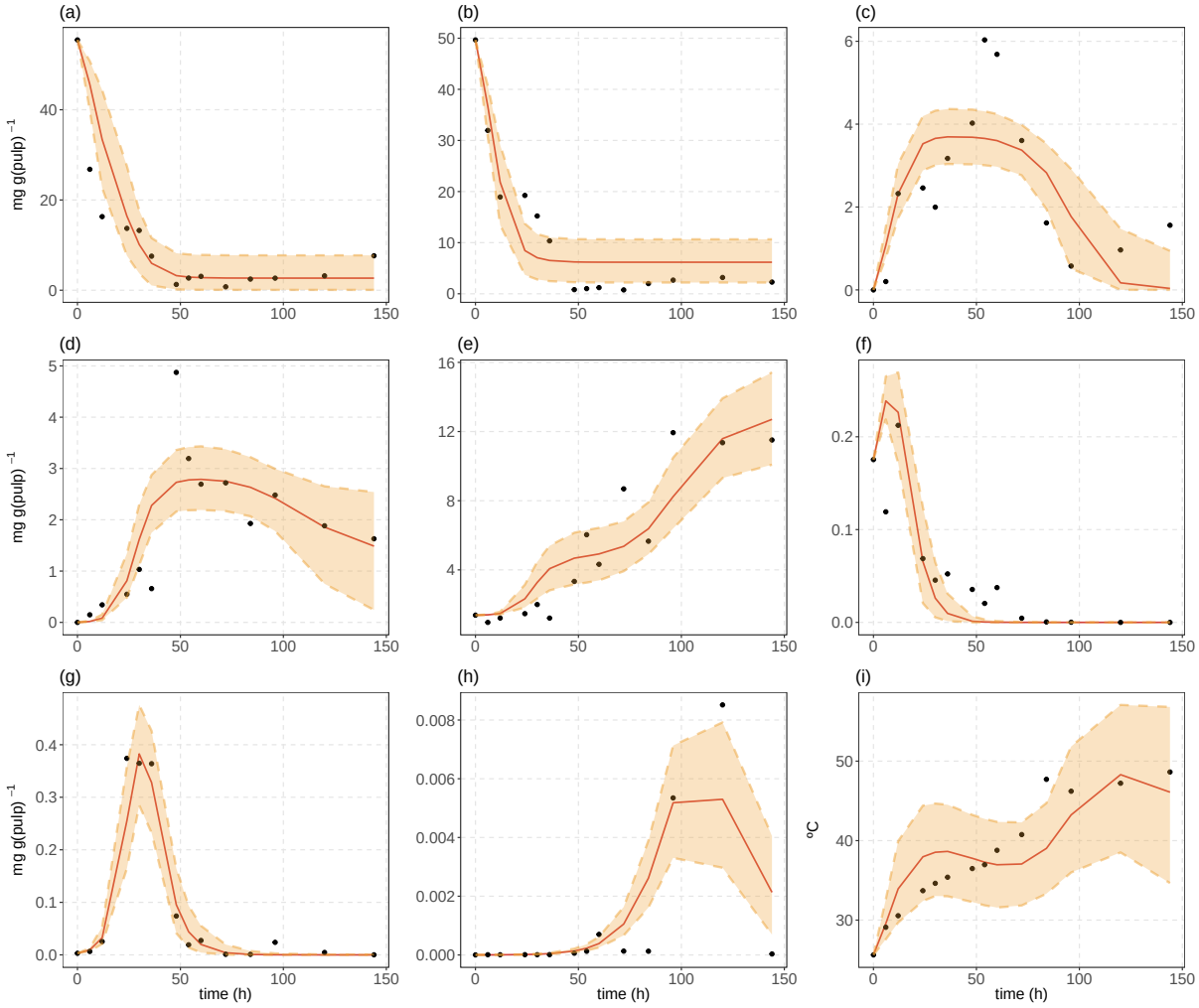


Figure 4.3: Posterior predictions of MI(0+t) over dataset *brwb1* reported by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. (i) Temperature. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

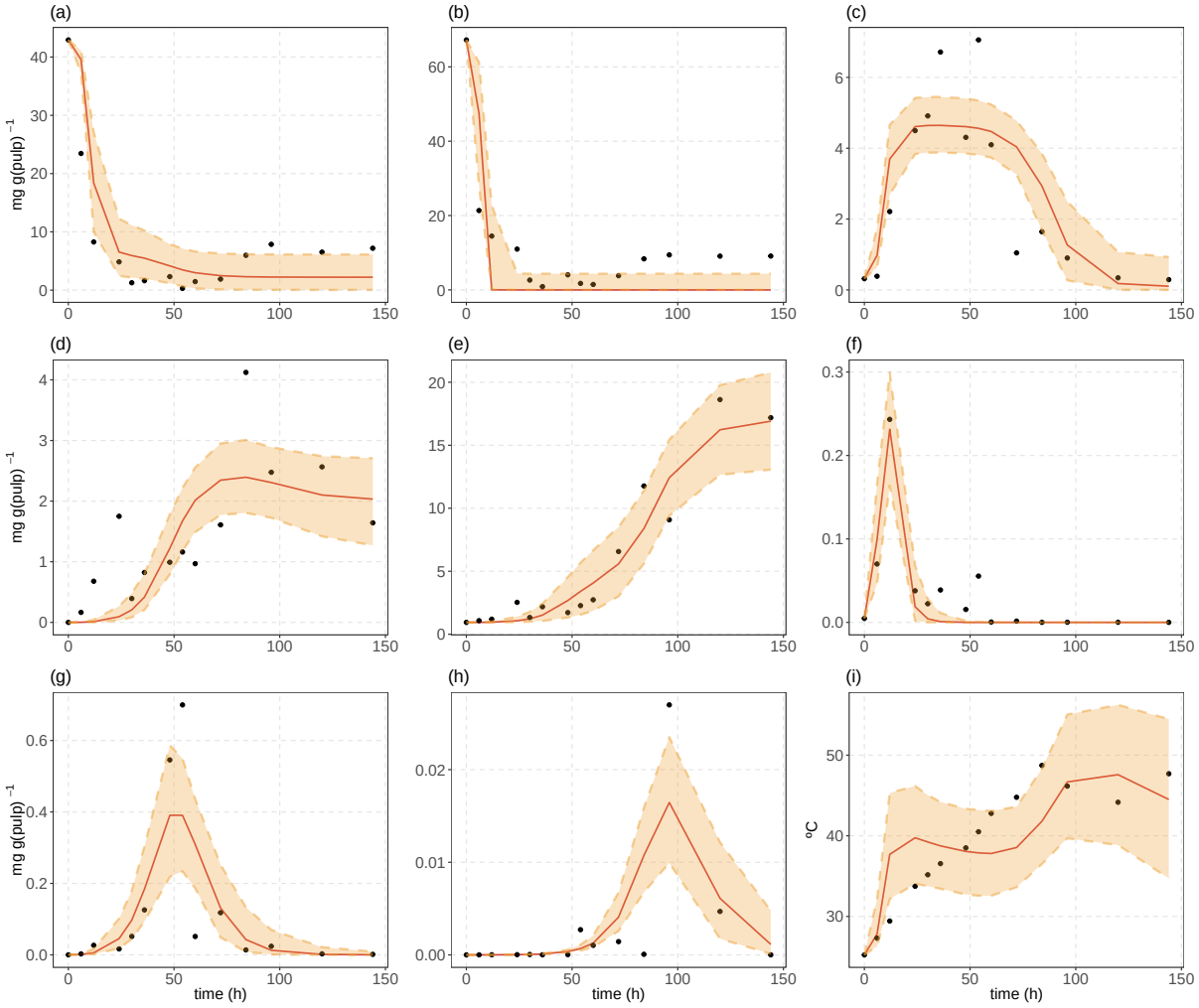


Figure 4.4: Posterior predictions of MI(0+t) over dataset *brwb2* reported by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. (i) Temperature. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

Chapter 5

Conclusions and outlook

As I have described it in previous chapters, cocoa bean fermentation represents a biological system more complicated than other typical fermentation processes in the food industry. Besides variation given by methodologies in its conduction, the major difference with simpler fermentations lies on the sequential growth of three dominating microbial groups which interact between each other as well with their environment in a bundle of ways that are still not entirely known. This imposes a varied group of obstacles for its quantitative representation due to, among others, heterogeneity of the fermenting mass, difficulty in measuring microbial populations separately, and little control over environmental conditions; which combined, result in complicated trial setups with low sampling rates and considerable experimental errors.

For such reasons, mathematical modeling of the process becomes a complicated task that, before this work, has been rarely performed [22, 23]. Instead, a considerable bibliography regarding modeling of post-processing steps of cocoa beans is available, e.g., drying kinetics [58, 59, 139, 140] and moisture adsorption [141], along with elucidation of microbial interactions by metabolic flux balance analysis over simulated conditions [19, 60, 142].

As a result, improvement in cocoa bean fermentation direct control and optimization has been appraised by qualitative descriptions only, opposed to other food fermentation processes where their understanding by modeling techniques have significantly enhanced their performance in industrial settings, as it is the case for wine [54, 143]. In this sense, this thesis has described a novel way in modeling the process by the use of Bayesian methods that could I will discuss from two broad perspectives: (1) their implications in modeling cocoa bean fermentation, and (2) challenges of their use towards refinement of further model iterations.

5.1 Implications of a Bayesian framework in modelling cocoa bean fermentation

In recent years, application of Bayesian inference in solving biological problems has gained acceptance in disciplines such as ecology, epidemiology [35], and lately, food science [144]. This rise in its popularity is thanks to the increment in machine's computational power and their proven efficiency in handling short and noisy time series usually found in such disciplines [35, 103].

Contrary to commonly used estimation methods, e.g., maximum-likelihood, ordinary least squares and non-linear least squares, Bayesian algorithms, e.g., MCMC-NUTS and Gibbs sampler, are less prone to suffer from non-identifiability issues in solving parameter estimation problems. On the one hand, classical methods sometimes lead to unrealistic parameter values ranging from estimates collapsed in zero to highly inflated because of their poor capacity of dealing with large experimental noise forcing them to stop at local minima regions in the

parameter space [103, 145].

On the other hand, a consequence of the formulation of the Bayes' rule, where the assumption of a vector of parameters, θ, giving rise to observational data, $\mathcal{Y}$, permits the determination of a region in the parameter space where these agree with $\mathcal{Y}$. This distinctive trait explains the growing use in Biology of Bayesian over frequentist approaches in three main aspects: (1) chances of parameters' non-identifiability because of convergence over local minima of the parameter space are reduced to some extent by a more thoughtfully exploration thanks to their sampling algorithms, (2) the uncertainty got around the parameters' posterior distributions can be interpreted in a more natural way as probabilities, and (3) the possibility of incorporate and update parameters of a model with prior information [146].

With this in mind, obtained parameter posterior distributions can serve to understand biological aspects of the system subject of modeling. In this precise case, the posterior distributions of parameters have served for discussing, besides their biological plausibility, about the likeliness of occurrence of certain mechanisms and how physical features between fermentation trials could have affected them (see Chapter 2).

In addition, under the assumption that practical non-identifiable parameters do not play an extremely important role in models' convergence, assessing candidate models using a simple statistic of their success (ESR) describing several datasets from distinct trials, constitutes a unique contribution of this thesis in extending the Bayesian paradigm into a systematic exploration scheme of hypotheses previously reported in literature. Hence, I present the novel possibility of using posterior estimates of models fitted over fermentation trials' data as an alternative to chemical fingerprinting of cocoa beans (see Chapter 3).

Furthermore, the Bayesian framework here employed, demonstrates its suitable use for assessing the influence of temperature over the prediction accuracy and improvement of biological interpretability of the parameters (see Chapter 4) concluding that for at least these two aspects, its inclusion in the modeling does not bring advantage in neither of both.

All together, this work shows the worthiness of using Bayesian methods for parameter estimation, fermentation dynamics simulations and interpretability from a biological perspective that has allowed a deeper understanding of the system. Specifically, methodologies such as convergence statistics (i.e., $\widehat{R}$ and ESS), and model comparison measures (i.e., PSIS-LOO and WAIC) that allowed to define first, a fully functional baseline model, and second, a scheme of candidate mechanisms' exploration for elucidating the likeliness of occurrence and co-occurrence of long-time hypothesized interactions. However, one should know a plethora of challenges remain still open to be solved, as I will discuss in the following section.

5.2 Outlook

From findings of this work, it is noticeable that further research in the mechanistic understanding of cocoa bean fermentation should turn around the improvement of both experimental conditions in which trials take place, and mathematical tools for its modeling.

On one side, as thoroughly discussed in this thesis, the spontaneous way in which the fermentation of cocoa beans takes place is undoubtedly the major source of experimental impediments in its further comprehension. This is clear in several ways, such as presence of weakly informative observations and outliers which reflect a highly heterogeneity of the system, combined in most cases with low sample rates. Thus, reducing variation (i.e., outliers and weakly informative observations) is a task that would be hardly solved *in situ* for spontaneous trials, considering that within a solid state fermentation matrix, homogenous samples are seldom workable. However, less separated measurements along the fermentation could bring richer information about its dynamics in a better time resolution that would lower the impact of outlying observations.

In the same regard, separate characterizations of individual components in the process would also bring light to its better understanding. Taking as an example the network representation of the baseline model (Figure 2.2), the entire system turns around of three major components, namely Y, LAB and AAB. For each of them, independent kinetic studies could be performed under controlled conditions in order to determine in better detail more interactions and equally important, parameter estimates of the system. Fortunately, initial steps in this direction have started. For instance, Kouamé *et al.* [147] have developed a first model for alcoholic fermentation of cocoa beans under laboratory conditions, where kinetic parameter estimates have been obtained from better quality data by using a liquid simulated fermentation matrix.

From the perspective of improving mathematical approaches, we can consider two main courses of further action. First, optimization of experimental conditions in fermentation trials and conduction of modeling by separate components of the system would bring sufficient information to in fact take advantage of one primary goal of Bayesian inference, which is the possibility to feed the model with prior knowledge. As previously discussed in Chapters 2 and Chapter 3, parameters suspect of suffering from practical non-identifiability are likely the result of both weakly informative priors, and observations. Thus, by constraining the parameter space with more informed priors and less noisy measurements will improve a Bayesian framework not only from a computational point of view (e.g., reduction of simulation times), but as an extension to plan further hypotheses in cases whether here proposed mechanisms agree no more with new collected evidence.

Second, the rest of challenges to consider are more in hand with possible different formulations of the Bayesian framework and current software and computational limitations. Regarding a possibility of different formulations of the problem, how Bayes' rule works, leaves the door open to a diverse number of combinations of likelihood and prior distributions to test with the aim to obtain better posterior distributions for the parameters and predictions. For instance, using a Beta likelihood function in Eq. (1.25) instead of a normal one would increase predictive accuracy by allowing MCMC to sample from the posterior nearer to outlying observations thanks to the heavier tails of the former. However, these experiments with distinct likelihood and prior distributions needs special attention into the aspect of its implementation that might require different parametrizations of the model.

Finally, regarding software and computational matters, Bayesian inference is still a field in continuous development where certain aspects have not yet been solved. In the specific case of ODE systems, their implementation used to be cumbersome in the few available Bayesian software before the release of Stan [50] which introduced built-in ODE solvers along with its previously described sampling algorithms (see Section 1.5.3). In general, MCMC sampling methods can take considerable amounts of time until their convergence depending on the model's complexity (e.g., models implemented in this work took between $\approx$2 to $\approx$72 hours of running time).

Consequently, parallel computation routines of MCMC have been developed in Stan to overcome this obstacle. Nevertheless, until the conduction of this work, their application was quite hard to accomplish and currently, even with the last release of Stan with new and easier-to-implement parallel computation functions, it still relies on accessibility of the user to powerful computational machinery such as multi-core cluster facilities [148]. In terms of cocoa bean fermentation, the aforesaid leads to expect larger computational times whether more complexity is added to the model (e.g., compartmentalization would easily need two-fold number of parameters) leading to an even slower process in testing new mechanisms, unless proper parallelization is achieved.

5.3 Final conclusion

In summary, this thesis has introduced a novel mathematical modeling approach in cocoa bean fermentation showing besides its implementation, their potential use in exploring regulatory mechanisms with an important emphasis on biological plausibility, and development of further fingerprinting methods. Also, it makes up a basis over which future iterations improving modeling are possible based on using prior information from new findings regarding experimentally obtained parameter estimates.

Appendix A

A mathematical model of coca bean fermentation – Supplementary material[7]

A.1 Geometric derivation of conversion factors between CFU to mass

According to the Bergey's Manual of Systematic Bacteriology [64, 65], the diameter of *Lactobacillus plantarum* (representative species of lactic acid bacteria in cocoa bean fermentation) is between 0.9 to 1.2 µm and its length between 3 to 8 µm. For the case of the *Acetobacter* genus, its usual diameter is between 0.6 to 0.9 µm and its length is between 1.0 to 4.0 µm.

Given the volume of a spherocylinder as

$$V = \pi\, r^2 \left(\frac{4}{3} r + a\right) ,$$

a midpoint from the above ranges can be computed in order to determine a theoretical volume for a single cell, where a is equal to the length minus two times the radius (r). Thus, for lactic acid bacteria (LAB) the volume is equal to 4.46 µm^3 and for acetic acid bacteria (AAB) it is $\approx$ 1 µm^3. Therefore; using as reference the dry weight of a single cell of *E. coli* reported by Neidhardt & Curtiss [66] of 2.8×10^{-10} mg and its volume of 1 µm^3 [67], the conversion factor between CFU to dry biomass of LAB and AAB were determined as 1.25 and 0.28 pg CFU^{-1} respectively.

A.2 Pre-modeling

The modeling choices of using a Contois [29] equation as growth rate of AAB on LA, and non-linear mortality rates, are the result of a pre-modeling stage where several pre-model versions were assessed for different mechanisms. Here, 3 pre-models from this stage are presented and compared with the final model version discussed in Chapter 2.

Pre-model 1 (PM1)

In a first instance, pre-model 1 (PM1) is proposed in an oversimplified manner. For PM1, the main substrates (Glc and Fru) were combined in a single state variable named 'monosaccharides' (M). Furthermore, it comprehends simple coupled interactions among the state variables reduced to single uptake of substrates and production of metabolites. A network diagram of PM1 is shown in Figure A.1.

[7] This chapter is based on the Supplementary material of Moreno-Zambrano *et al.* [1]

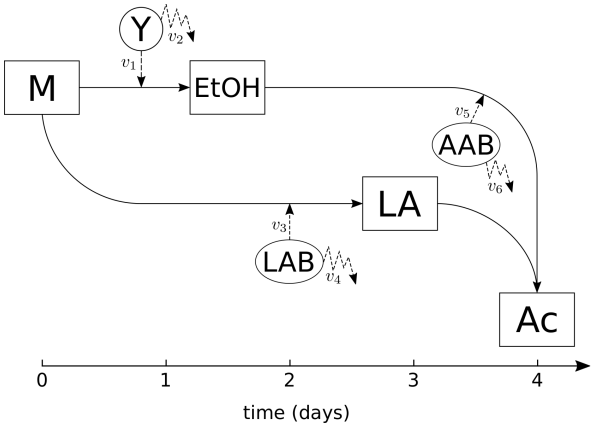


Figure A.1: Network diagram of the cocoa bean fermentation pre-model 1 (PM1). Microbial groups: yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB) are represented as circles. Metabolites: monosaccharides (M), ethanol (EtOH), lactic acid (LA) and acetic acid (Ac) are represented as squares. The growth rates of yeast on monosaccharides (v_1), of lactic acid bacteria on monosaccharides (v_3), and of acetic acid bacteria on ethanol (v_5) are represented as straight dashed arrows. The mortality rates of yeast (v_2), lactic acid bacteria (v_4), and acetic acid bacteria (v_6) are represented as zigzag dashed arrows. Solid straight arrows show the direction in which the conversion of metabolites occurs.

The system of ODE's for PM1 comprises Monod [27] equations for the growth rates and first-order reaction kinetics for the mortality rates, as shown in Table A.1.

Table A.1: Growth and mortality rate equations for the cocoa bean fermentation pre-model (PM1). Microbial groups are represented as yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB). Substrates are represented as monosaccharides (M) and ethanol (EtOH). Biomass and concentration of substrates, both are shown within square brackets [ ]. Maximum specific growth rates $\mu_{\max}$ are shown of the form $\mu^i_{\max}$, where i can be either Y, LAB and AAB. 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 M or EtOH. 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
$v_1 = \dfrac{\mu^{\mathrm{Y}}_{\max}\,[\mathrm{M}]}{[\mathrm{M}] + K^{\mathrm{Y}}_{\mathrm{M}}}\,[\mathrm{Y}]$	$v_2 = k_{\mathrm{Y}}\,[\mathrm{Y}]$
$v_3 = \dfrac{\mu^{\mathrm{LAB}}_{\max}\,[\mathrm{M}]}{[\mathrm{M}] + K^{\mathrm{LAB}}_{\mathrm{M}}}\,[\mathrm{LAB}]$	$v_4 = k_{\mathrm{LAB}}\,[\mathrm{LAB}]$
$v_5 = \dfrac{\mu^{\mathrm{AAB}}_{\max}\,[\mathrm{EtOH}]}{[\mathrm{EtOH}] + K^{\mathrm{AAB}}_{\mathrm{EtOH}}}\,[\mathrm{AAB}]$	$v_6 = k_{\mathrm{AAB}}\,[\mathrm{AAB}]$

In this way, PM1 is constructed as a system of ODE's with seven state variables with six yield coefficients, as expressed in Eqs. (A.1) to (A.7).

$$\frac{d\,[\mathrm{M}]}{dt} = -Y_{\mathrm{M|Y}}\,v_1 - Y_{\mathrm{M|LAB}}\,v_3 \tag{A.1}$$

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y_{\mathrm{EtOH|Y}}\,v_1 - Y_{\mathrm{EtOH|AAB}}\,v_5 \tag{A.2}$$

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\, v_3 \tag{A.3}$$

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y_{\mathrm{Ac|AAB}}\, v_5 \tag{A.4}$$

$$\frac{d\,[\mathrm{Y}]}{dt} = v_1 - v_2 \tag{A.5}$$

$$\frac{d\,[\mathrm{LAB}]}{dt} = v_3 - v_4 \tag{A.6}$$

$$\frac{d\,[\mathrm{AAB}]}{dt} = v_5 - v_6 \tag{A.7}$$

Pre-models 2 and 3 (PM2 and PM3)

For the formulation of pre-model 2 (PM2) and pre-model 3 (PM3), the 'monosaccharides' state variable in PM1 was considered by their separate components (Glc and Fru) and the addition of a growth rate of AAB on LA. In this manner, pre-models PM2 and PM3 are similar to the final model stated in Eqs. (2.1) to (2.8) by having 8 state variables. However, PM2 and PM3 differ with the latter regarding possible mechanisms of growth as well as mortality rates used.

In the case of PM2, all comprehended growth rates were expressed as Monod [27] terms and maintaining simple first-order reaction kinetics for mortality rates (a mortality rate constant affecting the concentration of a certain microbial group). With respect to PM3, the only difference with PM2 lies on the use of second- and third-order reaction kinetics with the correspondent products that each microbial group produces in the process [32]. A representation in the form of network diagrams of these two pre-models can be depicted in Figure A.2 in panel (a) for PM2 and panel (b) for PM3. Notice how the graphical representation of PM3 corresponds to the one of the final model stated in Chapter 2.

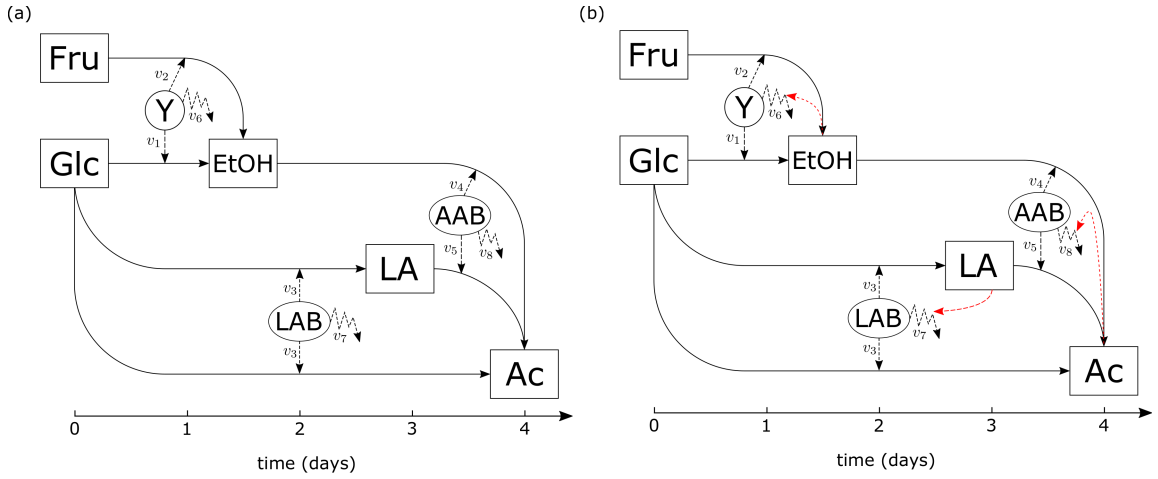


Figure A.2: Network diagram of the cocoa bean fermentation model. (a) Considering linear mortality rates. (b) Considering non-linear mortality rates (red dashed arrows). Microbial groups: yeast (Y), lactic acid bacteria (LAB) and acetic acid bacteria (AAB) are represented as circles. Metabolites: Glucose (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 lactic 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. Solid straight arrows show the direction in which the conversion of metabolites occur.

Final model (FM)

The main difference between pre-models PM2, PM3 and the model discussed in Chapter 2, from her called final model (FM), is given by the use of another growth rate term for AAB on LA. Specifically for FM, the Contois [29] equation was taken into account as candidate for this growth rate (equation v_5, in Table 2.1). This assumption relies on the following premise: given that few species of AAB are capable of catabolize LA [21, 60], it can be argued that the growth rate of these species is a function of their population size. In that manner, the cell growth of AAB on LA is reduced along the population size of these particular AAB species increases during the fermentation process.

The characteristics of each the previous model versions, including FM, is shown in Table A.2.

Table A.2: Summary of pre-models (PM) including final model (FM). Check-marks and cross-marks indicates whether the model includes use of multiple substrates (Glc and Fru), product toxicity interactions (non-linear mortality rates) and population size effect for the consumption of lactic acid (LA) in the form of a Contois term for the growth of acetic acid bacteria (AAB) on LA.

Model	Multiple substrate	Product toxicity	Population size effect for LA consumption
PM1	✗	✗	✗
PM2	✓	✗	✗
PM3	✓	✓	✗
FM	✓	✓	✓

Models' comparison

Pre-models as defined previously were assessed with the FM by fitting them to experimental data reported by Camu *et al.* [12] and Papalexandratou *et al.* [62]. Such a comparison was performed by deviance values obtained for PSIS-LOO and widely applicable information criterion (WAIC), as described by Vehtari *et al.* [48] and Watanabe [47], respectively.

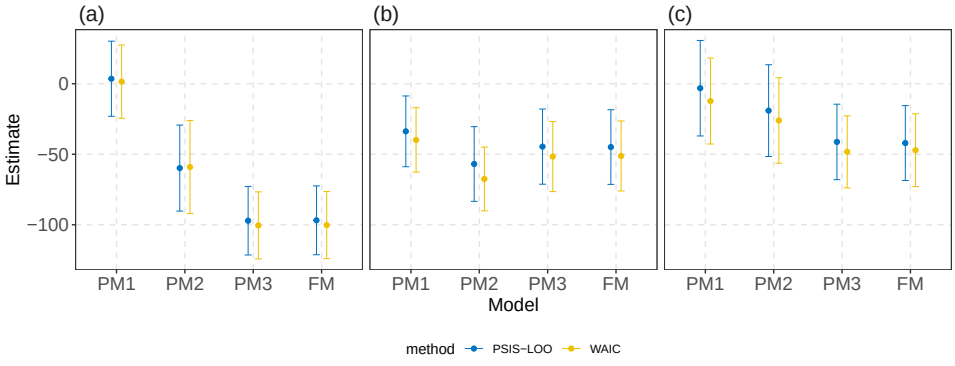


Figure A.3: Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO) and widely applicable information criterion (WAIC) for pre-models PM1, PM2, PM3 and final model FM. Dots represents the estimates for PSIS-LOO and WAIC, lines represent their standard error. (a) Camu *et al.* [12] dataset, (b) Box 1 of Papalexandratou *et al.* [62] and (c) Box 2 of Papalexandratou *et al.* [62].

In this fashion, the estimated values of PSIS-LOO and WAIC (Figure A.3) for PM1 have worse performances than for the rest of models. PM3 and FM showed better results for all

datasets, with the exception of data corresponding to Box 1 (Figure A.3, panel (b)) where PM2 showed a better performance.

However, FM seems to be the most plausible model among the three datasets because PM3 showed a bimodal posterior distribution for the maximum specific growth rate of AAB on EtOH ($\mu_{\max}^{\mathrm{AAB_{EtOH}}}$) with the data from Camu *et al.* [12] as shown in Figure A.4.

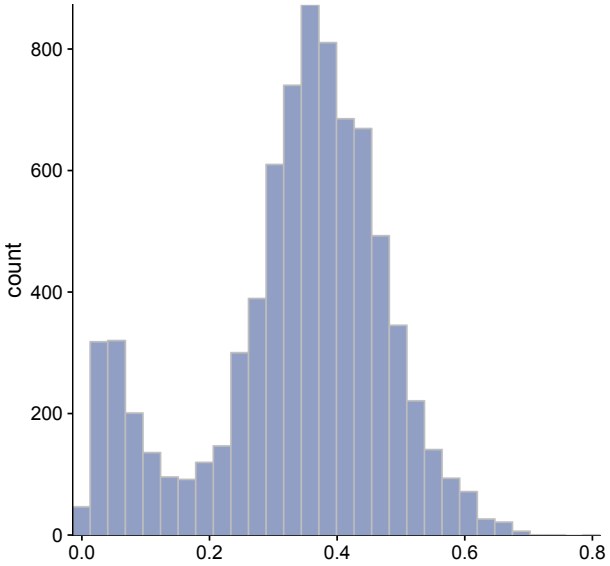


Figure A.4: Posterior distribution of the maximum specific growth rate of AAB on EtOH ($\mu_{\max}^{\mathrm{AAB_{EtOH}}}$) of pre-model PM3 with the data from Camu *et al.* [12].

A.3 Analytical determination of conversion factors for parameter estimates

Consider the kinetics of any x state variable in Eqs. (2.1) to (2.8), its concentration $[x]$ in a certain time and the maximum concentration $[x]_{max}$ within the given time interval in which the fermentation process took place.

Let's x' represent a transformation of the estate variable x expressed as

$$x' = \frac{[x]}{[x]_{max}}, \tag{A.8}$$

then, the time derivative of x' will be given by

$$\frac{dx'}{dt} = \frac{1}{[x]_{max}} \frac{d\,[x]}{dt}, \tag{A.9}$$

where $[x]_{max}$ is either the maximum concentration of glucose, fructose, ethanol, lactic acid, acetic acid, yeast, lactic acid bacteria or acetic acid bacteria, depending on which ODE of the proposed model is expressed in the form of Eq. (A.9).
Generalizing, by expressing all ODEs of the model in the form of Eq. (A.9), one gets a system of equations as

$$\frac{1}{[\mathrm{Glc}]_{max}}\frac{d\,[\mathrm{Glc}]}{dt} = -Y_1\,\frac{\mu_1\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_1}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}} - Y_2\,\frac{\mu_3\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_3}\,\frac{[\mathrm{LAB}]}{[\mathrm{LAB}]_{max}}, \tag{A.10}$$

$$\frac{1}{[\mathrm{Fru}]_{max}}\frac{d\,[\mathrm{Fru}]}{dt} = -Y_3\,\frac{\mu_2\,\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}}{\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}+Ks_2}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}}, \tag{A.11}$$

$$\begin{aligned}\frac{1}{[\mathrm{EtOH}]_{max}}\frac{d\,[\mathrm{EtOH}]}{dt} &= Y_4\,\frac{\mu_1\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_1}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}} + Y_5\,\frac{\mu_2\,\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}}{\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}+Ks_2}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}} \\ &\quad - Y_6\,\frac{\mu_4\,\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}}{\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}+Ks_4}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}},\end{aligned} \tag{A.12}$$

$$\frac{1}{[\mathrm{LA}]_{max}}\frac{d\,[\mathrm{LA}]}{dt} = Y_7\,\frac{\mu_3\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_3}\,\frac{[\mathrm{LAB}]}{[\mathrm{LAB}]_{max}} - Y_8\,\frac{\mu_5\,\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}}{\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}+Ks_5\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}}}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}}, \tag{A.13}$$

$$\begin{aligned}\frac{1}{[\mathrm{Ac}]_{max}}\frac{d\,[\mathrm{Ac}]}{dt} &= Y_9\,\frac{\mu_3\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_3}\,\frac{[\mathrm{LAB}]}{[\mathrm{LAB}]_{max}} + Y_{10}\,\frac{\mu_4\,\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}}{\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}+Ks_4}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}} \\ &\quad + Y_{11}\,\frac{\mu_5\,\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}}{\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}+Ks_5\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}}}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}},\end{aligned} \tag{A.14}$$

$$\frac{1}{[\mathrm{Y}]_{max}}\frac{d\,[\mathrm{Y}]}{dt} = \frac{\mu_1\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_1}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}} + \frac{\mu_2\,\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}}{\frac{[\mathrm{Fru}]}{[\mathrm{Fru}]_{max}}+Ks_2}\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}} - k_1\,\frac{[\mathrm{Y}]}{[\mathrm{Y}]_{max}}\,\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}, \tag{A.15}$$

$$\frac{1}{[\mathrm{LAB}]_{max}}\frac{d\,[\mathrm{LAB}]}{dt} = \frac{\mu_3\,\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}}{\frac{[\mathrm{Glc}]}{[\mathrm{Glc}]_{max}}+Ks_3}\,\frac{[\mathrm{LAB}]}{[\mathrm{LAB}]_{max}} - k_2\,\frac{[\mathrm{LAB}]}{[\mathrm{LAB}]_{max}}\,\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}, \tag{A.16}$$

$$\begin{aligned}\frac{1}{[\mathrm{AAB}]_{max}}\frac{d\,[\mathrm{AAB}]}{dt} &= \frac{\mu_4\,\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}}{\frac{[\mathrm{EtOH}]}{[\mathrm{EtOH}]_{max}}+Ks_4}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}} + \frac{\mu_5\,\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}}{\frac{[\mathrm{LA}]}{[\mathrm{LA}]_{max}}+Ks_5\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}}}\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}} \\ &\quad - k_3\,\frac{[\mathrm{AAB}]}{[\mathrm{AAB}]_{max}}\,\frac{[\mathrm{Ac}]^2}{[\mathrm{Ac}]^2_{max}},\end{aligned} \tag{A.17}$$

which constitutes a scaled version of the proposed model, where μ_i, Ks_i, Y_i and k_i are the solutions of the ODEs upon their correspondent scaled time series (according to Eq. (A.8))

for specific maximum growth rates, saturation constants, yield coefficients and mortality rates, respectively.

Once the parameters of Eqs. (A.10) to (A.17) are estimated as a result of the Bayesian optimization routine, obtaining conversion factors to their actual units is accomplished by properly working out the equations in such a way that the maximum values of each time series are multiplied by the parameters' scaled solutions.

For instance, Eq. (A.10) can be simplified to:

$$\frac{d\,[\mathrm{Glc}]}{dt} = -Y_1 \frac{[\mathrm{Glc}]_{max}}{[\mathrm{Y}]_{max}} \frac{\mu_1\,[\mathrm{Glc}]}{[\mathrm{Glc}] + Ks_1\,[\mathrm{Glc}]_{max}}\,[\mathrm{Y}] - Y_2 \frac{[\mathrm{Glc}]_{max}}{[\mathrm{LAB}]_{max}} \frac{\mu_3\,[\mathrm{Glc}]}{[\mathrm{Glc}] + Ks_3\,[\mathrm{Glc}]_{max}}\,[\mathrm{LAB}]\,, \tag{A.18}$$

where the terms μ_1, μ_3, $Ks_1\,[\mathrm{Glc}]_{max}$, $Ks_2\,[\mathrm{Glc}]_{max}$, $Y_1\,\frac{[\mathrm{Glc}]_{max}}{[\mathrm{Y}]_{max}}$ and $Y_2\,\frac{[\mathrm{Glc}]_{max}}{[\mathrm{LAB}]_{max}}$ are equivalent to $\mu_{\max}^{\mathrm{Y_{Glc}}}$, $\mu_{\max}^{\mathrm{LAB_{Glc}}}$, $K_{\mathrm{Glc}}^{\mathrm{Y}}$, $K_{\mathrm{Glc}}^{\mathrm{LAB}}$, $Y_{\mathrm{Glc|Y}}$ and $Y_{\mathrm{Glc|LAB}}$ respectively and the left hand side of the equation corresponds to the state variable with no scaling.

Thus, similarly to Eq. (A.18), conversion factors between the scaled parameters to its actual units can be determined as

$$\begin{aligned}
\mu_{\max}^{\mathrm{Y_{Glc}}} &= \mu_1 \\
\mu_{\max}^{\mathrm{Y_{Fru}}} &= \mu_2 \\
\mu_{\max}^{\mathrm{LAB_{Glc}}} &= \mu_3 \\
\mu_{\max}^{\mathrm{AAB_{EtOH}}} &= \mu_4 \\
\mu_{\max}^{\mathrm{AAB_{LA}}} &= \mu_5 \\
K_{\mathrm{Glc}}^{\mathrm{Y}} &= Ks_1\,[\mathrm{Glc}]_{\max} \\
K_{\mathrm{Fru}}^{\mathrm{Y}} &= Ks_2\,[\mathrm{Fru}]_{\max} \\
K_{\mathrm{Glc}}^{\mathrm{LAB}} &= Ks_3\,[\mathrm{Glc}]_{\max} \\
K_{\mathrm{EtOH}}^{\mathrm{AAB}} &= Ks_4\,[\mathrm{EtOH}]_{\max} \\
K_{\mathrm{LA}}^{\mathrm{AAB}} &= Ks_5 \frac{[\mathrm{LA}]_{\max}}{[\mathrm{AAB}]_{\max}} \\
k_{\mathrm{Y}} &= \frac{k_1}{[\mathrm{EtOH}]_{\max}} \\
k_{\mathrm{LAB}} &= \frac{k_2}{[\mathrm{LA}]_{\max}} \\
k_{\mathrm{AAB}} &= \frac{k_3}{[\mathrm{Ac}]_{\max}^2} \\
Y_{\mathrm{Glc|Y}} &= Y_1 \frac{[\mathrm{Glc}]_{\max}}{[\mathrm{Y}]_{\max}}
\end{aligned}
\qquad
\begin{aligned}
Y_{\mathrm{Glc|LAB}} &= Y_2 \frac{[\mathrm{Glc}]_{\max}}{[\mathrm{LAB}]_{\max}} \\
Y_{\mathrm{Fru|Y}} &= Y_3 \frac{[\mathrm{Fru}]_{\max}}{[\mathrm{Y}]_{\max}} \\
Y_{\mathrm{EtOH|Y}}^{\mathrm{Glc}} &= Y_4 \frac{[\mathrm{EtOH}]_{\max}}{[\mathrm{Y}]_{\max}} \\
Y_{\mathrm{EtOH|Y}}^{\mathrm{Fru}} &= Y_5 \frac{[\mathrm{EtOH}]_{\max}}{[\mathrm{Y}]_{\max}} \\
Y_{\mathrm{EtOH|AAB}} &= Y_6 \frac{[\mathrm{EtOH}]_{\max}}{[\mathrm{AAB}]_{\max}} \\
Y_{\mathrm{LA|LAB}}^{\mathrm{Glc}} &= Y_7 \frac{[\mathrm{LA}]_{\max}}{[\mathrm{LAB}]_{\max}} \\
Y_{\mathrm{LA|AAB}} &= Y_8 \frac{[\mathrm{LA}]_{\max}}{[\mathrm{AAB}]_{\max}} \\
Y_{\mathrm{Ac|LAB}}^{\mathrm{Glc}} &= Y_9 \frac{[\mathrm{Ac}]_{\max}}{[\mathrm{LAB}]_{\max}} \\
Y_{\mathrm{Ac|AAB}}^{\mathrm{EtOH}} &= Y_{10} \frac{[\mathrm{Ac}]_{\max}}{[\mathrm{AAB}]_{\max}} \\
Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}} &= Y_{11} \frac{[\mathrm{Ac}]_{\max}}{[\mathrm{AAB}]_{\max}}
\end{aligned}$$

In this way, the prior θ in Eq. (2.9) is equivalent to sample from an unscaled prior distribution determined by the conversion factors above derived. For each of the k parameters we then have

$$\theta_{k_u} \sim \mathcal{N}(c_k\,0.5,\ c_k\,0.3),\ \theta_{k_u} > 0, \tag{A.19}$$

where θ_{k_u} represents the unscaled prior distribution for parameter k in θ and c is the conversion factor for each k parameter.

In other words, Eq. (A.19) defines the original unscaled ranges of the parameters in which the scaled priors for θ in Eq. (2.9) are equivalently sampled. A summary of the original ranges, from which the parameters are sampled, is given as unscaled prior distributions in Table A.3.

Table A.3: Prior distributions in the range of the original units of the parameters of the cocoa bean fermentation model for the data of Camu *et al.* [12] and the fermentation boxes 1 and 2 of Papalexandratou *et al.* [62]. All priors are constrained in the positive set of real numbers.

Parameter	Camu	Box 1	Box 2
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	$\mathcal{N}(25.9, 15.6)$	$\mathcal{N}(27.7, 16.6)$	$\mathcal{N}(21.5, 12.9)$
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	$\mathcal{N}(28.9, 17.3)$	$\mathcal{N}(24.8, 14.9)$	$\mathcal{N}(33.6, 20.2)$
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	$\mathcal{N}(25.9, 15.6)$	$\mathcal{N}(27.7, 16.6)$	$\mathcal{N}(21.5, 12.9)$
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	$\mathcal{N}(11.2, 6.8)$	$\mathcal{N}(3.0, 1.8)$	$\mathcal{N}(3.5, 2.1)$
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	$\mathcal{N}(2243.7, 1346.2)$	$\mathcal{N}(286.2, 171.7)$	$\mathcal{N}(76.4, 45.8)$
k_{Y}	$\mathcal{N}(0.02, 0.01)$	$\mathcal{N}(0.08, 0.05)$	$\mathcal{N}(0.07, 0.04)$
k_{LAB}	$\mathcal{N}(0.06, 0.03)$	$\mathcal{N}(0.10, 0.06)$	$\mathcal{N}(0.12, 0.07)$
k_{AAB}	$\mathcal{N}(0.01, 0.008)$	$\mathcal{N}(0.003, 0.002)$	$\mathcal{N}(0.001, 0.0009)$
$Y_{\mathrm{Glc\|Y}}$	$\mathcal{N}(27.2, 16.3)$	$\mathcal{N}(130.6, 78.4)$	$\mathcal{N}(88.2, 52.9)$
$Y_{\mathrm{Glc\|LAB}}$	$\mathcal{N}(45.9, 27.5)$	$\mathcal{N}(74.2, 44.5)$	$\mathcal{N}(30.7, 18.4)$
$Y_{\mathrm{Fru\|Y}}$	$\mathcal{N}(30.2, 18.1)$	$\mathcal{N}(116.9, 70.2)$	$\mathcal{N}(138.2, 82.9)$
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	$\mathcal{N}(11.8, 7.1)$	$\mathcal{N}(14.2, 8.5)$	$\mathcal{N}(14.5, 8.7)$
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	$\mathcal{N}(11.8, 7.1)$	$\mathcal{N}(14.2, 8.5)$	$\mathcal{N}(14.5, 8.7)$
$Y_{\mathrm{EtOH\|AAB}}$	$\mathcal{N}(5872.7, 3523.6)$	$\mathcal{N}(354.3, 212.6)$	$\mathcal{N}(130.7, 78.4)$
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	$\mathcal{N}(7.6, 4.6)$	$\mathcal{N}(6.5, 3.9)$	$\mathcal{N}(2.9, 1.8)$
$Y_{\mathrm{LA\|AAB}}$	$\mathcal{N}(2243.7, 1346.2)$	$\mathcal{N}(286.2, 171.7)$	$\mathcal{N}(76.4, 45.8)$
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	$\mathcal{N}(5.4, 3.3)$	$\mathcal{N}(15.9, 9.6)$	$\mathcal{N}(13.3, 7.9)$
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	$\mathcal{N}(1602.9, 961.7)$	$\mathcal{N}(700.9, 420.5)$	$\mathcal{N}(345.1, 207.0)$
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	$\mathcal{N}(1602.9, 961.7)$	$\mathcal{N}(700.9, 420.5)$	$\mathcal{N}(345.1, 207.0)$

A.4 Model's diagnostics

Table A.4: Scaled posterior moments and quantiles of parameter estimates for the Camu *et al.* [12] dataset. Number of effective sample size and $\widehat{R}$ statistic are also shown.

Parameter	Mean	SE mean	sd	2.5%	25%	50%	75%	97.5%	n-eff	$\widehat{R}$
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	0.253	0.002	0.094	0.098	0.184	0.242	0.314	0.465	3369	1.00
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	0.358	0.002	0.106	0.173	0.282	0.352	0.429	0.584	4280	1.00
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	0.358	0.001	0.067	0.237	0.311	0.354	0.402	0.500	5427	1.00
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	0.380	0.001	0.092	0.214	0.317	0.376	0.437	0.577	4194	1.00
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	0.008	0.000	0.012	0.000	0.002	0.005	0.009	0.038	1597	1.00
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	0.680	0.003	0.266	0.181	0.493	0.669	0.858	1.215	8000	1.00
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	0.615	0.003	0.264	0.125	0.424	0.607	0.791	1.157	6066	1.00
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	0.731	0.003	0.238	0.294	0.560	0.718	0.887	1.220	5931	1.00
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	0.714	0.003	0.251	0.252	0.539	0.705	0.879	1.236	6190	1.00
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	0.559	0.003	0.275	0.076	0.358	0.544	0.748	1.132	8000	1.00
k_{Y}	0.748	0.002	0.114	0.544	0.669	0.741	0.820	0.989	5105	1.00
k_{LAB}	0.047	0.000	0.012	0.028	0.038	0.045	0.053	0.076	3816	1.00
k_{AAB}	0.259	0.001	0.053	0.167	0.223	0.255	0.292	0.377	5838	1.00
$Y_{\mathrm{Glc\|Y}}$	0.614	0.003	0.207	0.275	0.461	0.592	0.745	1.057	4527	1.00
$Y_{\mathrm{Glc\|LAB}}$	0.319	0.002	0.118	0.128	0.235	0.304	0.390	0.588	4277	1.00
$Y_{\mathrm{Fru\|Y}}$	0.680	0.002	0.185	0.382	0.544	0.658	0.796	1.085	5325	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	0.316	0.003	0.192	0.025	0.169	0.295	0.438	0.747	5586	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	0.252	0.002	0.142	0.024	0.145	0.238	0.342	0.569	5744	1.00
$Y_{\mathrm{EtOH\|AAB}}$	0.110	0.001	0.054	0.039	0.076	0.101	0.133	0.236	2407	1.00
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	0.700	0.002	0.102	0.517	0.628	0.695	0.763	0.919	4186	1.00
$Y_{\mathrm{LA\|AAB}}$	0.430	0.003	0.271	0.022	0.215	0.402	0.614	1.010	8000	1.00
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	0.517	0.001	0.085	0.363	0.458	0.513	0.573	0.698	4553	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	0.032	0.001	0.031	0.001	0.013	0.026	0.044	0.097	2640	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	0.445	0.004	0.270	0.026	0.233	0.423	0.625	1.030	5728	1.00
σ	0.149	0.000	0.010	0.131	0.142	0.148	0.156	0.171	8000	1.00

Table A.5: Scaled posterior moments and quantiles of parameter estimates for the Papalexandratou *et al.* [62] dataset of Box 1. Number of effective sample size and $\widehat{R}$ statistic are also shown.

Parameter	Mean	SE mean	sd	2.5%	25%	50%	75%	97.5%	n-eff	$\widehat{R}$
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	0.063	0.000	0.025	0.024	0.046	0.060	0.077	0.122	5307	1.00
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	0.083	0.000	0.030	0.038	0.062	0.079	0.100	0.154	5569	1.00
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	0.414	0.001	0.096	0.246	0.346	0.408	0.477	0.617	4494	1.00
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	0.150	0.001	0.051	0.061	0.113	0.148	0.184	0.258	4873	1.00
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	0.025	0.000	0.017	0.002	0.012	0.021	0.034	0.065	4384	1.00
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	0.619	0.003	0.270	0.123	0.425	0.615	0.800	1.170	8000	1.00
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	0.504	0.004	0.286	0.040	0.281	0.483	0.697	1.101	5535	1.00
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	0.571	0.004	0.250	0.123	0.390	0.558	0.737	1.093	4238	1.00
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	0.633	0.003	0.271	0.125	0.440	0.628	0.823	1.170	8000	1.00
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	0.545	0.003	0.267	0.074	0.350	0.531	0.721	1.104	8000	1.00
k_{Y}	0.312	0.001	0.072	0.203	0.261	0.302	0.352	0.479	5425	1.00
k_{LAB}	0.310	0.002	0.089	0.159	0.250	0.304	0.362	0.509	3086	1.00
k_{AAB}	0.061	0.000	0.027	0.022	0.042	0.056	0.076	0.130	4145	1.00
$Y_{\mathrm{Glc\|Y}}$	0.922	0.003	0.229	0.495	0.765	0.911	1.072	1.390	8000	1.00
$Y_{\mathrm{Glc\|LAB}}$	0.136	0.002	0.089	0.017	0.073	0.118	0.177	0.362	3166	1.00
$Y_{\mathrm{Fru\|Y}}$	1.044	0.002	0.198	0.689	0.904	1.034	1.172	1.456	8000	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	0.420	0.003	0.213	0.052	0.261	0.406	0.562	0.865	5429	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	0.394	0.003	0.176	0.063	0.267	0.393	0.515	0.745	4619	1.00
$Y_{\mathrm{EtOH\|AAB}}$	0.534	0.003	0.215	0.169	0.380	0.517	0.671	0.996	5536	1.00
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	0.214	0.001	0.051	0.137	0.179	0.207	0.241	0.333	3420	1.00
$Y_{\mathrm{LA\|AAB}}$	0.502	0.003	0.247	0.078	0.321	0.483	0.662	1.024	8000	1.00
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	0.103	0.000	0.035	0.044	0.079	0.098	0.123	0.182	4363	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	0.411	0.003	0.194	0.075	0.270	0.399	0.536	0.820	4561	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	0.476	0.003	0.255	0.040	0.287	0.458	0.644	1.014	8000	1.00
σ	0.168	0.000	0.014	0.143	0.158	0.167	0.176	0.198	8000	1.00

Table A.6: Scaled posterior moments and quantiles of parameter estimates for the Papalexandratou *et al.* [62] dataset of Box 2. Number of effective sample size and $\widehat{R}$ statistic are also shown.

Parameter	Mean	SE mean	sd	2.5%	25%	50%	75%	97.5%	n-eff	$\widehat{R}$
$\mu_{\max}^{\mathrm{Y_{Glc}}}$	0.368	0.002	0.131	0.162	0.269	0.354	0.447	0.664	3979	1.00
$\mu_{\max}^{\mathrm{Y_{Fru}}}$	0.572	0.002	0.151	0.301	0.465	0.562	0.671	0.887	4323	1.00
$\mu_{\max}^{\mathrm{LAB_{Glc}}}$	0.499	0.002	0.152	0.227	0.388	0.492	0.600	0.815	4436	1.00
$\mu_{\max}^{\mathrm{AAB_{EtOH}}}$	0.168	0.001	0.052	0.079	0.130	0.165	0.202	0.280	4578	1.00
$\mu_{\max}^{\mathrm{AAB_{LA}}}$	0.022	0.000	0.016	0.001	0.009	0.019	0.031	0.060	4019	1.00
$K_{\mathrm{Glc}}^{\mathrm{Y}}$	0.699	0.003	0.254	0.231	0.519	0.688	0.867	1.213	8000	1.00
$K_{\mathrm{Fru}}^{\mathrm{Y}}$	0.615	0.003	0.257	0.132	0.434	0.609	0.788	1.142	6094	1.00
$K_{\mathrm{Glc}}^{\mathrm{LAB}}$	0.449	0.004	0.247	0.046	0.263	0.430	0.615	0.972	4583	1.00
$K_{\mathrm{EtOH}}^{\mathrm{AAB}}$	0.575	0.004	0.281	0.073	0.369	0.567	0.767	1.143	5840	1.00
$K_{\mathrm{LA}}^{\mathrm{AAB}}$	0.537	0.003	0.271	0.064	0.334	0.528	0.715	1.101	8000	1.00
k_{Y}	0.649	0.003	0.175	0.343	0.519	0.644	0.772	0.994	3747	1.00
k_{LAB}	0.283	0.002	0.103	0.106	0.210	0.275	0.347	0.510	3237	1.00
k_{AAB}	0.128	0.001	0.053	0.053	0.091	0.119	0.155	0.255	5813	1.00
$Y_{\mathrm{Glc\|Y}}$	0.678	0.003	0.199	0.350	0.532	0.660	0.804	1.109	5925	1.00
$Y_{\mathrm{Glc\|LAB}}$	0.054	0.001	0.056	0.002	0.016	0.037	0.073	0.209	4101	1.00
$Y_{\mathrm{Fru\|Y}}$	0.841	0.003	0.228	0.448	0.676	0.828	0.989	1.322	6484	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Glc}}$	0.283	0.002	0.155	0.025	0.168	0.271	0.384	0.621	4052	1.00
$Y_{\mathrm{EtOH\|Y}}^{\mathrm{Fru}}$	0.207	0.002	0.134	0.011	0.104	0.190	0.292	0.507	4731	1.00
$Y_{\mathrm{EtOH\|AAB}}$	0.652	0.003	0.215	0.296	0.496	0.630	0.787	1.132	6241	1.00
$Y_{\mathrm{LA\|LAB}}^{\mathrm{Glc}}$	0.363	0.002	0.136	0.185	0.268	0.333	0.424	0.717	3402	1.00
$Y_{\mathrm{LA\|AAB}}$	0.412	0.003	0.252	0.027	0.215	0.386	0.575	0.968	5647	1.00
$Y_{\mathrm{Ac\|LAB}}^{\mathrm{Glc}}$	0.109	0.001	0.071	0.011	0.060	0.095	0.142	0.286	4521	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{EtOH}}$	0.466	0.003	0.192	0.096	0.337	0.462	0.589	0.866	4820	1.00
$Y_{\mathrm{Ac\|AAB}}^{\mathrm{LA}}$	0.558	0.003	0.257	0.083	0.377	0.551	0.730	1.080	8000	1.00
σ	0.168	0.000	0.013	0.144	0.158	0.167	0.176	0.195	8000	1.00

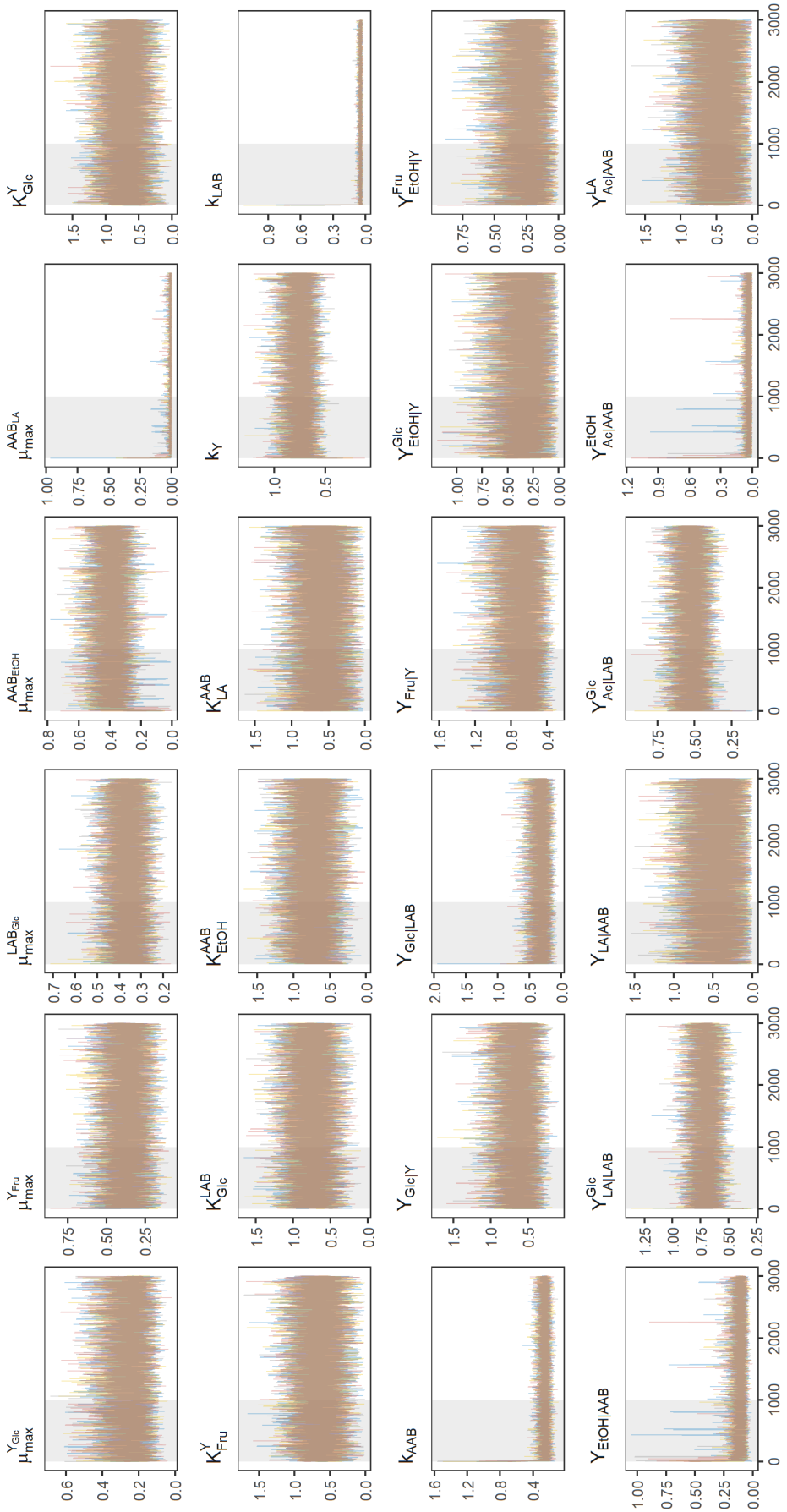


Figure A.5: Scaled trace plots for the 24 parameters of the model estimated from the dataset of Camu *et al.* [12]. Different color lines, represent different MCMC. The shadowed area represents the 1000 iterations used as warm-up.

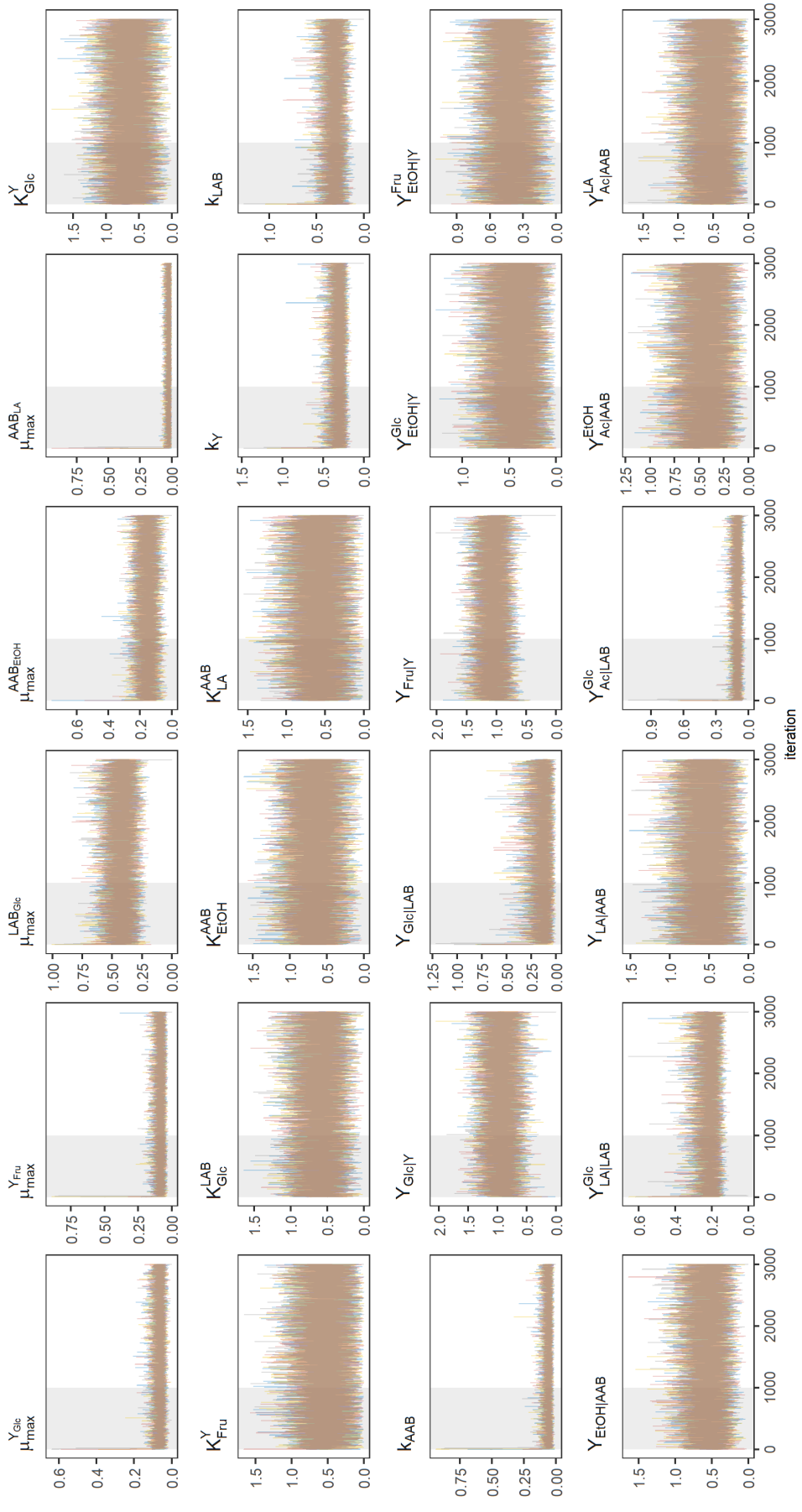


Figure A.6: Scaled trace plots for the 24 parameters of the model estimated from the dataset of Box 1 of Papalexandratou *et al.* [62]. Different color lines, represent different MCMC. The shadowed area represents the 1000 iterations used as warm-up.

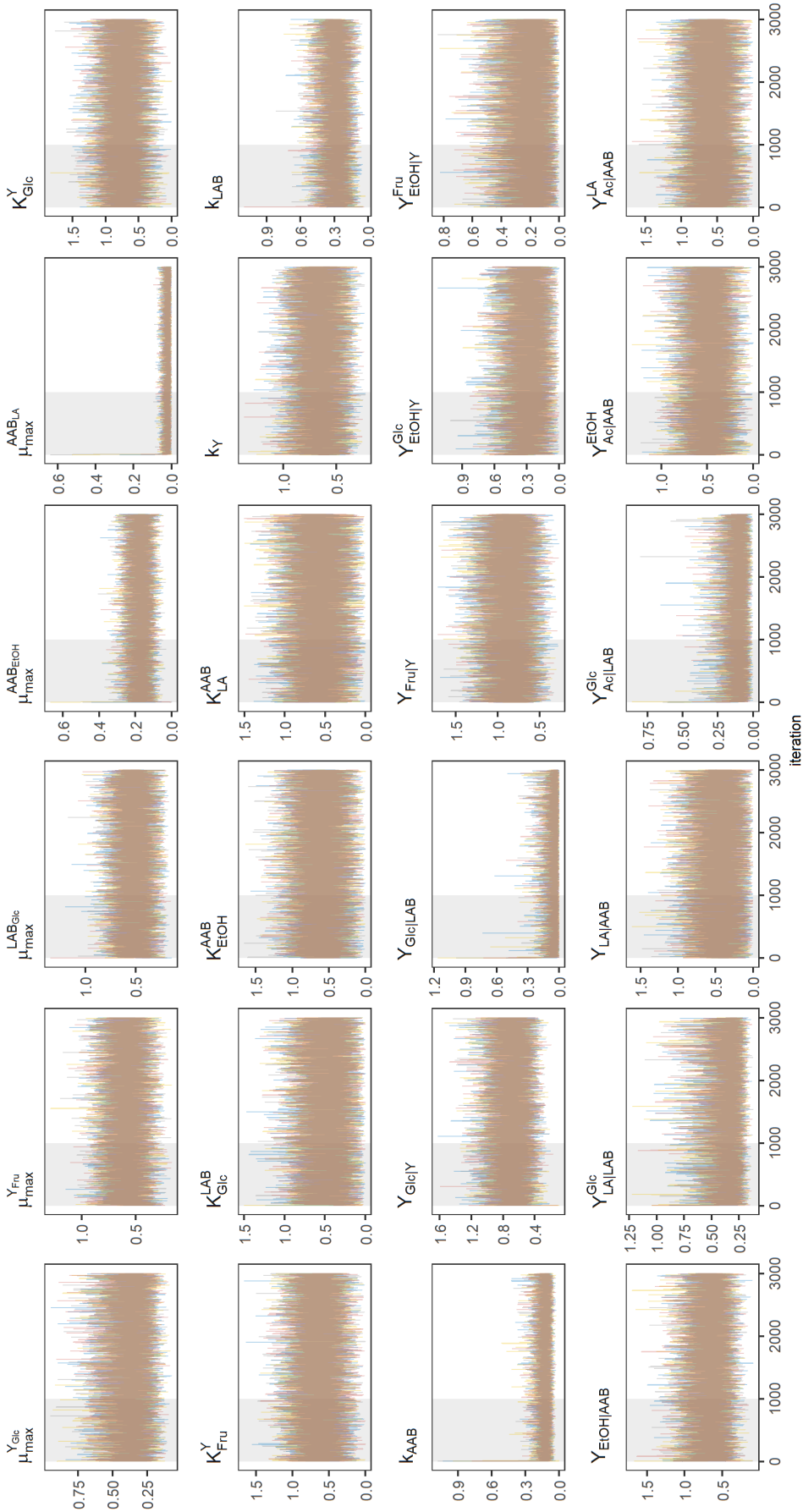


Figure A.7: Scaled trace plots for the 24 parameters of the model estimated from the dataset of Box 2 of Papalexandratou *et al.* [62]. Different color lines, represent different MCMC. The shadowed area represents the 1000 iterations used as warm-up.

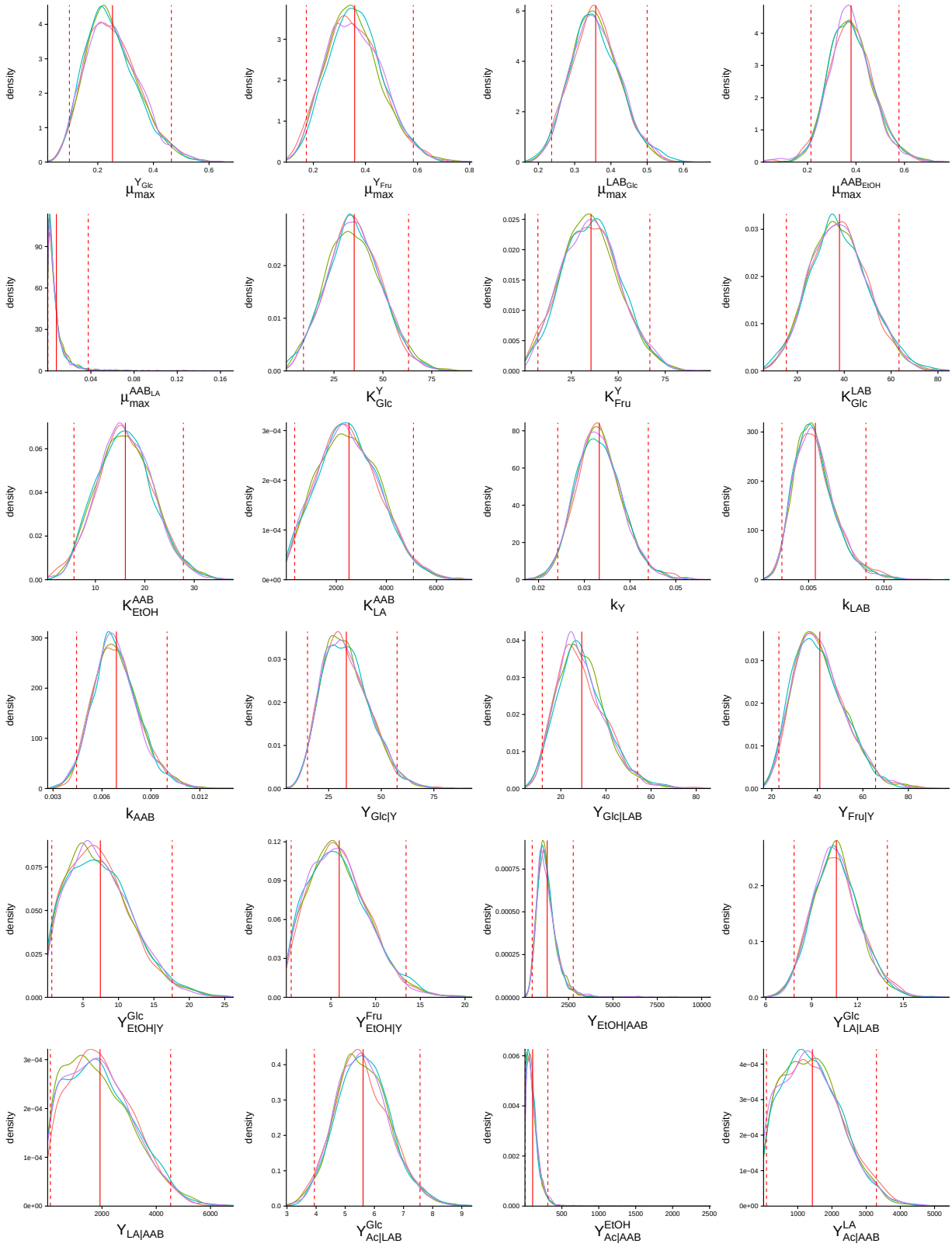


Figure A.8: Unscaled posterior density distributions for the estimated parameters of the proposed models for the data reported by Camu *et al.* [12]. A different color represents a single Markov Chain. The vertical solid vertical lines represent the means of the distributions. The vertical dashed lines denote the 95% credible intervals.

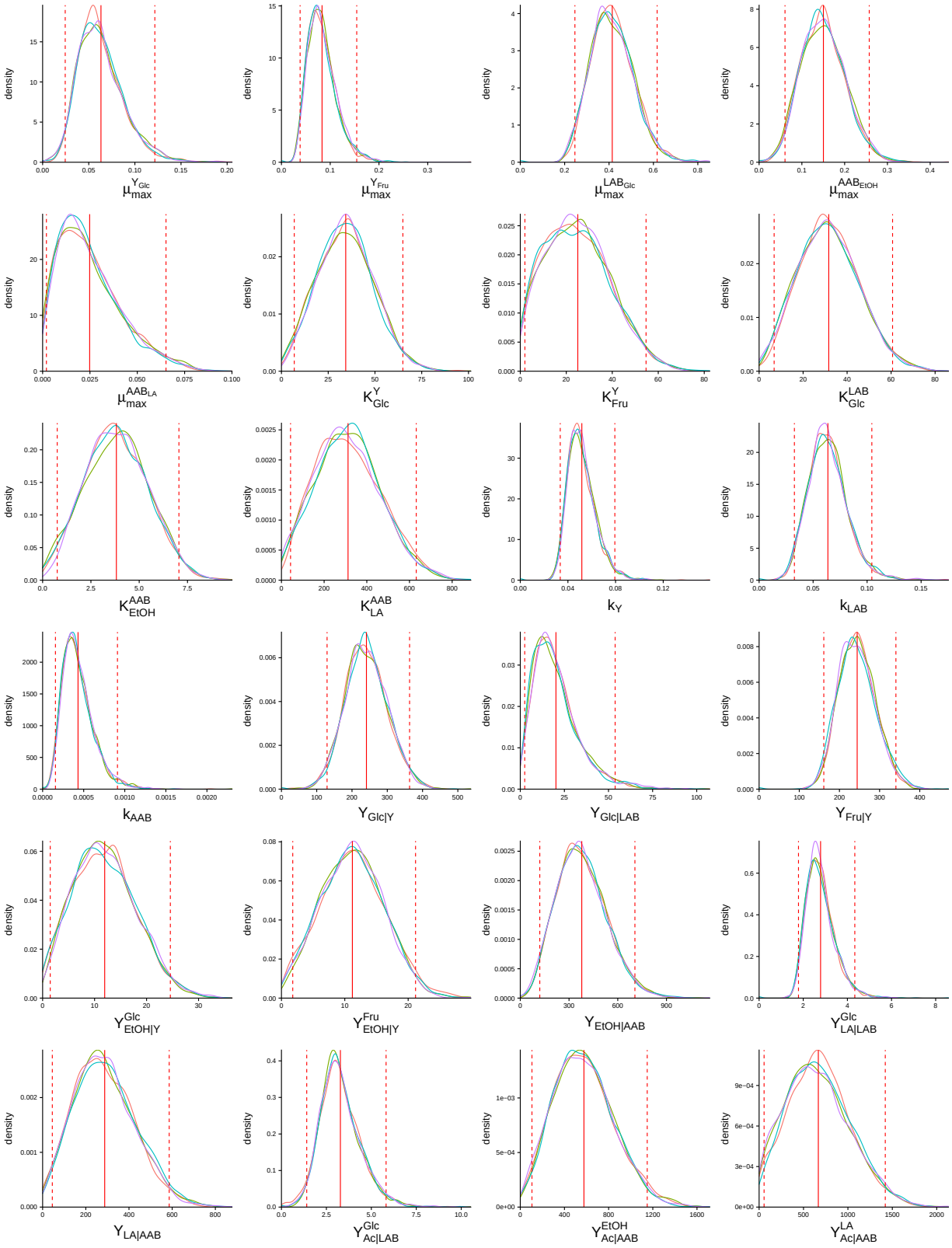


Figure A.9: Unscaled posterior density distributions for the estimated parameters of the proposed models for the data reported for box 1 by Papalexandratou *et al.* [62]. A different color represents a single Markov Chain. The vertical solid vertical lines represent the means of the distributions. The vertical dashed lines denote the 95% credible intervals.

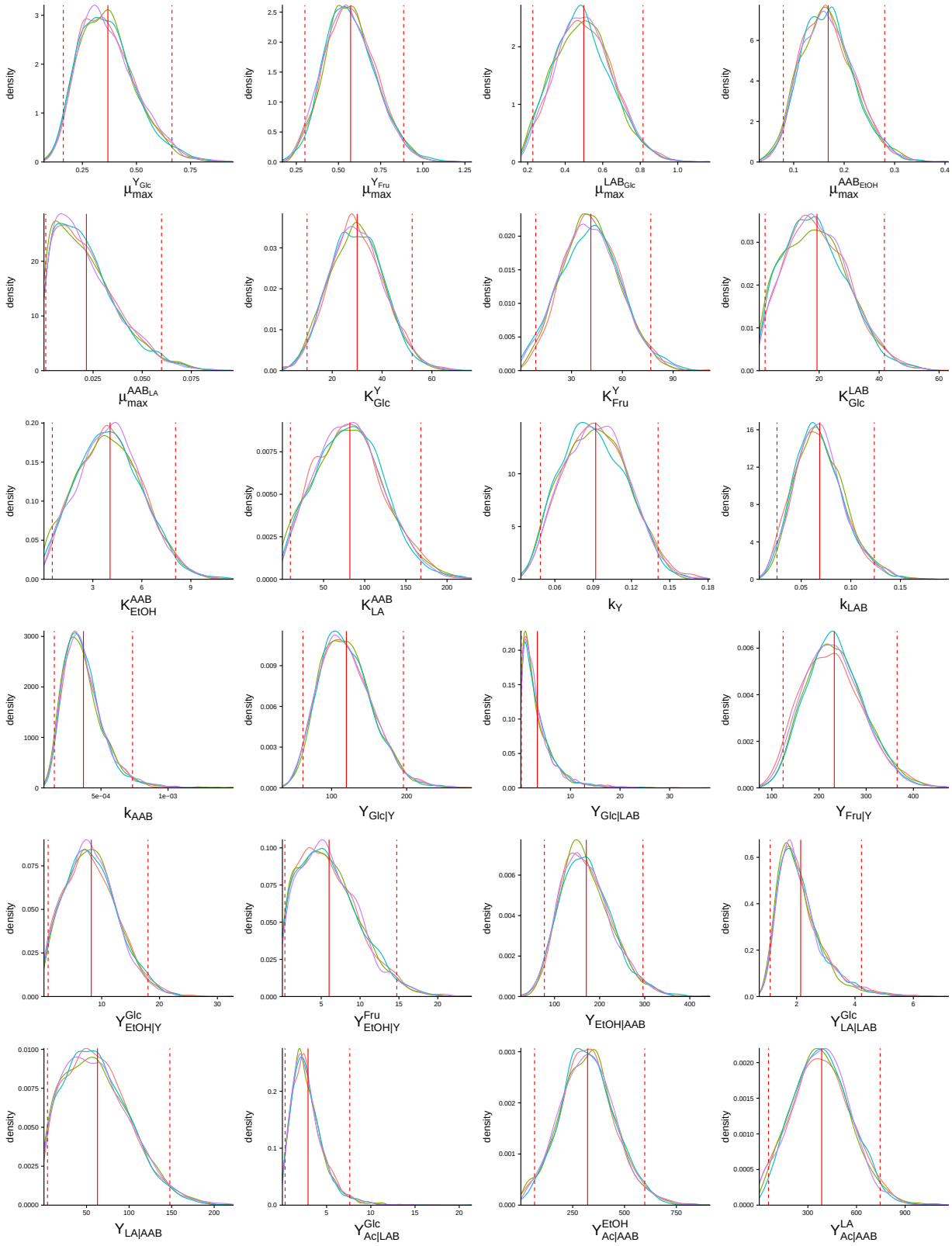


Figure A.10: Unscaled posterior density distributions for the estimated parameters of the proposed models for the data reported for box 2 by Papalexandratou *et al.* [62]. A different color represents a single Markov Chain. The vertical solid vertical lines represent the means of the distributions. The vertical dashed lines denote the 95% credible intervals.

A.5 Asymptotic behavior

The biologically relevant aspect of the dynamics of this fermentation model is the transient, with its sequential activation of different populations. It is, however, also instructive to analyze the asymptotic behavior of the model.

Figure A.11 shows a longer time series for each of the parameter vectors derived from the three data sets.

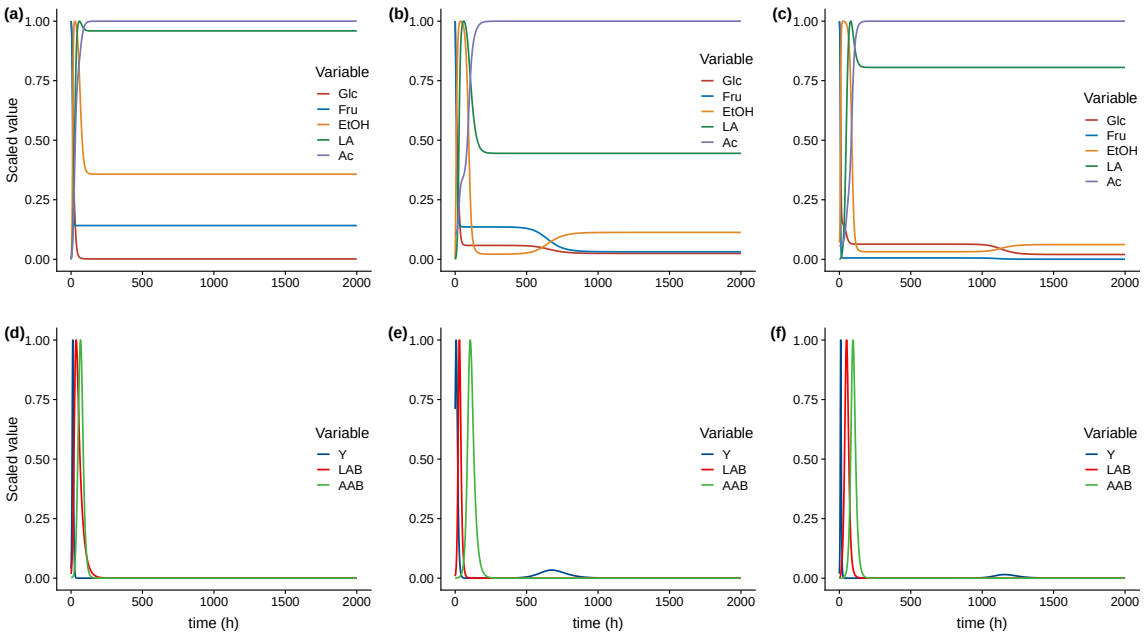


Figure A.11: Long time behaviour of the model. Values scaled to one are used for visualization purposes. Metabolite kinetics for the data reported by (a) Camu *et al.* [12], (b) Box 1, Papalexandratou *et al.* [62] and (c) Box 2, Papalexandratou *et al.* [62]. Microbial community dynamics for the data reported by (d) Camu *et al.* [12], (e) Box 1, Papalexandratou *et al.* [62] and (f) Box 2, Papalexandratou *et al.* [62].

Analytically, one can see that the fixed point of the model has the following form: For some dynamical variables the fixed point value is zero, while others are undetermined or fulfill algebraic relationships, e.g.:

$$
\begin{aligned}
&[\mathbf{Glc}] = [\mathrm{Glc}],\ [\mathbf{Fru}] = 0,\ [\mathbf{EtOH}] = \frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}} + [\mathrm{Glc}])},\ [\mathbf{LA}] = \frac{Y_{\mathrm{Glc|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}(Y_{\mathrm{EtOH|AAB}} + [\mathrm{Glc}])},\\
&[\mathbf{Ac}] = \frac{\sqrt{Y_{\mathrm{Fru|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}} + Y_{\mathrm{Fru|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}[\mathrm{Glc}]Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}} + k_{\mathrm{LAB}}Y_{\mathrm{Glc|LAB}}[\mathrm{Glc}] + k_{\mathrm{LAB}}Y_{\mathrm{Fru|Y}}[\mathrm{Glc}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}} + k_{\mathrm{LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}} + Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}}},\\
&[\mathbf{Y}] = 0,\ [\mathbf{LAB}] = 0,\ [\mathbf{AAB}] = 0
\end{aligned}
\tag{A.20}
$$

as $\frac{dx_i}{dt} = 0$ is an under-determined system.

As several combinations of variables being zero can lead to a fixed point, in total 28 such fixed points exist, all following the pattern described above. In Table A.7 they are listed.

Table A.7: Analytical solution for the fixed points of the model.

[Glc]	[Fru]	[EtOH]	[LA]	[Ac]	[Y]	[LAB]	[AAB]
[Glc]	[Fru]	[EtOH]	[LA]	[Ac]	0	0	0
0	[Fru]	[EtOH]	[LA]	[Ac]	0	0	0
[Glc]	[Fru]	[EtOH]	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	[Ac]	0	0	0
[Glc]	[Fru]	[EtOH]	[LA]	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	[Fru]	[EtOH]	[LA]	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
0	0	0	0	0	[Y]	[LAB]	[AAB]
0	[Fru]	0	0	0	0	[LAB]	[AAB]
0	0	0	[LA]	[Ac]	[Y]	0	0
[Glc]	[Fru]	0	0	0	0	0	[AAB]
0	[Fru]	[EtOH]	[LA]	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
0	[Fru]	[EtOH]	[LA]	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	[LA]	[Ac]	0	0	0
[Glc]	[Fru]	[EtOH]	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	[Fru]	[EtOH]	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}+Y_{\mathrm{Glc\|LAB}}[\mathrm{EtOH}]+Y_{\mathrm{Fru\|Y}}[\mathrm{EtOH}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+[\mathrm{EtOH}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
0	0	0	0	0	[Y]	$-\frac{\mu^{Y_{\mathrm{Glc}}}_{\mathrm{max}}k_{\mathrm{LAB}}Y_{\mathrm{EtOH\|AAB}}[\mathrm{Y}]}{Y_{\mathrm{Glc\|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}c_2}$	[AAB]
0	[Fru]	0	0	0	0	0	[AAB]
0	0	0	0	0	[Y]	0	[AAB]
0	[Fru]	0	0	0	0	0	[AAB]
0	0	0	[LA]	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}}}{\sqrt{Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	[Y]	0	0
0	0	0	[LA]	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}}}{\sqrt{Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	[Y]	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	[LA]	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}[\mathrm{Glc}]Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+k_{\mathrm{LAB}}Y_{\mathrm{Glc\|LAB}}[\mathrm{Glc}]+k_{\mathrm{LAB}}Y_{\mathrm{Fru\|Y}}[\mathrm{Glc}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+k_{\mathrm{LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	[LA]	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}[\mathrm{Glc}]Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+k_{\mathrm{LAB}}Y_{\mathrm{Glc\|LAB}}[\mathrm{Glc}]+k_{\mathrm{LAB}}Y_{\mathrm{Fru\|Y}}[\mathrm{Glc}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+k_{\mathrm{LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	[Ac]	0	0	0
0	0	0	0	0	0	0	[AAB]
0	0	0	[LA]	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}}}{\sqrt{Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
0	0	0	[LA]	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}}}{\sqrt{Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	$-\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}[\mathrm{Glc}]Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+k_{\mathrm{LAB}}Y_{\mathrm{Glc\|LAB}}[\mathrm{Glc}]+k_{\mathrm{LAB}}Y_{\mathrm{Fru\|Y}}[\mathrm{Glc}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+k_{\mathrm{LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0
[Glc]	0	$\frac{k_{\mathrm{LAB}}[\mathrm{Glc}]}{Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}(Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}+[\mathrm{Glc}])}$	$\frac{Y_{\mathrm{Glc\|Y}}[\mathrm{Glc}]}{Y^{\mathrm{EtOH}}_{\mathrm{Ac\|AAB}}(Y_{\mathrm{EtOH\|AAB}}+[\mathrm{Glc}])}$	$\frac{\sqrt{Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+Y_{\mathrm{Fru\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}[\mathrm{Glc}]Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}+k_{\mathrm{LAB}}Y_{\mathrm{Glc\|LAB}}[\mathrm{Glc}]+k_{\mathrm{LAB}}Y_{\mathrm{Fru\|Y}}[\mathrm{Glc}]}}{\sqrt{Y^{\mathrm{Glc}}_{\mathrm{EtOH\|Y}}Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+k_{\mathrm{LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}+Y^{\mathrm{Glc}}_{\mathrm{LA\|LAB}}Y^{\mathrm{Glc}}_{\mathrm{Ac\|LAB}}[\mathrm{Glc}]Y^{\mathrm{LA}}_{\mathrm{Ac\|AAB}}}}$	0	0	0

For those cases, where positive values of all dynamical variables are compatible with the 'template' shown in Table A.7, computation of the eigenvalues of the Jacobian matrix in these fixed points yields zeroes and negative values, consistent with an under-determined system with stable fixed points.

A.6 Measurement errors

As mentioned in Chapter 1, a standard deviation σ was necessary for fitting the model on each data set. Thus, the standard deviation in all three datasets showed lower values at decreased values of the negative log-likelihood conditioned on the posterior distribution (see Figure A.12). At first glance, this result might contradict the general tendency to favor the negative log-likelihood by adding variance to the posterior probabilities of the parameter estimates. However, that kind of behavior would have been present if instead of estimating a total standard deviation for the whole model, separate standard deviations for each state variable were computed.

To show that effect, consider the following total posterior probability:

$$P\left(\theta \mid \mathcal{Y}\right) \propto \prod_{i=1}^{T}\prod_{j=1}^{N}\mathcal{N}\left(f\left(x_{i,j},\theta\right),\sigma_i\right)\, P\left(\theta\right). \tag{A.21}$$

Eq. (A.21) differs from Eq. (1.24) by σ_i, which is representing a standard deviation for each

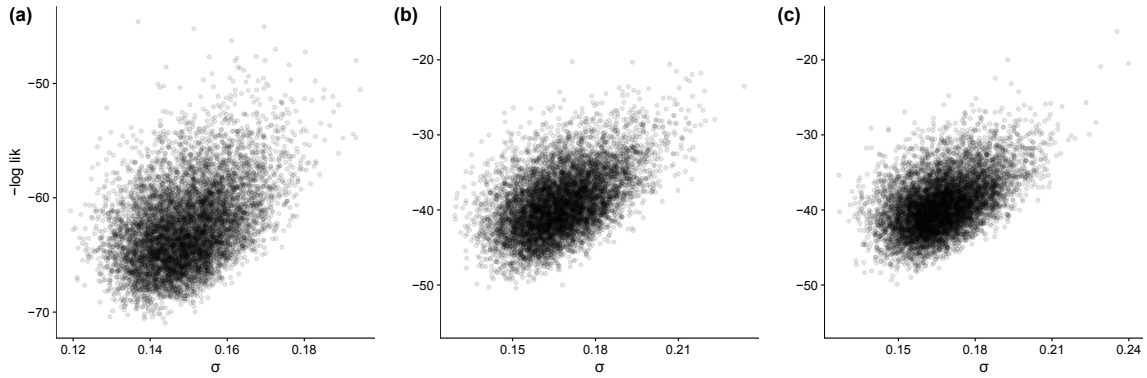


Figure A.12: Scatter plots of the negative log-likelihood conditioned on the posterior distribution vs. scaled estimated standard deviation σ for each of the data sets. (a) Data reported by Camu *et al.* [12], (b) Data reported for Box 1 Papalexandratou *et al.* [62] and (c) Data reported for Box 2 Papalexandratou *et al.* [62].

i state variable of the model. Applying Eq. (A.21) into our proposed model, would mean to estimate 8 standard deviations that in practice are likely to inflate the log-likelihood. As an example, by fitting a model according to Eq. (A.21) with the data reported by Camu *et al.* [12] the log-likelihood is increased from a mean of 62.48 obtained by FM reported in Chapter 2, to a mean of 72.91 (Figure A.13).

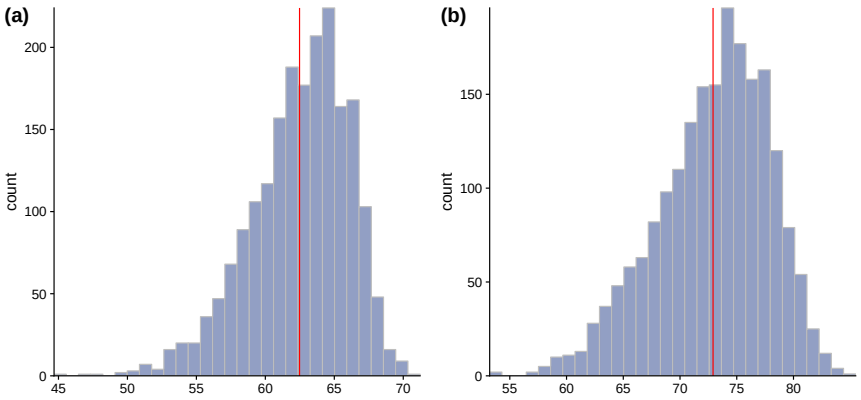


Figure A.13: Log-likelihood conditioned on the posterior distribution with data from Camu *et al.* [12]. Solid vertical red line represents the mean of the distribution. (a) Model with one σ (b) Model with one σ per state variable

This increment of the log-likelihood is likely related to an inflation of certain parameters, namely $\mu_{\max}^{\mathrm{AAB_{LA}}}$, $Y_{\mathrm{EtOH|AAB}}$ and $Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}}$ (Figure A.14).

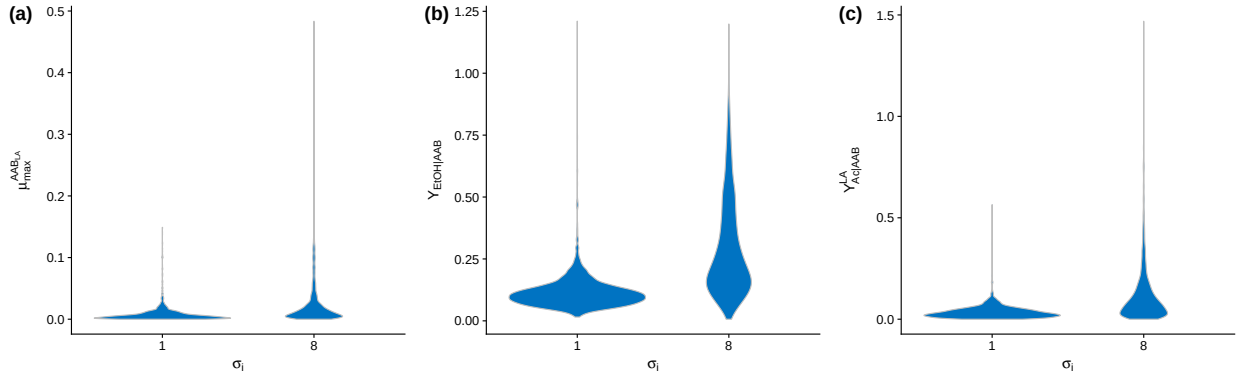


Figure A.14: Posterior distributions for scaled parameters (a) $\mu_{\max}^{\mathrm{AAB_{LA}}}$, (b) $Y_{\mathrm{EtOH|AAB}}$ and (c) $Y_{\mathrm{Ac|AAB}}^{\mathrm{LA}}$ with data from Camu *et al.* [12]. The x axis is representing whether the model was fitted with either one or eight σ

.

A.7 Parameters' correlation

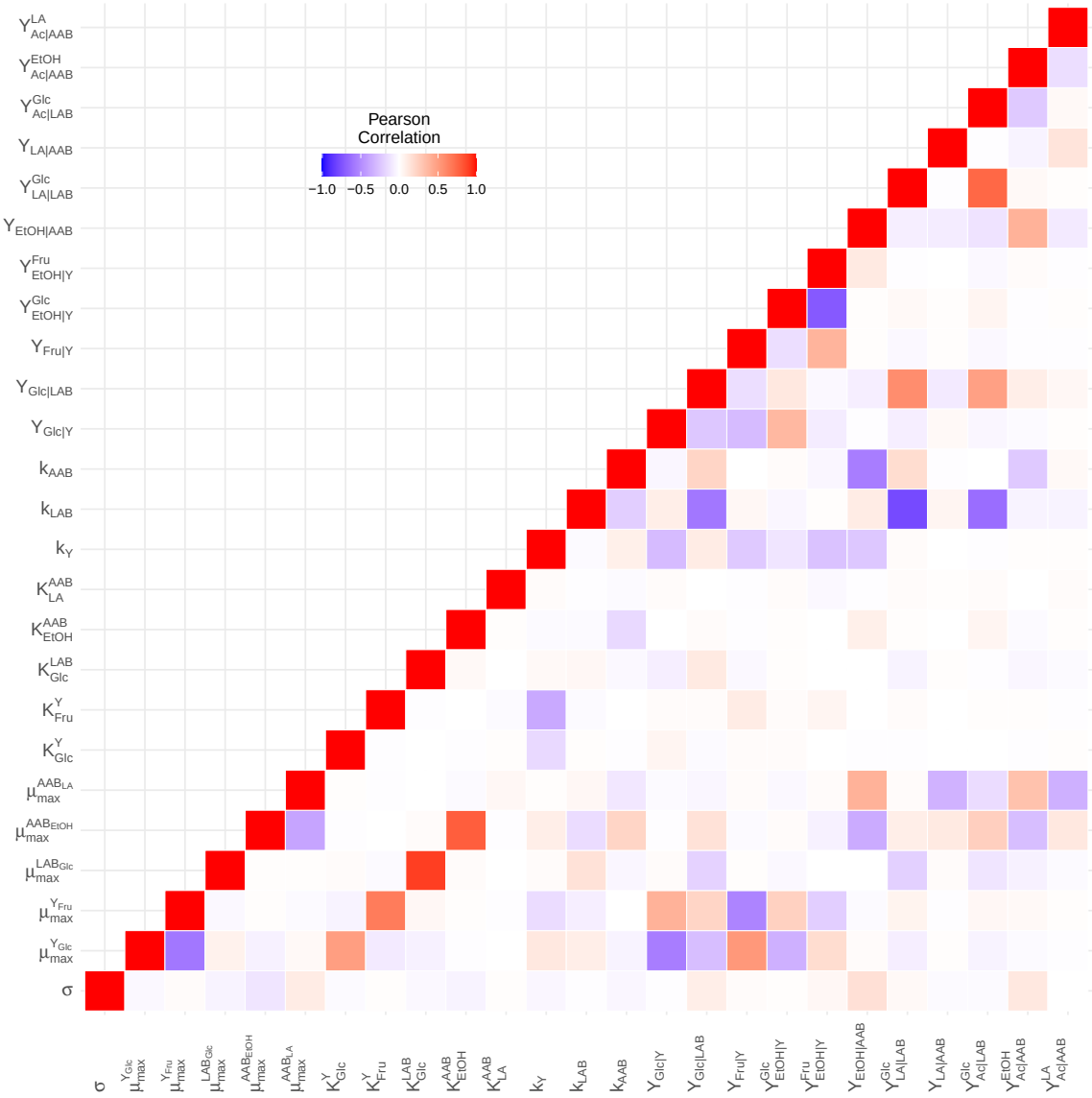


Figure A.15: Heatmap based on Pearson's correlation derived from the data reported by Camu *et al.* [12].

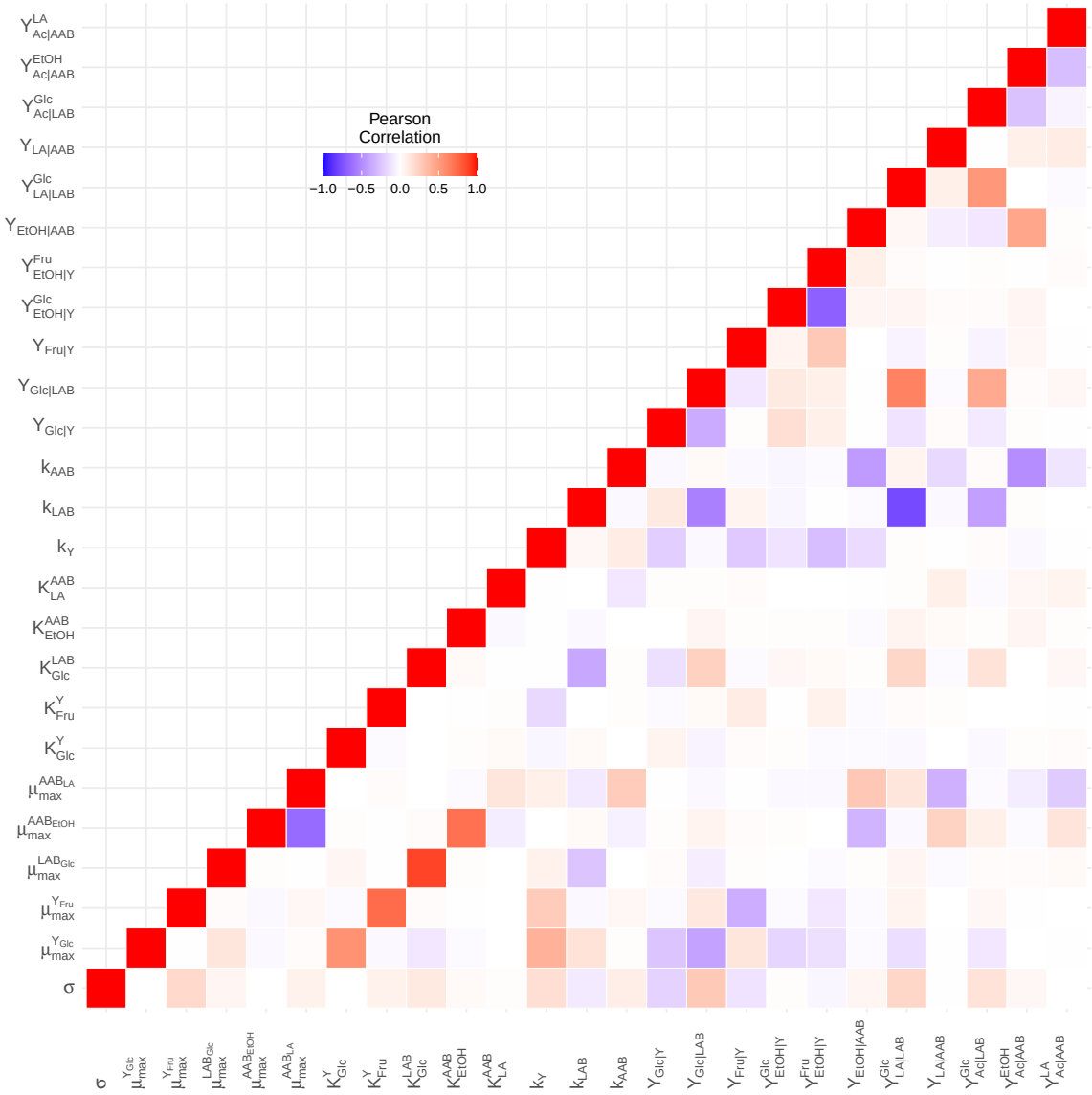


Figure A.16: Heatmap based on Pearson's correlation derived from the data reported by Papalexandratou *et al.* [62] for Box 1.

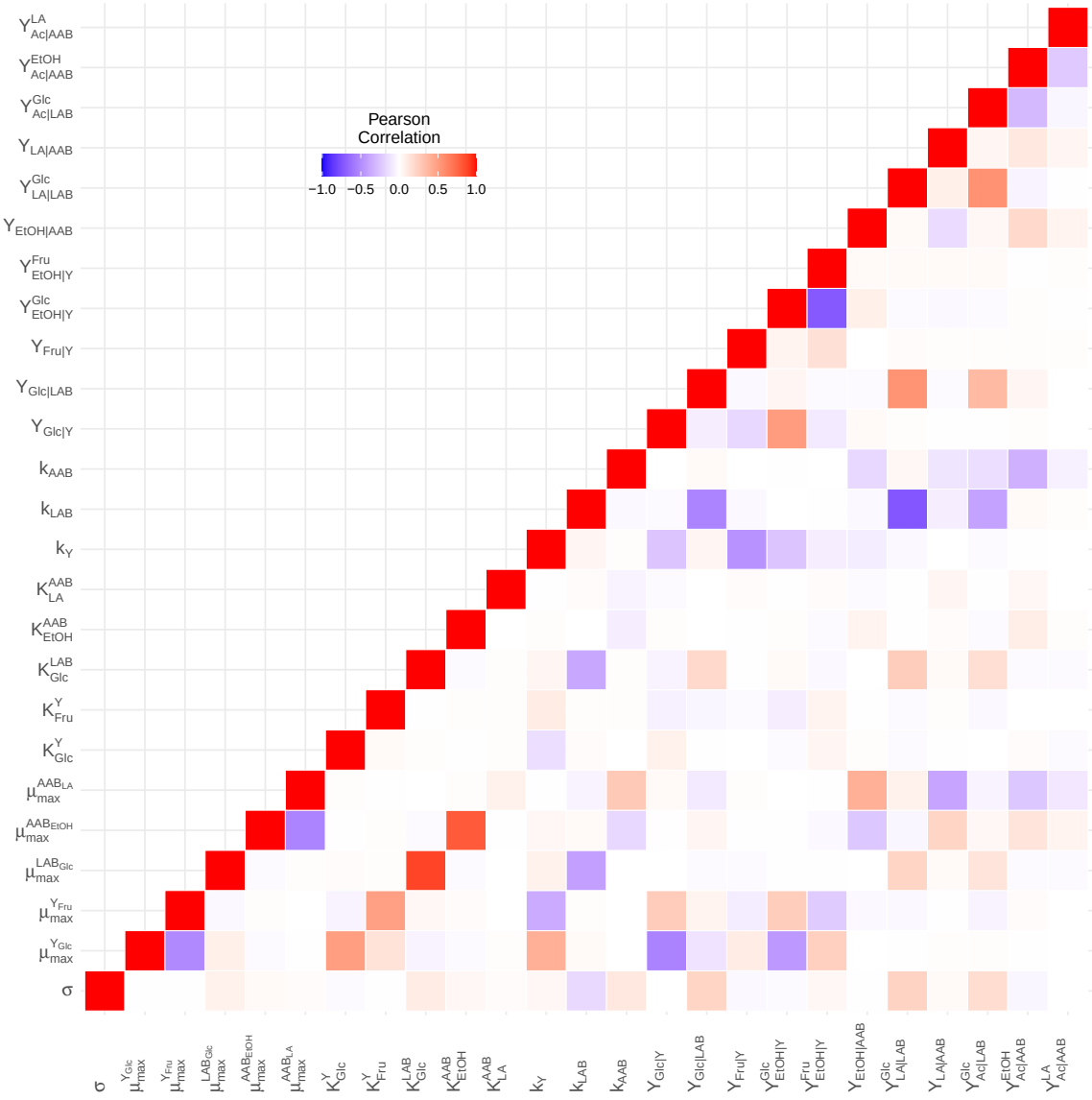


Figure A.17: Heatmap based on Pearson's correlation derived from the data reported by Papalexandratou *et al.* [62] for Box 2.

Appendix B

Exploring cocoa bean fermentation mechanisms by kinetic modeling – Supplementary material[8]

B.1 Fermentation data

From bibliographical review, 15 published research papers met the inclusion criteria described in Section 3.2.1 [12–15, 62, 68, 112–117, 126–128]. We also considered an additional study lacking time series for LA due to the limited number of works where heap method was used [17]. For this paper, LA's time series were simulated during models calibrations. Overall, data from 32 fermentation trials were gathered. After the calibration of models, 9 out of 32 datasets did not result in successful fits for any MI (see Table B.1).

Fermentation trials that resulted in successful fits, according to features mentioned in Section 3.2.6, were grouped by country in 8 experiments from Brazil [14, 15, 62, 68], 5 from Ghana [12, 17], 5 from Ecuador [113, 116], 2 from Nicaragua [117], 1 from Dominican Republic [112], 1 from Malaysia [114] and 1 from Honduras [115]. Regarding the cocoa cultivar, 4 experiments were performed using Nacional/Trinitario [113], 3 using Criollo/Forastero [12, 62], 1 using Trinitario [112], 1 using PH16 [15], 1 using a mix of IMC-67, UF-29 and UF-668 [115], 1 using Criollo [116], 1 using Nugu [117], 1 using O'payo [117] and 1 using unspecified mixed hybrids [114]. Furthermore, 2 studies with 6 trials in total did not mention any identification about used cultivar [17, 68] and 1 remaining study reported 3 trials with the use of unspecified mixed hybrids [14]. With the aim of not losing information from such studies, cultivars were here categorized as different unknowns based on their authorship. Hence, 4 trials with unidentified cultivars performed by Camu *et al.* [17] and 2 conducted by Pereira *et al.* [68] were coded as *un1* and *un2*, respectively. Finally, 3 trials using unspecified mixed hybrids performed by Pereira *et al.* [14] were coded as *un3*.

Regarding the fermentation method, trials were grouped in 11 from wooden boxes [14, 15, 62, 112–115, 117], 5 from heaps [12, 17], 2 from platforms [117], 2 from stainless steel tanks [14, 68] and 2 from plastic boxes [68, 116]. About turning of fermenting mass, 16 experiments reported to have turned cocoa beans during fermentation [14, 17, 62, 68, 112–115, 117], 5 did not conduct turning [12, 17, 113] and 2 did not report whether turning of fermenting cocoa beans was done [15, 116]. Regarding temperature control, only 3 trials reported to have controlled it [68, 116]. Finally, with respect to the use of a starter culture, data gathered from the only study that suited inclusion criteria and reported its use [126], could not be fitted to any MI. A

[8]This chapter is based on the supplementary material of Moreno-Zambrano *et al.* [2]

summary of fermentation experiments that could be fitted to at least one MI is shown in Table 3.1.

B.2 Mathematical formulation of candidate mechanisms

M1: Decay of fermentation products

M1 can be formulated by considering a linear term of the form

$$d_i = b_X[X], \tag{B.1}$$

where d represents the decay i at a rate b of metabolite X, can be added to Eqs. (2.3), (2.4) and (2.5) to account for decrease of EtOH, LA and Ac, respectively:

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}\,v_2 - Y_{\mathrm{EtOH|AAB}}\,v_4 - d_1, \tag{B.2}$$

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\,v_3 - Y_{\mathrm{LA|AAB}}\,v_5 - d_2 \tag{B.3}$$

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\,v_3 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\,v_4 + Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}\,v_5 - d_3. \tag{B.4}$$

Therefore, the addition of these terms provides three extra parameters to the baseline model.

M2: Consumption of fructose (Fru) by lactic acid bacteria (LAB)

M2 can be easily implemented into the baseline model in three steps. First, an extra growth rate for LAB is needed to describe uptake of Fru as carbon source. Such a growth rate can be defined by a Monod type equation of the form

$$v_9 = \frac{\mu^{\mathrm{LAB_{Fru}}}_{\mathrm{max}}\;[\mathrm{Fru}]}{[\mathrm{Fru}] + K^{\mathrm{LAB}}_{\mathrm{Fru}}}\;[\mathrm{LAB}]\,, \tag{B.5}$$

where, $\mu^{\mathrm{LAB_{Fru}}}_{\mathrm{max}}$ represents the maximum specific growth rate of LAB on Fru and $K^{\mathrm{LAB}}_{\mathrm{Fru}}$ is the substrate saturation constant of LAB growth on Fru. Then, v_9 must be included in Eq. (2.2) with aid of a LAB-to-Fru yield coefficient ($Y_{\mathrm{Fru|LAB}}$), giving

$$\frac{d\,[\mathrm{Fru}]}{dt} = -Y_{\mathrm{Fru|Y}}\,v_2 - Y_{\mathrm{Fru|LAB}}\,v_9\,. \tag{B.6}$$

Second, contributions of LAB in producing EtOH, LA and Ac from growth rates v_3 and v_9 are included in Eqs. (2.3), (2.4) and (2.5) with help of four extra parameters, namely LAB-to-EtOH from Glc ($Y^{\mathrm{Glc}}_{\mathrm{EtOH|LAB}}$), LAB-to-EtOH from Fru ($Y^{\mathrm{Fru}}_{\mathrm{EtOH|LAB}}$), LAB-to-LA from Fru ($Y^{\mathrm{Fru}}_{\mathrm{LA|LAB}}$), and LAB-to-Ac from Fru ($Y^{\mathrm{Fru}}_{\mathrm{Ac|LAB}}$) yield coefficients:

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}\,v_2 + Y^{\mathrm{Glc}}_{\mathrm{EtOH|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|LAB}}\,v_9 - Y_{\mathrm{EtOH|AAB}}\,v_4 \tag{B.7}$$

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{LA|LAB}}\,v_9 - Y_{\mathrm{LA|AAB}}\,v_5 \tag{B.8}$$

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\,v_3 + Y^{\mathrm{Fru}}_{\mathrm{Ac|LAB}}\,v_9 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\,v_4. \tag{B.9}$$

Finally, increase on microbial population by growth rate v_9 can be added to Eq. (2.7) as

$$\frac{d\,[\mathrm{LAB}]}{dt} = v_3 + v_9 - v_7. \tag{B.10}$$

In total, implementation of Eqs. (B.6) to (B.10) in the baseline model adds seven extra parameters.

M3: Production of acetic acid (Ac) by yeasts (Y)

Given that the baseline model already accounts for growth rates of Y on Glc (v_1) and Fru (v_2), deployment of this new mechanism in Eq. (2.5) needs Y-to-Ac from Glc ($Y^{\mathrm{Glc}}_{\mathrm{Ac|Y}}$) and Y-to-Ac from Fru ($Y^{\mathrm{Fru}}_{\mathrm{Ac|Y}}$) yield coefficients:

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\,v_3 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\,v_4 + Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}\,v_5 + Y^{\mathrm{Glc}}_{\mathrm{Ac|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{Ac|Y}}\,v_2, \tag{B.11}$$

adding two more parameters to the baseline model.

M4: Production of lactic acid (LA) by yeasts (Y)

Similar to M2, M4 can be incorporated to the baseline model in three steps. First, a new growth rate for Y on LA is defined by a Monod type equation of the form

$$v_{10} = \frac{\mu^{\mathrm{Y_{LA}}}_{\mathrm{max}}\,[\mathrm{LA}]}{[\mathrm{LA}] + K^{\mathrm{Y}}_{\mathrm{LA}}}\,[\mathrm{Y}]\ , \tag{B.12}$$

where, $\mu^{\mathrm{Y_{LA}}}_{\mathrm{max}}$ is the maximum specific growth rate of Y on LA and $K^{\mathrm{Y}}_{\mathrm{LA}}$ represents the substrate saturation constant of Y growth on LA. v_{10} is then added to Eq. (2.4) with a Y-to-LA yield coefficient ($Y_{\mathrm{LA|Y}}$):

$$\frac{d\,[\mathrm{LA}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{LA|LAB}}\,v_3 - Y_{\mathrm{LA|AAB}}\,v_5 - Y_{\mathrm{LA|Y}}\,v_{10}. \tag{B.13}$$

Second, to account the production of EtOH via consumption of LA by Y, v_{10} must be part of Eq. (2.3) with companion of a Y-to-EtOH from LA yield coefficient ($Y^{\mathrm{LA}}_{\mathrm{EtOH|Y}}$):

$$\frac{d\,[\mathrm{EtOH}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{EtOH|Y}}\,v_1 + Y^{\mathrm{Fru}}_{\mathrm{EtOH|Y}}\,v_2 + Y^{\mathrm{LA}}_{\mathrm{EtOH|Y}}\,v_{10} - Y_{\mathrm{EtOH|AAB}}\,v_4. \tag{B.14}$$

Finally, the population size Y (Eq. (2.6)) is also affected by v_{10}:

$$\frac{d\,[\mathrm{Y}]}{dt} = v_1 + v_2 + v_{10} - v_6. \tag{B.15}$$

In total, M4 contributes 4 parameters to the baseline model.

M5: Over-oxidation of acetic acid (Ac) by acetic acid bacteria (AAB)

Similar to M2 and M4, an extra growth rate is required to describe AAB growth on its own main product, Ac:

$$v_{11} = \frac{\mu^{\mathrm{AAB_{Ac}}}_{\mathrm{max}}\,[\mathrm{Ac}]}{[\mathrm{Ac}] + K^{\mathrm{AAB}}_{\mathrm{Ac}}}\,[\mathrm{AAB}]\ , \tag{B.16}$$

where $\mu^{\mathrm{AAB_{Ac}}}_{\mathrm{max}}$ depicts the maximum specific growth rate of AAB on Ac and $K^{\mathrm{AAB}}_{\mathrm{Ac}}$ is the substrate saturation constant of AAB growth on Ac. Then, v_{11} enters in Eq. (2.5) with an AAB-to-Ac yield coefficient ($Y_{\mathrm{Ac|AAB}}$):

$$\frac{d\,[\mathrm{Ac}]}{dt} = Y^{\mathrm{Glc}}_{\mathrm{Ac|LAB}}\, v_3 + Y^{\mathrm{EtOH}}_{\mathrm{Ac|AAB}}\, v_4 + Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}\, v_5 - Y_{\mathrm{Ac|AAB}}\, v_{11}. \tag{B.17}$$

Lastly, population of AAB described in Eq. (2.8) is complemented with v_{11}:

$$\frac{d\,[\mathrm{AAB}]}{dt} = v_4 + v_5 + v_{11} - v_8\,. \tag{B.18}$$

Since we are not interested in simulating carbon dioxide production, no extra equation is needed and M5 ends up adding three extra parameters to the baseline model.

B.3 Supplementary tables and figures

Table B.1: Detailed description of the considered data. Author, year of publication, cocoa country of origin, cocoa cultivar, used methodology, code name given in the original paper, re-coded name given in this research, use of starter culture, turning of the fermenting mass and controlled temperature are shown.

Reference	Year	Country	Cultivar	Method	Trial	Code	Starter	Turning	Ctrl. Temp.
Camu *et al.* [12]	2007	Ghana	Criollo/Forastero	heap	heap 5	ghhp1	✗	✗	✗
Lagunes Gálvez *et al.* [112]	2007	Dominican Republic	Trinitario	wooden box	NA	dowb1	✗	✓	✗
Camu *et al.* [17]	2008	Ghana	NA	heap	heap 10	ghhp2	✗	✓	✗
					heap 11	ghhp3	✗	✗	✗
					heap 12	ghhp4	✗	✓	✗
					heap 13	ghhp5	✗	✗	✗
Lefeber *et al.* [13]	2011	Ivory Coast	NA	plastic box	NA	cipb1	✗	✗	✗
Papalexandratou *et al.* [62]	2011	Brazil	Criollo/Forastero	wooden box	box 1	brwb1	✗	✓	✗
					box 2	brwb2	✗	✓	✗
Papalexandratou *et al.* [113]	2011	Ecuador	Nacional/Trinitario	platform	P1	ecpt1	✗	✗	✗
					P2	ecpt2	✗	✗	✗
				wooden box	B1	ecwb1	✗	✓	✗
					B2	ecwb2	✗	✓	✗
Pereira *et al.* [68]	2012	Brazil	NA	plastic box	PC	brpb1	✗	✓	✓
				stainless tank	ST	brst1	✗	✓	✓
Lefeber *et al.* [126]	2012	Malaysia	NA	wooden box	NA	mywb1	✓	✓	NA
					NA	mywb2	✓	✓	NA
		Ghana	NA	heap	NA	ghhp6	✗	NA	NA
					NA	ghhp7	✓	✓	NA
Pereira *et al.* [14]	2013	Brazil	Mixed hybrids	wooden box	WB1	brwb3	✗	✓	✗
					WB2	brwb4	✗	✓	✗
				stainless tank	SST	brst2	✗	✓	✗
Moreira *et al.* [15]	2013	Brazil	PH9, PH15, PH16	wooden box	PH9	brwb5	✗	NA	✗
					PH15	brwb6	✗	NA	✗
					PH16	brwb7	✗	NA	✗
Papalexandratou *et al.* [114]	2013	Malaysia	Mixed hybrids	wooden box	box 2	mywb3	✗	✓	✗
Bastos *et al.* [127]	2018	Brazil	TSH565	wooden box	NA	brwb8	✗	✓	✗
Romanens *et al.* [115]	2018	Honduras	IMC-67, UF-29, UF-668	wooden box	OF-F	hnwb1	✗	✓	✗
†Lee *et al.* [116]	2019	Ecuador	Criollo	plastic box	NA	ecpb1	✗	NA	✓
Papalexandratou *et al.* [117]	2019	Nicaragua	Nugu/O'payo	wooden box	NUGU	niwb1	✗	✓	✗
					O'PAYO	niwb2	✗	✓	✗
†Racine *et al.* [128]	2019	Ecuador	Criollo	plastic box	NA	ecpb2	✗	✓	✓

† Simulated fermentation.

Table B.2: Prior distributions in the range of original units for parameters of all model iterations per dataset. All priors are truncated to the positive set of real numbers.

Parameter	ghhp1	dowb1	ghhp2	ghhp3	ghhp4	ghhp5	cipb1	brwb1
$\mu_{\max}^{Y_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{EtOH}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{Ac}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
K_{Glc}^{Y}	$\mathcal{N}(25.98, 15.59)$	$\mathcal{N}(28.45, 17.07)$	$\mathcal{N}(23.98, 14.39)$	$\mathcal{N}(20.12, 12.07)$	$\mathcal{N}(20.79, 12.47)$	$\mathcal{N}(24.65, 14.79)$	$\mathcal{N}(47.8, 28.68)$	$\mathcal{N}(27.74, 16.64)$
K_{Fru}^{Y}	$\mathcal{N}(28.87, 17.32)$	$\mathcal{N}(44.4, 26.64)$	$\mathcal{N}(25.69, 15.42)$	$\mathcal{N}(25.43, 15.26)$	$\mathcal{N}(22.11, 13.26)$	$\mathcal{N}(25.7, 15.42)$	$\mathcal{N}(91, 54.6)$	$\mathcal{N}(24.83, 14.9)$
K_{LA}^{Y}	$\mathcal{N}(28.87, 17.32)$	$\mathcal{N}(44.4, 26.64)$	$\mathcal{N}(25.69, 15.42)$	$\mathcal{N}(25.43, 15.26)$	$\mathcal{N}(22.11, 13.26)$	$\mathcal{N}(25.7, 15.42)$	$\mathcal{N}(91, 54.6)$	$\mathcal{N}(24.83, 14.9)$
K_{Glc}^{LAB}	$\mathcal{N}(25.98, 15.59)$	$\mathcal{N}(28.45, 17.07)$	$\mathcal{N}(23.98, 14.39)$	$\mathcal{N}(20.12, 12.07)$	$\mathcal{N}(20.79, 12.47)$	$\mathcal{N}(24.65, 14.79)$	$\mathcal{N}(47.8, 28.68)$	$\mathcal{N}(27.74, 16.64)$
K_{Fru}^{LAB}	$\mathcal{N}(28.87, 17.32)$	$\mathcal{N}(44.4, 26.64)$	$\mathcal{N}(25.69, 15.42)$	$\mathcal{N}(25.43, 15.26)$	$\mathcal{N}(22.11, 13.26)$	$\mathcal{N}(25.7, 15.42)$	$\mathcal{N}(91, 54.6)$	$\mathcal{N}(24.83, 14.9)$
K_{EtOH}^{AAB}	$\mathcal{N}(11.25, 6.75)$	$\mathcal{N}(4.85, 2.91)$	$\mathcal{N}(5.25, 3.15)$	$\mathcal{N}(6.71, 4.03)$	$\mathcal{N}(9.14, 5.48)$	$\mathcal{N}(10.13, 6.08)$	$\mathcal{N}(8.3, 4.98)$	$\mathcal{N}(3.02, 1.81)$
K_{LA}^{AAB}	$\mathcal{N}(2243.66, 2.58)$	$\mathcal{N}(24.95, 0.67)$	-	-	-	-	$\mathcal{N}(55.21, 4.89)$	$\mathcal{N}(286.16, 1.46)$
K_{Ac}^{AAB}	$\mathcal{N}(3.07, 1.84)$	$\mathcal{N}(11.24, 6.74)$	$\mathcal{N}(7.9, 4.74)$	$\mathcal{N}(2.15, 1.29)$	$\mathcal{N}(6.04, 3.62)$	$\mathcal{N}(5.88, 3.53)$	$\mathcal{N}(19.25, 11.55)$	$\mathcal{N}(5.97, 3.58)$
k_Y	$\mathcal{N}(0.0222, 0.0133)$	$\mathcal{N}(0.0515, 0.0309)$	$\mathcal{N}(0.0477, 0.0286)$	$\mathcal{N}(0.0373, 0.0224)$	$\mathcal{N}(0.0274, 0.0164)$	$\mathcal{N}(0.0247, 0.0148)$	$\mathcal{N}(0.0301, 0.0181)$	$\mathcal{N}(0.0829, 0.0497)$
k_{LAB}	$\mathcal{N}(0.0582, 0.0349)$	$\mathcal{N}(0.2222, 0.1333)$	-	-	-	-	$\mathcal{N}(0.0307, 0.0184)$	$\mathcal{N}(0.1026, 0.0616)$
k_{AAB}	$\mathcal{N}(0.0133, 0.008)$	$\mathcal{N}(0.001, 6 \times 10^{-4})$	$\mathcal{N}(0.002, 0.0012)$	$\mathcal{N}(0.027, 0.0162)$	$\mathcal{N}(0.0034, 0.0021)$	$\mathcal{N}(0.0036, 0.0022)$	$\mathcal{N}(3 \times 10^{-4}, 2 \times 10^{-4})$	$\mathcal{N}(0.0035, 0.0021)$
$Y_{Glc\|Y}$	$\mathcal{N}(27.2, 16.32)$	$\mathcal{N}(29.72, 17.83)$	$\mathcal{N}(47.18, 28.31)$	$\mathcal{N}(56.05, 33.63)$	$\mathcal{N}(32.41, 19.45)$	$\mathcal{N}(73.93, 44.36)$	$\mathcal{N}(316.47, 189.88)$	$\mathcal{N}(130.63, 78.38)$
$Y_{Glc\|LAB}$	$\mathcal{N}(45.89, 27.54)$	$\mathcal{N}(306.37, 183.82)$	$\mathcal{N}(59.55, 35.73)$	$\mathcal{N}(25.28, 15.17)$	$\mathcal{N}(25, 15)$	$\mathcal{N}(14.55, 8.73)$	$\mathcal{N}(57.08, 34.25)$	$\mathcal{N}(74.17, 44.5)$
$Y_{Fru\|Y}$	$\mathcal{N}(30.22, 18.13)$	$\mathcal{N}(46.37, 27.82)$	$\mathcal{N}(50.55, 30.33)$	$\mathcal{N}(70.82, 42.49)$	$\mathcal{N}(34.47, 20.68)$	$\mathcal{N}(77.05, 46.23)$	$\mathcal{N}(602.49, 361.49)$	$\mathcal{N}(116.94, 70.16)$
$Y_{Fru\|LAB}$	$\mathcal{N}(51, 30.6)$	$\mathcal{N}(478.02, 286.81)$	$\mathcal{N}(63.81, 38.29)$	$\mathcal{N}(31.94, 19.17)$	$\mathcal{N}(26.58, 15.95)$	$\mathcal{N}(15.17, 9.1)$	$\mathcal{N}(108.68, 65.21)$	$\mathcal{N}(66.4, 39.84)$
$Y_{EtOH\|Y}^{Glc}$	$\mathcal{N}(11.77, 7.06)$	$\mathcal{N}(5.07, 3.04)$	$\mathcal{N}(10.32, 6.19)$	$\mathcal{N}(18.69, 11.21)$	$\mathcal{N}(14.24, 8.55)$	$\mathcal{N}(30.37, 18.22)$	$\mathcal{N}(54.95, 32.97)$	$\mathcal{N}(14.2, 8.52)$
$Y_{EtOH\|Y}^{Fru}$	$\mathcal{N}(11.77, 7.06)$	$\mathcal{N}(5.07, 3.04)$	$\mathcal{N}(10.32, 6.19)$	$\mathcal{N}(18.69, 11.21)$	$\mathcal{N}(14.24, 8.55)$	$\mathcal{N}(30.37, 18.22)$	$\mathcal{N}(54.95, 32.97)$	$\mathcal{N}(14.2, 8.52)$
$Y_{EtOH\|Y}^{LA}$	$\mathcal{N}(11.77, 7.06)$	$\mathcal{N}(5.07, 3.04)$	$\mathcal{N}(10.32, 6.19)$	$\mathcal{N}(18.69, 11.21)$	$\mathcal{N}(14.24, 8.55)$	$\mathcal{N}(30.37, 18.22)$	$\mathcal{N}(54.95, 32.97)$	$\mathcal{N}(14.2, 8.52)$
$Y_{EtOH\|LAB}^{Glc}$	$\mathcal{N}(19.86, 11.92)$	$\mathcal{N}(52.23, 31.34)$	$\mathcal{N}(13.03, 7.82)$	$\mathcal{N}(8.43, 5.06)$	$\mathcal{N}(10.99, 6.59)$	$\mathcal{N}(5.98, 3.59)$	$\mathcal{N}(9.91, 5.95)$	$\mathcal{N}(8.06, 4.84)$
$Y_{EtOH\|LAB}^{Fru}$	$\mathcal{N}(19.86, 11.92)$	$\mathcal{N}(52.23, 31.34)$	$\mathcal{N}(13.03, 7.82)$	$\mathcal{N}(8.43, 5.06)$	$\mathcal{N}(10.99, 6.59)$	$\mathcal{N}(5.98, 3.59)$	$\mathcal{N}(9.91, 5.95)$	$\mathcal{N}(8.06, 4.84)$
$Y_{EtOH\|AAB}$	$\mathcal{N}(5872.73, 3523.64)$	$\mathcal{N}(107.57, 64.54)$	$\mathcal{N}(712.24, 427.35)$	$\mathcal{N}(481.43, 288.86)$	$\mathcal{N}(795.38, 477.23)$	$\mathcal{N}(1036.07, 621.64)$	$\mathcal{N}(56.22, 33.73)$	$\mathcal{N}(354.28, 212.57)$
$Y_{LA\|LAB}^{Glc}$	$\mathcal{N}(7.59, 4.55)$	$\mathcal{N}(12.11, 7.27)$	-	-	-	-	$\mathcal{N}(9.73, 5.84)$	$\mathcal{N}(6.51, 3.91)$
$Y_{LA\|LAB}^{Fru}$	$\mathcal{N}(7.59, 4.55)$	$\mathcal{N}(12.11, 7.27)$	-	-	-	-	$\mathcal{N}(9.73, 5.84)$	$\mathcal{N}(6.51, 3.91)$
$Y_{LA\|AAB}$	$\mathcal{N}(2243.66, 1346.19)$	$\mathcal{N}(24.95, 14.97)$	-	-	-	-	$\mathcal{N}(55.21, 33.12)$	$\mathcal{N}(286.16, 171.7)$
$Y_{LA\|Y}$	$\mathcal{N}(4.5, 2.7)$	$\mathcal{N}(1.18, 0.71)$	-	-	-	-	$\mathcal{N}(53.96, 32.38)$	$\mathcal{N}(11.47, 6.88)$
$Y_{Ac\|LAB}^{Glc}$	$\mathcal{N}(5.42, 3.25)$	$\mathcal{N}(120.98, 72.59)$	$\mathcal{N}(19.62, 11.77)$	$\mathcal{N}(2.7, 1.62)$	$\mathcal{N}(7.26, 4.36)$	$\mathcal{N}(3.47, 2.08)$	$\mathcal{N}(22.99, 13.79)$	$\mathcal{N}(15.95, 9.57)$
$Y_{Ac\|LAB}^{Fru}$	$\mathcal{N}(5.42, 3.25)$	$\mathcal{N}(120.98, 72.59)$	$\mathcal{N}(19.62, 11.77)$	$\mathcal{N}(2.7, 1.62)$	$\mathcal{N}(7.26, 4.36)$	$\mathcal{N}(3.47, 2.08)$	$\mathcal{N}(22.99, 13.79)$	$\mathcal{N}(15.95, 9.57)$
$Y_{Ac\|AAB}^{EtOH}$	$\mathcal{N}(1602.91, 961.75)$	$\mathcal{N}(249.15, 149.49)$	$\mathcal{N}(1072.81, 643.69)$	$\mathcal{N}(154.45, 92.67)$	$\mathcal{N}(525.57, 315.34)$	$\mathcal{N}(601.38, 360.83)$	$\mathcal{N}(130.4, 78.24)$	$\mathcal{N}(700.87, 420.52)$
$Y_{Ac\|AAB}^{LA}$	$\mathcal{N}(1602.91, 961.75)$	$\mathcal{N}(249.15, 149.49)$	$\mathcal{N}(1072.81, 643.69)$	$\mathcal{N}(154.45, 92.67)$	$\mathcal{N}(525.57, 315.34)$	$\mathcal{N}(601.38, 360.83)$	$\mathcal{N}(130.4, 78.24)$	$\mathcal{N}(700.87, 420.52)$
$Y_{Ac\|Y}^{Glc}$	$\mathcal{N}(3.21, 1.93)$	$\mathcal{N}(11.74, 7.04)$	$\mathcal{N}(15.55, 9.33)$	$\mathcal{N}(6, 3.6)$	$\mathcal{N}(9.41, 5.65)$	$\mathcal{N}(17.63, 10.58)$	$\mathcal{N}(127.45, 76.47)$	$\mathcal{N}(28.1, 16.86)$
$Y_{Ac\|Y}^{Fru}$	$\mathcal{N}(3.21, 1.93)$	$\mathcal{N}(11.74, 7.04)$	$\mathcal{N}(15.55, 9.33)$	$\mathcal{N}(6, 3.6)$	$\mathcal{N}(9.41, 5.65)$	$\mathcal{N}(17.63, 10.58)$	$\mathcal{N}(127.45, 76.47)$	$\mathcal{N}(28.1, 16.86)$
$Y_{Ac\|AAB}$	$\mathcal{N}(1602.91, 961.75)$	$\mathcal{N}(249.15, 149.49)$	$\mathcal{N}(1072.81, 643.69)$	$\mathcal{N}(154.45, 92.67)$	$\mathcal{N}(525.57, 315.34)$	$\mathcal{N}(601.38, 360.83)$	$\mathcal{N}(130.4, 78.24)$	$\mathcal{N}(700.87, 420.52)$
b_{EtOH}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{LA}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{Ac}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$

Table B.2 **(cont.)**: Prior distributions in the range of original units for parameters of all model iterations per dataset. All priors are truncated to the positive set of real numbers.

Parameter	brwb2	ecpt1	ecpt2	ecwb1	ecwb2	brpb1	brst1	mywb1
$\mu_{\max}^{Y_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{EtOH}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{Ac}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
K_{Glc}^{Y}	$\mathcal{N}(21.47, 12.88)$	$\mathcal{N}(24.56, 14.74)$	$\mathcal{N}(21.33, 12.8)$	$\mathcal{N}(21.04, 12.62)$	$\mathcal{N}(18.34, 11)$	$\mathcal{N}(51.17, 30.7)$	$\mathcal{N}(66.24, 39.74)$	$\mathcal{N}(23.09, 13.85)$
K_{Fru}^{Y}	$\mathcal{N}(33.62, 20.17)$	$\mathcal{N}(26.97, 16.18)$	$\mathcal{N}(19.46, 11.68)$	$\mathcal{N}(26.95, 16.17)$	$\mathcal{N}(16.44, 9.87)$	$\mathcal{N}(47.53, 28.52)$	$\mathcal{N}(61.27, 36.76)$	$\mathcal{N}(27.89, 16.73)$
K_{LA}^{Y}	$\mathcal{N}(33.62, 20.17)$	$\mathcal{N}(26.97, 16.18)$	$\mathcal{N}(19.46, 11.68)$	$\mathcal{N}(26.95, 16.17)$	$\mathcal{N}(16.44, 9.87)$	$\mathcal{N}(47.53, 28.52)$	$\mathcal{N}(61.27, 36.76)$	$\mathcal{N}(27.89, 16.73)$
K_{Glc}^{LAB}	$\mathcal{N}(21.47, 12.88)$	$\mathcal{N}(24.56, 14.74)$	$\mathcal{N}(21.33, 12.8)$	$\mathcal{N}(21.04, 12.62)$	$\mathcal{N}(18.34, 11)$	$\mathcal{N}(51.17, 30.7)$	$\mathcal{N}(66.24, 39.74)$	$\mathcal{N}(23.09, 13.85)$
K_{Fru}^{LAB}	$\mathcal{N}(33.62, 20.17)$	$\mathcal{N}(26.97, 16.18)$	$\mathcal{N}(19.46, 11.68)$	$\mathcal{N}(26.95, 16.17)$	$\mathcal{N}(16.44, 9.87)$	$\mathcal{N}(47.53, 28.52)$	$\mathcal{N}(61.27, 36.76)$	$\mathcal{N}(27.89, 16.73)$
K_{EtOH}^{AAB}	$\mathcal{N}(3.53, 2.12)$	$\mathcal{N}(6.27, 3.76)$	$\mathcal{N}(4.81, 2.89)$	$\mathcal{N}(8.92, 5.35)$	$\mathcal{N}(6.26, 3.76)$	$\mathcal{N}(32.66, 19.6)$	$\mathcal{N}(39.08, 23.45)$	$\mathcal{N}(8.61, 5.16)$
K_{LA}^{AAB}	$\mathcal{N}(76.39, 1.24)$	$\mathcal{N}(173.95, 1.81)$	$\mathcal{N}(41.5, 0.49)$	$\mathcal{N}(929.25, 1.52)$	$\mathcal{N}(471.83, 1.55)$	$\mathcal{N}(26006.84, 9.43)$	$\mathcal{N}(58000.95, 7.04)$	$\mathcal{N}(1844.4, 1.7)$
K_{Ac}^{AAB}	$\mathcal{N}(9.31, 5.59)$	$\mathcal{N}(1.96, 1.18)$	$\mathcal{N}(8.67, 5.2)$	$\mathcal{N}(3.92, 2.35)$	$\mathcal{N}(12.05, 7.23)$	$\mathcal{N}(34.36, 20.62)$	$\mathcal{N}(9.36, 5.61)$	$\mathcal{N}(3.07, 1.84)$
k_{Y}	$\mathcal{N}(0.0709, 0.0425)$	$\mathcal{N}(0.0399, 0.0239)$	$\mathcal{N}(0.052, 0.0312)$	$\mathcal{N}(0.028, 0.0168)$	$\mathcal{N}(0.0399, 0.024)$	$\mathcal{N}(0.0077, 0.0046)$	$\mathcal{N}(0.0064, 0.0038)$	$\mathcal{N}(0.0291, 0.0174)$
k_{LAB}	$\mathcal{N}(0.1213, 0.0728)$	$\mathcal{N}(0.0829, 0.0497)$	$\mathcal{N}(0.306, 0.1836)$	$\mathcal{N}(0.0988, 0.0593)$	$\mathcal{N}(0.097, 0.0582)$	$\mathcal{N}(0.0159, 0.0095)$	$\mathcal{N}(0.0213, 0.0128)$	$\mathcal{N}(0.0881, 0.0529)$
k_{AAB}	$\mathcal{N}(0.0014, 9 \times 10^{-4})$	$\mathcal{N}(0.0325, 0.0195)$	$\mathcal{N}(0.0017, 0.001)$	$\mathcal{N}(0.0081, 0.0049)$	$\mathcal{N}(9 \times 10^{-4}, 5 \times 10^{-4})$	$\mathcal{N}(1 \times 10^{-4}, 1 \times 10^{-4})$	$\mathcal{N}(0.0014, 9 \times 10^{-4})$	$\mathcal{N}(0.0133, 0.008)$
$Y_{Glc\|Y}$	$\mathcal{N}(88.25, 52.95)$	$\mathcal{N}(22.45, 13.47)$	$\mathcal{N}(24.48, 14.69)$	$\mathcal{N}(25, 15)$	$\mathcal{N}(27.68, 16.61)$	$\mathcal{N}(93.31, 55.99)$	$\mathcal{N}(90.37, 54.22)$	$\mathcal{N}(52.51, 31.51)$
$Y_{Glc\|LAB}$	$\mathcal{N}(30.68, 18.41)$	$\mathcal{N}(44.09, 26.45)$	$\mathcal{N}(71.12, 42.67)$	$\mathcal{N}(32.52, 19.51)$	$\mathcal{N}(28.6, 17.16)$	$\mathcal{N}(152.45, 91.47)$	$\mathcal{N}(594.54, 356.73)$	$\mathcal{N}(28.41, 17.04)$
$Y_{Fru\|Y}$	$\mathcal{N}(138.22, 82.93)$	$\mathcal{N}(24.65, 14.79)$	$\mathcal{N}(22.34, 13.4)$	$\mathcal{N}(32.02, 19.21)$	$\mathcal{N}(24.82, 14.89)$	$\mathcal{N}(86.67, 52)$	$\mathcal{N}(83.59, 50.15)$	$\mathcal{N}(63.44, 38.07)$
$Y_{Fru\|LAB}$	$\mathcal{N}(48.06, 28.83)$	$\mathcal{N}(48.41, 29.04)$	$\mathcal{N}(64.9, 38.94)$	$\mathcal{N}(41.65, 24.99)$	$\mathcal{N}(25.65, 15.39)$	$\mathcal{N}(141.6, 84.96)$	$\mathcal{N}(549.96, 329.98)$	$\mathcal{N}(34.32, 20.59)$
$Y_{EtOH\|Y}^{Glc}$	$\mathcal{N}(14.5, 8.7)$	$\mathcal{N}(5.73, 3.44)$	$\mathcal{N}(5.52, 3.31)$	$\mathcal{N}(10.6, 6.36)$	$\mathcal{N}(9.45, 5.67)$	$\mathcal{N}(59.56, 35.74)$	$\mathcal{N}(53.31, 31.99)$	$\mathcal{N}(19.57, 11.74)$
$Y_{EtOH\|Y}^{Fru}$	$\mathcal{N}(14.5, 8.7)$	$\mathcal{N}(5.73, 3.44)$	$\mathcal{N}(5.52, 3.31)$	$\mathcal{N}(10.6, 6.36)$	$\mathcal{N}(9.45, 5.67)$	$\mathcal{N}(59.56, 35.74)$	$\mathcal{N}(53.31, 31.99)$	$\mathcal{N}(19.57, 11.74)$
$Y_{EtOH\|Y}^{LA}$	$\mathcal{N}(14.5, 8.7)$	$\mathcal{N}(5.73, 3.44)$	$\mathcal{N}(5.52, 3.31)$	$\mathcal{N}(10.6, 6.36)$	$\mathcal{N}(9.45, 5.67)$	$\mathcal{N}(59.56, 35.74)$	$\mathcal{N}(53.31, 31.99)$	$\mathcal{N}(19.57, 11.74)$
$Y_{EtOH\|LAB}^{Glc}$	$\mathcal{N}(5.04, 3.03)$	$\mathcal{N}(11.25, 6.75)$	$\mathcal{N}(16.04, 9.62)$	$\mathcal{N}(13.79, 8.28)$	$\mathcal{N}(9.76, 5.86)$	$\mathcal{N}(97.31, 58.38)$	$\mathcal{N}(350.75, 210.45)$	$\mathcal{N}(10.59, 6.35)$
$Y_{EtOH\|LAB}^{Fru}$	$\mathcal{N}(5.04, 3.03)$	$\mathcal{N}(11.25, 6.75)$	$\mathcal{N}(16.04, 9.62)$	$\mathcal{N}(13.79, 8.28)$	$\mathcal{N}(9.76, 5.86)$	$\mathcal{N}(97.31, 58.38)$	$\mathcal{N}(350.75, 210.45)$	$\mathcal{N}(10.59, 6.35)$
$Y_{EtOH\|AAB}$	$\mathcal{N}(130.75, 78.45)$	$\mathcal{N}(361.41, 216.85)$	$\mathcal{N}(244.26, 146.56)$	$\mathcal{N}(3276.25, 1965.75)$	$\mathcal{N}(1146.34, 687.8)$	$\mathcal{N}(54062.78, 32437.67)$	$\mathcal{N}(193086.14, 115851.68)$	$\mathcal{N}(5592.66, 3355.59)$
$Y_{LA\|LAB}^{Glc}$	$\mathcal{N}(2.95, 1.77)$	$\mathcal{N}(5.42, 3.25)$	$\mathcal{N}(2.72, 1.63)$	$\mathcal{N}(3.91, 2.35)$	$\mathcal{N}(4.02, 2.41)$	$\mathcal{N}(46.81, 28.09)$	$\mathcal{N}(105.36, 63.22)$	$\mathcal{N}(3.49, 2.1)$
$Y_{LA\|LAB}^{Fru}$	$\mathcal{N}(2.95, 1.77)$	$\mathcal{N}(5.42, 3.25)$	$\mathcal{N}(2.72, 1.63)$	$\mathcal{N}(3.91, 2.35)$	$\mathcal{N}(4.02, 2.41)$	$\mathcal{N}(46.81, 28.09)$	$\mathcal{N}(105.36, 63.22)$	$\mathcal{N}(3.49, 2.1)$
$Y_{LA\|AAB}$	$\mathcal{N}(76.39, 45.83)$	$\mathcal{N}(173.95, 104.37)$	$\mathcal{N}(41.5, 24.9)$	$\mathcal{N}(929.25, 557.55)$	$\mathcal{N}(471.83, 283.1)$	$\mathcal{N}(26006.84, 15604.1)$	$\mathcal{N}(58000.95, 34800.57)$	$\mathcal{N}(1844.4, 1106.64)$
$Y_{LA\|Y}$	$\mathcal{N}(8.47, 5.08)$	$\mathcal{N}(2.76, 1.65)$	$\mathcal{N}(0.94, 0.56)$	$\mathcal{N}(3.01, 1.8)$	$\mathcal{N}(3.89, 2.33)$	$\mathcal{N}(28.65, 17.19)$	$\mathcal{N}(16.01, 9.61)$	$\mathcal{N}(6.46, 3.87)$
$Y_{Ac\|LAB}^{Glc}$	$\mathcal{N}(13.31, 7.99)$	$\mathcal{N}(3.52, 2.11)$	$\mathcal{N}(28.91, 17.35)$	$\mathcal{N}(6.06, 3.64)$	$\mathcal{N}(18.79, 11.28)$	$\mathcal{N}(102.37, 61.42)$	$\mathcal{N}(83.97, 50.38)$	$\mathcal{N}(3.77, 2.26)$
$Y_{Ac\|LAB}^{Fru}$	$\mathcal{N}(13.31, 7.99)$	$\mathcal{N}(3.52, 2.11)$	$\mathcal{N}(28.91, 17.35)$	$\mathcal{N}(6.06, 3.64)$	$\mathcal{N}(18.79, 11.28)$	$\mathcal{N}(102.37, 61.42)$	$\mathcal{N}(83.97, 50.38)$	$\mathcal{N}(3.77, 2.26)$
$Y_{Ac\|AAB}^{EtOH}$	$\mathcal{N}(345.07, 207.04)$	$\mathcal{N}(113.06, 67.84)$	$\mathcal{N}(440.34, 264.2)$	$\mathcal{N}(1439.59, 863.75)$	$\mathcal{N}(2206.86, 1324.12)$	$\mathcal{N}(56876.56, 34125.94)$	$\mathcal{N}(46225.84, 27735.5)$	$\mathcal{N}(1993.55, 1196.13)$
$Y_{Ac\|AAB}^{LA}$	$\mathcal{N}(345.07, 207.04)$	$\mathcal{N}(113.06, 67.84)$	$\mathcal{N}(440.34, 264.2)$	$\mathcal{N}(1439.59, 863.75)$	$\mathcal{N}(2206.86, 1324.12)$	$\mathcal{N}(56876.56, 34125.94)$	$\mathcal{N}(46225.84, 27735.5)$	$\mathcal{N}(1993.55, 1196.13)$
$Y_{Ac\|Y}^{Glc}$	$\mathcal{N}(38.28, 22.97)$	$\mathcal{N}(1.79, 1.08)$	$\mathcal{N}(9.95, 5.97)$	$\mathcal{N}(4.66, 2.8)$	$\mathcal{N}(18.19, 10.91)$	$\mathcal{N}(62.66, 37.6)$	$\mathcal{N}(12.76, 7.66)$	$\mathcal{N}(6.98, 4.19)$
$Y_{Ac\|Y}^{Fru}$	$\mathcal{N}(38.28, 22.97)$	$\mathcal{N}(1.79, 1.08)$	$\mathcal{N}(9.95, 5.97)$	$\mathcal{N}(4.66, 2.8)$	$\mathcal{N}(18.19, 10.91)$	$\mathcal{N}(62.66, 37.6)$	$\mathcal{N}(12.76, 7.66)$	$\mathcal{N}(6.98, 4.19)$
$Y_{Ac\|AAB}$	$\mathcal{N}(345.07, 207.04)$	$\mathcal{N}(113.06, 67.84)$	$\mathcal{N}(440.34, 264.2)$	$\mathcal{N}(1439.59, 863.75)$	$\mathcal{N}(2206.86, 1324.12)$	$\mathcal{N}(56876.56, 34125.94)$	$\mathcal{N}(46225.84, 27735.5)$	$\mathcal{N}(1993.55, 1196.13)$
b_{EtOH}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{LA}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{Ac}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$

Table B.2 **(cont.)**: Prior distributions in the range of original units for parameters of all model iterations per dataset. All priors are truncated to the positive set of real numbers.

Parameter	mywb2	ghhp6	ghhp7	brwb3	brwb4	brst2	brwb5	brwb6
$\mu_{\max}^{Y_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{EtOH}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{Ac}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
K_{Glc}^{Y}	$\mathcal{N}(14.2, 8.52)$	$\mathcal{N}(21.27, 12.76)$	$\mathcal{N}(25.54, 15.32)$	$\mathcal{N}(73.29, 43.98)$	$\mathcal{N}(66.49, 39.89)$	$\mathcal{N}(61.44, 36.87)$	$\mathcal{N}(5.02, 3.01)$	$\mathcal{N}(6.71, 4.03)$
K_{Fru}^{Y}	$\mathcal{N}(14.74, 8.84)$	$\mathcal{N}(39.14, 23.49)$	$\mathcal{N}(29.67, 17.8)$	$\mathcal{N}(81.08, 48.65)$	$\mathcal{N}(74.5, 44.7)$	$\mathcal{N}(68.28, 40.97)$	$\mathcal{N}(6.31, 3.78)$	$\mathcal{N}(8.04, 4.82)$
K_{LA}^{Y}	$\mathcal{N}(14.74, 8.84)$	$\mathcal{N}(39.14, 23.49)$	$\mathcal{N}(29.67, 17.8)$	$\mathcal{N}(81.08, 48.65)$	$\mathcal{N}(74.5, 44.7)$	$\mathcal{N}(68.28, 40.97)$	$\mathcal{N}(6.31, 3.78)$	$\mathcal{N}(8.04, 4.82)$
K_{Glc}^{LAB}	$\mathcal{N}(14.2, 8.52)$	$\mathcal{N}(21.27, 12.76)$	$\mathcal{N}(25.54, 15.32)$	$\mathcal{N}(73.29, 43.98)$	$\mathcal{N}(66.49, 39.89)$	$\mathcal{N}(61.44, 36.87)$	$\mathcal{N}(5.02, 3.01)$	$\mathcal{N}(6.71, 4.03)$
K_{Fru}^{LAB}	$\mathcal{N}(14.74, 8.84)$	$\mathcal{N}(39.14, 23.49)$	$\mathcal{N}(29.67, 17.8)$	$\mathcal{N}(81.08, 48.65)$	$\mathcal{N}(74.5, 44.7)$	$\mathcal{N}(68.28, 40.97)$	$\mathcal{N}(6.31, 3.78)$	$\mathcal{N}(8.04, 4.82)$
K_{EtOH}^{AAB}	$\mathcal{N}(6.69, 4.02)$	$\mathcal{N}(4.59, 2.75)$	$\mathcal{N}(9.06, 5.44)$	$\mathcal{N}(31.81, 19.09)$	$\mathcal{N}(23.92, 14.35)$	$\mathcal{N}(35.5, 21.3)$	$\mathcal{N}(1.54, 0.92)$	$\mathcal{N}(1.99, 1.19)$
K_{LA}^{AAB}	$\mathcal{N}(1126.25, 3.27)$	$\mathcal{N}(33.82, 1.04)$	$\mathcal{N}(108.9, 0.85)$	$\mathcal{N}(9973.04, 5.08)$	$\mathcal{N}(17956.46, 6.62)$	$\mathcal{N}(444524.64, 7.13)$	$\mathcal{N}(2072.3, 0.25)$	$\mathcal{N}(6615.52, 0.27)$
K_{Ac}^{AAB}	$\mathcal{N}(3.84, 2.31)$	$\mathcal{N}(3.4, 2.04)$	$\mathcal{N}(4.14, 2.49)$	$\mathcal{N}(18.09, 10.85)$	$\mathcal{N}(13.46, 8.08)$	$\mathcal{N}(28.71, 17.23)$	$\mathcal{N}(0.8, 0.48)$	$\mathcal{N}(0.85, 0.51)$
k_{Y}	$\mathcal{N}(0.0374, 0.0224)$	$\mathcal{N}(0.0545, 0.0327)$	$\mathcal{N}(0.0276, 0.0166)$	$\mathcal{N}(0.0079, 0.0047)$	$\mathcal{N}(0.0105, 0.0063)$	$\mathcal{N}(0.007, 0.0042)$	$\mathcal{N}(0.1627, 0.0976)$	$\mathcal{N}(0.1258, 0.0755)$
k_{LAB}	$\mathcal{N}(0.0458, 0.0275)$	$\mathcal{N}(0.1444, 0.0867)$	$\mathcal{N}(0.1757, 0.1054)$	$\mathcal{N}(0.0295, 0.0177)$	$\mathcal{N}(0.0227, 0.0136)$	$\mathcal{N}(0.021, 0.0126)$	$\mathcal{N}(0.5903, 0.3542)$	$\mathcal{N}(0.5482, 0.3289)$
k_{AAB}	$\mathcal{N}(0.0085, 0.0051)$	$\mathcal{N}(0.0108, 0.0065)$	$\mathcal{N}(0.0073, 0.0044)$	$\mathcal{N}(4 \times 10^{-4}, 2 \times 10^{-4})$	$\mathcal{N}(7 \times 10^{-4}, 4 \times 10^{-4})$	$\mathcal{N}(2 \times 10^{-4}, 1 \times 10^{-4})$	$\mathcal{N}(0.197, 0.1182)$	$\mathcal{N}(0.1742, 0.1045)$
$Y_{Glc\|Y}$	$\mathcal{N}(72.33, 43.4)$	$\mathcal{N}(154.4, 92.64)$	$\mathcal{N}(21.14, 12.69)$	$\mathcal{N}(38.72, 23.23)$	$\mathcal{N}(122.92, 73.75)$	$\mathcal{N}(199.25, 119.55)$	$\mathcal{N}(833.73, 500.24)$	$\mathcal{N}(556.73, 334.04)$
$Y_{Glc\|LAB}$	$\mathcal{N}(351.94, 211.16)$	$\mathcal{N}(26.29, 15.78)$	$\mathcal{N}(14.01, 8.4)$	$\mathcal{N}(587.71, 352.62)$	$\mathcal{N}(195.35, 117.21)$	$\mathcal{N}(1070.46, 642.27)$	$\mathcal{N}(2119.36, 1271.62)$	$\mathcal{N}(10588.37, 6353.02)$
$Y_{Fru\|Y}$	$\mathcal{N}(75.05, 45.03)$	$\mathcal{N}(284.17, 170.5)$	$\mathcal{N}(24.56, 14.74)$	$\mathcal{N}(42.84, 25.7)$	$\mathcal{N}(137.74, 82.65)$	$\mathcal{N}(221.42, 132.85)$	$\mathcal{N}(1048.14, 628.89)$	$\mathcal{N}(666.52, 399.91)$
$Y_{Fru\|LAB}$	$\mathcal{N}(365.19, 219.11)$	$\mathcal{N}(48.39, 29.03)$	$\mathcal{N}(16.27, 9.76)$	$\mathcal{N}(650.11, 390.06)$	$\mathcal{N}(218.9, 131.34)$	$\mathcal{N}(1189.6, 713.76)$	$\mathcal{N}(2664.41, 1598.65)$	$\mathcal{N}(12676.39, 7605.84)$
$Y_{EtOH\|Y}^{Glc}$	$\mathcal{N}(34.08, 20.45)$	$\mathcal{N}(33.32, 19.99)$	$\mathcal{N}(7.5, 4.5)$	$\mathcal{N}(16.81, 10.08)$	$\mathcal{N}(44.22, 26.53)$	$\mathcal{N}(115.12, 69.07)$	$\mathcal{N}(255.47, 153.28)$	$\mathcal{N}(164.85, 98.91)$
$Y_{EtOH\|Y}^{Fru}$	$\mathcal{N}(34.08, 20.45)$	$\mathcal{N}(33.32, 19.99)$	$\mathcal{N}(7.5, 4.5)$	$\mathcal{N}(16.81, 10.08)$	$\mathcal{N}(44.22, 26.53)$	$\mathcal{N}(115.12, 69.07)$	$\mathcal{N}(255.47, 153.28)$	$\mathcal{N}(164.85, 98.91)$
$Y_{EtOH\|Y}^{LA}$	$\mathcal{N}(34.08, 20.45)$	$\mathcal{N}(33.32, 19.99)$	$\mathcal{N}(7.5, 4.5)$	$\mathcal{N}(16.81, 10.08)$	$\mathcal{N}(44.22, 26.53)$	$\mathcal{N}(115.12, 69.07)$	$\mathcal{N}(255.47, 153.28)$	$\mathcal{N}(164.85, 98.91)$
$Y_{EtOH\|LAB}^{Glc}$	$\mathcal{N}(165.82, 99.49)$	$\mathcal{N}(5.67, 3.4)$	$\mathcal{N}(4.97, 2.98)$	$\mathcal{N}(255.09, 153.06)$	$\mathcal{N}(70.28, 42.17)$	$\mathcal{N}(618.5, 371.1)$	$\mathcal{N}(649.41, 389.65)$	$\mathcal{N}(3135.19, 1881.12)$
$Y_{EtOH\|LAB}^{Fru}$	$\mathcal{N}(165.82, 99.49)$	$\mathcal{N}(5.67, 3.4)$	$\mathcal{N}(4.97, 2.98)$	$\mathcal{N}(255.09, 153.06)$	$\mathcal{N}(70.28, 42.17)$	$\mathcal{N}(618.5, 371.1)$	$\mathcal{N}(649.41, 389.65)$	$\mathcal{N}(3135.19, 1881.12)$
$Y_{EtOH\|AAB}$	$\mathcal{N}(1381.65, 828.99)$	$\mathcal{N}(89.68, 53.81)$	$\mathcal{N}(693.53, 416.12)$	$\mathcal{N}(37450.12, 22470.07)$	$\mathcal{N}(38957.79, 23374.68)$	$\mathcal{N}(1327683.83, 796610.3)$	$\mathcal{N}(7520.97, 4512.58)$	$\mathcal{N}(28841.36, 17304.81)$
$Y_{LA\|LAB}^{Glc}$	$\mathcal{N}(135.17, 81.1)$	$\mathcal{N}(2.14, 1.28)$	$\mathcal{N}(0.78, 0.47)$	$\mathcal{N}(67.93, 40.76)$	$\mathcal{N}(32.39, 19.44)$	$\mathcal{N}(207.08, 124.25)$	$\mathcal{N}(178.94, 107.36)$	$\mathcal{N}(719.14, 431.48)$
$Y_{LA\|LAB}^{Fru}$	$\mathcal{N}(135.17, 81.1)$	$\mathcal{N}(2.14, 1.28)$	$\mathcal{N}(0.78, 0.47)$	$\mathcal{N}(67.93, 40.76)$	$\mathcal{N}(32.39, 19.44)$	$\mathcal{N}(207.08, 124.25)$	$\mathcal{N}(178.94, 107.36)$	$\mathcal{N}(719.14, 431.48)$
$Y_{LA\|AAB}$	$\mathcal{N}(1126.25, 675.75)$	$\mathcal{N}(33.82, 20.29)$	$\mathcal{N}(108.9, 65.34)$	$\mathcal{N}(9973.04, 5983.83)$	$\mathcal{N}(17956.46, 10773.88)$	$\mathcal{N}(444524.64, 266714.78)$	$\mathcal{N}(2072.3, 1243.38)$	$\mathcal{N}(6615.52, 3969.31)$
$Y_{LA\|Y}$	$\mathcal{N}(27.78, 16.67)$	$\mathcal{N}(12.57, 7.54)$	$\mathcal{N}(1.18, 0.71)$	$\mathcal{N}(4.48, 2.69)$	$\mathcal{N}(20.38, 12.23)$	$\mathcal{N}(38.54, 23.13)$	$\mathcal{N}(70.39, 42.23)$	$\mathcal{N}(37.81, 22.69)$
$Y_{Ac\|LAB}^{Glc}$	$\mathcal{N}(95.23, 57.14)$	$\mathcal{N}(4.2, 2.52)$	$\mathcal{N}(2.27, 1.36)$	$\mathcal{N}(145.04, 87.02)$	$\mathcal{N}(39.55, 23.73)$	$\mathcal{N}(500.19, 300.11)$	$\mathcal{N}(336.54, 201.92)$	$\mathcal{N}(1335.77, 801.46)$
$Y_{Ac\|LAB}^{Fru}$	$\mathcal{N}(95.23, 57.14)$	$\mathcal{N}(4.2, 2.52)$	$\mathcal{N}(2.27, 1.36)$	$\mathcal{N}(145.04, 87.02)$	$\mathcal{N}(39.55, 23.73)$	$\mathcal{N}(500.19, 300.11)$	$\mathcal{N}(336.54, 201.92)$	$\mathcal{N}(1335.77, 801.46)$
$Y_{Ac\|AAB}^{EtOH}$	$\mathcal{N}(793.44, 476.06)$	$\mathcal{N}(66.39, 39.84)$	$\mathcal{N}(317.11, 190.27)$	$\mathcal{N}(21293.36, 12776.02)$	$\mathcal{N}(21923.17, 13153.9)$	$\mathcal{N}(1073699.13, 644219.48)$	$\mathcal{N}(3897.5, 2338.5)$	$\mathcal{N}(12288.04, 7372.83)$
$Y_{Ac\|AAB}^{LA}$	$\mathcal{N}(793.44, 476.06)$	$\mathcal{N}(66.39, 39.84)$	$\mathcal{N}(317.11, 190.27)$	$\mathcal{N}(21293.36, 12776.02)$	$\mathcal{N}(21923.17, 13153.9)$	$\mathcal{N}(1073699.13, 644219.48)$	$\mathcal{N}(3897.5, 2338.5)$	$\mathcal{N}(12288.04, 7372.83)$
$Y_{Ac\|Y}^{Glc}$	$\mathcal{N}(19.57, 11.74)$	$\mathcal{N}(24.67, 14.8)$	$\mathcal{N}(3.43, 2.06)$	$\mathcal{N}(9.56, 5.73)$	$\mathcal{N}(24.89, 14.93)$	$\mathcal{N}(93.1, 55.86)$	$\mathcal{N}(132.39, 79.43)$	$\mathcal{N}(70.23, 42.14)$
$Y_{Ac\|Y}^{Fru}$	$\mathcal{N}(19.57, 11.74)$	$\mathcal{N}(24.67, 14.8)$	$\mathcal{N}(3.43, 2.06)$	$\mathcal{N}(9.56, 5.73)$	$\mathcal{N}(24.89, 14.93)$	$\mathcal{N}(93.1, 55.86)$	$\mathcal{N}(132.39, 79.43)$	$\mathcal{N}(70.23, 42.14)$
$Y_{Ac\|AAB}$	$\mathcal{N}(793.44, 476.06)$	$\mathcal{N}(66.39, 39.84)$	$\mathcal{N}(317.11, 190.27)$	$\mathcal{N}(21293.36, 12776.02)$	$\mathcal{N}(21923.17, 13153.9)$	$\mathcal{N}(1073699.13, 644219.48)$	$\mathcal{N}(3897.5, 2338.5)$	$\mathcal{N}(12288.04, 7372.83)$
b_{EtOH}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{LA}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{Ac}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$

Table B.2 **(cont.)**: Prior distributions in the range of original units for parameters of all model iterations per dataset. All priors are truncated to the positive set of real numbers.

Parameter	brwb7	mywb3	brwb8	hnwb1	ecpb1	niwb1	niwb2	ecpb2
$\mu_{\max}^{Y_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{Y_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Glc}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{LAB_{Fru}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{EtOH}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{LA}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
$\mu_{\max}^{AAB_{Ac}}$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
K_{Glc}^{Y}	$\mathcal{N}(7.85, 4.71)$	$\mathcal{N}(19.96, 11.98)$	$\mathcal{N}(40.87, 24.52)$	$\mathcal{N}(21.78, 13.07)$	$\mathcal{N}(14.95, 8.97)$	$\mathcal{N}(12.18, 7.31)$	$\mathcal{N}(10.71, 6.43)$	$\mathcal{N}(14.89, 8.94)$
K_{Fru}^{Y}	$\mathcal{N}(9.29, 5.57)$	$\mathcal{N}(25.92, 15.55)$	$\mathcal{N}(43.24, 25.95)$	$\mathcal{N}(26.8, 16.08)$	$\mathcal{N}(18.32, 10.99)$	$\mathcal{N}(15.62, 9.38)$	$\mathcal{N}(12.77, 7.66)$	$\mathcal{N}(17.64, 10.59)$
K_{LA}^{Y}	$\mathcal{N}(9.29, 5.57)$	$\mathcal{N}(25.92, 15.55)$	$\mathcal{N}(43.24, 25.95)$	$\mathcal{N}(26.8, 16.08)$	$\mathcal{N}(18.32, 10.99)$	$\mathcal{N}(15.62, 9.38)$	$\mathcal{N}(12.77, 7.66)$	$\mathcal{N}(17.64, 10.59)$
K_{Glc}^{LAB}	$\mathcal{N}(7.85, 4.71)$	$\mathcal{N}(19.96, 11.98)$	$\mathcal{N}(40.87, 24.52)$	$\mathcal{N}(21.78, 13.07)$	$\mathcal{N}(14.95, 8.97)$	$\mathcal{N}(12.18, 7.31)$	$\mathcal{N}(10.71, 6.43)$	$\mathcal{N}(14.89, 8.94)$
K_{Fru}^{LAB}	$\mathcal{N}(9.29, 5.57)$	$\mathcal{N}(25.92, 15.55)$	$\mathcal{N}(43.24, 25.95)$	$\mathcal{N}(26.8, 16.08)$	$\mathcal{N}(18.32, 10.99)$	$\mathcal{N}(15.62, 9.38)$	$\mathcal{N}(12.77, 7.66)$	$\mathcal{N}(17.64, 10.59)$
K_{EtOH}^{AAB}	$\mathcal{N}(4.89, 2.93)$	$\mathcal{N}(8.61, 5.17)$	$\mathcal{N}(5.41, 3.25)$	$\mathcal{N}(10.93, 6.56)$	$\mathcal{N}(9.95, 5.97)$	$\mathcal{N}(8.71, 5.22)$	$\mathcal{N}(9.37, 5.62)$	$\mathcal{N}(19.3, 11.58)$
K_{LA}^{AAB}	$\mathcal{N}(1648.18, 0.3)$	$\mathcal{N}(453.45, 3.58)$	$\mathcal{N}(125.38, 3.35)$	$\mathcal{N}(242.02, 1.47)$	$\mathcal{N}(0.98, 3.55)$	$\mathcal{N}(596.41, 1.7)$	$\mathcal{N}(141.89, 2.33)$	$\mathcal{N}(13.6, 1.58)$
K_{Ac}^{AAB}	$\mathcal{N}(1.72, 1.03)$	$\mathcal{N}(4.96, 2.98)$	$\mathcal{N}(0.6, 0.36)$	$\mathcal{N}(7.15, 4.29)$	$\mathcal{N}(12.35, 7.41)$	$\mathcal{N}(7.92, 4.75)$	$\mathcal{N}(4.22, 2.53)$	$\mathcal{N}(4.01, 2.41)$
k_{Y}	$\mathcal{N}(0.0512, 0.0307)$	$\mathcal{N}(0.029, 0.0174)$	$\mathcal{N}(0.0462, 0.0277)$	$\mathcal{N}(0.0229, 0.0137)$	$\mathcal{N}(0.0251, 0.0151)$	$\mathcal{N}(0.0287, 0.0172)$	$\mathcal{N}(0.0267, 0.016)$	$\mathcal{N}(0.013, 0.0078)$
k_{LAB}	$\mathcal{N}(0.4995, 0.2997)$	$\mathcal{N}(0.0419, 0.0251)$	$\mathcal{N}(0.0448, 0.0269)$	$\mathcal{N}(0.1023, 0.0614)$	$\mathcal{N}(0.0423, 0.0254)$	$\mathcal{N}(0.0882, 0.0529)$	$\mathcal{N}(0.0644, 0.0386)$	$\mathcal{N}(0.0951, 0.0571)$
k_{AAB}	$\mathcal{N}(0.0423, 0.0254)$	$\mathcal{N}(0.0051, 0.003)$	$\mathcal{N}(0.3478, 0.2087)$	$\mathcal{N}(0.0024, 0.0015)$	$\mathcal{N}(8 \times 10^{-4}, 5 \times 10^{-4})$	$\mathcal{N}(0.002, 0.0012)$	$\mathcal{N}(0.007, 0.0042)$	$\mathcal{N}(0.0078, 0.0047)$
$Y_{Glc\|Y}$	$\mathcal{N}(771.56, 462.94)$	$\mathcal{N}(77.29, 46.38)$	$\mathcal{N}(2.85, 1.71)$	$\mathcal{N}(84.12, 50.47)$	$\mathcal{N}(0.2, 0.12)$	$\mathcal{N}(13.47, 8.08)$	$\mathcal{N}(9.2, 5.52)$	$\mathcal{N}(4.82, 2.89)$
$Y_{Glc\|LAB}$	$\mathcal{N}(5566.1, 3339.66)$	$\mathcal{N}(14.8, 8.88)$	$\mathcal{N}(191.21, 114.72)$	$\mathcal{N}(64.14, 38.48)$	$\mathcal{N}(0.55, 0.33)$	$\mathcal{N}(756.1, 453.66)$	$\mathcal{N}(52.82, 31.69)$	$\mathcal{N}(12.39, 7.43)$
$Y_{Fru\|Y}$	$\mathcal{N}(913.39, 548.04)$	$\mathcal{N}(100.35, 60.21)$	$\mathcal{N}(3.02, 1.81)$	$\mathcal{N}(103.53, 62.12)$	$\mathcal{N}(0.25, 0.15)$	$\mathcal{N}(17.29, 10.37)$	$\mathcal{N}(10.96, 6.58)$	$\mathcal{N}(5.71, 3.43)$
$Y_{Fru\|LAB}$	$\mathcal{N}(6589.31, 3953.58)$	$\mathcal{N}(19.22, 11.53)$	$\mathcal{N}(202.3, 121.38)$	$\mathcal{N}(78.93, 47.36)$	$\mathcal{N}(0.68, 0.41)$	$\mathcal{N}(970.31, 582.19)$	$\mathcal{N}(62.97, 37.78)$	$\mathcal{N}(14.68, 8.81)$
$Y_{EtOH\|Y}^{Glc}$	$\mathcal{N}(480.52, 288.31)$	$\mathcal{N}(33.35, 20.01)$	$\mathcal{N}(0.38, 0.23)$	$\mathcal{N}(42.23, 25.34)$	$\mathcal{N}(0.14, 0.08)$	$\mathcal{N}(9.63, 5.78)$	$\mathcal{N}(8.05, 4.83)$	$\mathcal{N}(6.24, 3.75)$
$Y_{EtOH\|Y}^{Fru}$	$\mathcal{N}(480.52, 288.31)$	$\mathcal{N}(33.35, 20.01)$	$\mathcal{N}(0.38, 0.23)$	$\mathcal{N}(42.23, 25.34)$	$\mathcal{N}(0.14, 0.08)$	$\mathcal{N}(9.63, 5.78)$	$\mathcal{N}(8.05, 4.83)$	$\mathcal{N}(6.24, 3.75)$
$Y_{EtOH\|Y}^{LA}$	$\mathcal{N}(480.52, 288.31)$	$\mathcal{N}(33.35, 20.01)$	$\mathcal{N}(0.38, 0.23)$	$\mathcal{N}(42.23, 25.34)$	$\mathcal{N}(0.14, 0.08)$	$\mathcal{N}(9.63, 5.78)$	$\mathcal{N}(8.05, 4.83)$	$\mathcal{N}(6.24, 3.75)$
$Y_{EtOH\|LAB}^{Glc}$	$\mathcal{N}(3466.49, 2079.89)$	$\mathcal{N}(6.39, 3.83)$	$\mathcal{N}(25.33, 15.2)$	$\mathcal{N}(32.19, 19.32)$	$\mathcal{N}(0.37, 0.22)$	$\mathcal{N}(540.64, 324.38)$	$\mathcal{N}(46.23, 27.74)$	$\mathcal{N}(16.06, 9.63)$
$Y_{EtOH\|LAB}^{Fru}$	$\mathcal{N}(3466.49, 2079.89)$	$\mathcal{N}(6.39, 3.83)$	$\mathcal{N}(25.33, 15.2)$	$\mathcal{N}(32.19, 19.32)$	$\mathcal{N}(0.37, 0.22)$	$\mathcal{N}(540.64, 324.38)$	$\mathcal{N}(46.23, 27.74)$	$\mathcal{N}(16.06, 9.63)$
$Y_{EtOH\|AAB}$	$\mathcal{N}(16093.18, 9655.91)$	$\mathcal{N}(654.59, 392.76)$	$\mathcal{N}(121.73, 73.04)$	$\mathcal{N}(1082.73, 649.64)$	$\mathcal{N}(1.64, 0.99)$	$\mathcal{N}(1830.88, 1098.53)$	$\mathcal{N}(342.53, 205.52)$	$\mathcal{N}(99.86, 59.92)$
$Y_{LA\|LAB}^{Glc}$	$\mathcal{N}(355.02, 213.01)$	$\mathcal{N}(4.42, 2.65)$	$\mathcal{N}(26.09, 15.65)$	$\mathcal{N}(7.2, 4.32)$	$\mathcal{N}(0.22, 0.13)$	$\mathcal{N}(176.11, 105.67)$	$\mathcal{N}(19.15, 11.49)$	$\mathcal{N}(2.19, 1.31)$
$Y_{LA\|LAB}^{Fru}$	$\mathcal{N}(355.02, 213.01)$	$\mathcal{N}(4.42, 2.65)$	$\mathcal{N}(26.09, 15.65)$	$\mathcal{N}(7.2, 4.32)$	$\mathcal{N}(0.22, 0.13)$	$\mathcal{N}(176.11, 105.67)$	$\mathcal{N}(19.15, 11.49)$	$\mathcal{N}(2.19, 1.31)$
$Y_{LA\|AAB}$	$\mathcal{N}(1648.18, 988.91)$	$\mathcal{N}(453.45, 272.07)$	$\mathcal{N}(125.38, 75.23)$	$\mathcal{N}(242.02, 145.21)$	$\mathcal{N}(0.98, 0.59)$	$\mathcal{N}(596.41, 357.85)$	$\mathcal{N}(141.89, 85.13)$	$\mathcal{N}(13.6, 8.16)$
$Y_{LA\|Y}$	$\mathcal{N}(49.21, 29.53)$	$\mathcal{N}(23.1, 13.86)$	$\mathcal{N}(0.39, 0.23)$	$\mathcal{N}(9.44, 5.66)$	$\mathcal{N}(0.08, 0.05)$	$\mathcal{N}(3.14, 1.88)$	$\mathcal{N}(3.33, 2)$	$\mathcal{N}(0.85, 0.51)$
$Y_{Ac\|LAB}^{Glc}$	$\mathcal{N}(1219.34, 731.6)$	$\mathcal{N}(3.68, 2.21)$	$\mathcal{N}(2.8, 1.68)$	$\mathcal{N}(21.06, 12.64)$	$\mathcal{N}(0.46, 0.27)$	$\mathcal{N}(492.02, 295.21)$	$\mathcal{N}(20.82, 12.49)$	$\mathcal{N}(3.34, 2)$
$Y_{Ac\|LAB}^{Fru}$	$\mathcal{N}(1219.34, 731.6)$	$\mathcal{N}(3.68, 2.21)$	$\mathcal{N}(2.8, 1.68)$	$\mathcal{N}(21.06, 12.64)$	$\mathcal{N}(0.46, 0.27)$	$\mathcal{N}(492.02, 295.21)$	$\mathcal{N}(20.82, 12.49)$	$\mathcal{N}(3.34, 2)$
$Y_{Ac\|AAB}^{EtOH}$	$\mathcal{N}(5660.77, 3396.46)$	$\mathcal{N}(377.21, 226.33)$	$\mathcal{N}(13.48, 8.09)$	$\mathcal{N}(708.29, 424.97)$	$\mathcal{N}(2.04, 1.22)$	$\mathcal{N}(1666.22, 999.73)$	$\mathcal{N}(154.28, 92.57)$	$\mathcal{N}(20.76, 12.46)$
$Y_{Ac\|AAB}^{LA}$	$\mathcal{N}(5660.77, 3396.46)$	$\mathcal{N}(377.21, 226.33)$	$\mathcal{N}(13.48, 8.09)$	$\mathcal{N}(708.29, 424.97)$	$\mathcal{N}(2.04, 1.22)$	$\mathcal{N}(1666.22, 999.73)$	$\mathcal{N}(154.28, 92.57)$	$\mathcal{N}(20.76, 12.46)$
$Y_{Ac\|Y}^{Glc}$	$\mathcal{N}(169.02, 101.41)$	$\mathcal{N}(19.22, 11.53)$	$\mathcal{N}(0.04, 0.03)$	$\mathcal{N}(27.62, 16.57)$	$\mathcal{N}(0.17, 0.1)$	$\mathcal{N}(8.77, 5.26)$	$\mathcal{N}(3.63, 2.18)$	$\mathcal{N}(1.3, 0.78)$
$Y_{Ac\|Y}^{Fru}$	$\mathcal{N}(169.02, 101.41)$	$\mathcal{N}(19.22, 11.53)$	$\mathcal{N}(0.04, 0.03)$	$\mathcal{N}(27.62, 16.57)$	$\mathcal{N}(0.17, 0.1)$	$\mathcal{N}(8.77, 5.26)$	$\mathcal{N}(3.63, 2.18)$	$\mathcal{N}(1.3, 0.78)$
$Y_{Ac\|AAB}$	$\mathcal{N}(5660.77, 3396.46)$	$\mathcal{N}(377.21, 226.33)$	$\mathcal{N}(13.48, 8.09)$	$\mathcal{N}(708.29, 424.97)$	$\mathcal{N}(2.04, 1.22)$	$\mathcal{N}(1666.22, 999.73)$	$\mathcal{N}(154.28, 92.57)$	$\mathcal{N}(20.76, 12.46)$
b_{EtOH}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{LA}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$
b_{Ac}	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$	$\mathcal{N}(0.5, 0.3)$

Table B.3: Summary of $\widehat{R}$ among successful fittings across 31 model iterations (MI). Maximum $\widehat{R}$ in the referred MI is reported. Column "MI( )" indicates combination of mechanisms deployed in MI.

MI( )	ghhp1	dowb1	ghhp2	ghhp3	ghhp4	ghhp5	brwb1	brwb2	ecpt1	ecpt2	ecwb1	ecwb2	brpb1	brst1	brwb3	brwb4	brst2	brwb7	mywb3	hnwb1	ecpb1	niwb1	niwb2
0	1.00	-	-	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	-	1.00
1	1.00	-	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.01	1.01	-	-	1.00	1.00	-	-	-	-	-	1.00
2	1.00	-	-	1.00	1.00	-	1.00	1.01	1.00	1.00	-	1.01	1.00	1.00	-	-	1.00	1.00	1.00	-	-	-	1.00
3	1.00	-	-	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00
4	1.00	-	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	-	1.00	1.01	-	1.00	1.00	-	-	1.00
5	-	-	-	-	-	-	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2	1.00	-	1.00	1.00	1.00	-	1.00	1.00	1.00	-	1.00	-	1.00	-	-	-	-	-	-	-	-	-	-
1,3	1.00	-	-	-	-	1.01	1.00	1.00	1.00	1.00	1.00	1.02	1.00	-	-	1.00	1.00	-	-	-	-	-	-
1,4	-	-	-	-	-	1.00	1.00	1.00	1.00	-	1.01	-	1.00	-	-	-	1.00	-	1.00	-	-	-	1.01
1,5	-	-	-	-	-	-	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.00	1.00	-	-	-	-	1.00	-	-	1.00	1.00
2,4	1.00	-	1.00	1.00	1.00	-	-	1.00	-	1.00	1.00	-	1.00	-	-	-	-	-	1.00	-	-	-	1.00
2,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4	1.00	-	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00	-	1.00	1.01	-	-	-	-	1.00	1.00	1.00	-	1.00
3,5	-	-	-	-	-	-	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3	1.00	-	1.00	1.00	-	-	1.00	1.00	-	-	1.00	-	1.00	-	-	-	-	-	-	-	-	-	-
1,2,4	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	-	1.00	-	-	-	-	-	-	-	-	-	-
1,2,5	-	-	-	-	-	-	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4	-	-	-	-	1.01	1.00	1.00	1.00	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-	-	-	-	-	-
1,3,5	-	-	-	-	-	-	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,4,5	-	-	-	-	-	-	-	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4	1.00	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	-	1.00	-	-	-	1.00	-	1.00	-	-	-	1.00
2,3,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4	1.00	-	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	-	1.00	-	-	-	-	-	-	-	-	-	-
1,2,3,5	-	-	-	-	-	-	1.00	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table B.4: Summary of bulk effective sample size (Bulk-ESS) among successful fittings across 31 model iterations (MI). Minimum bulk-ESS in the referred MI is reported. Column "MI( )" indicates combination of mechanisms deployed in MI.

MI( )	ghhp1	dowb1	ghhp2	ghhp3	ghhp4	ghhp5	brwb1	brwb2	ecpt1	ecpt2	ecwb1	ecwb2	brpb1	brst1	brwb3	brwb4	brst2	brwb7	mywb3	hnwb1	ecpb1	niwb1	niwb2
0	3385	-	-	-	2515	1159	3241	2975	3399	2475	3297	775	-	1908	1174	654	596	1415	1501	442	253	-	2548
1	636	-	-	-	407	992	2796	2006	-	848	1160	152	1311	-	-	721	328	-	-	-	-	-	2340
2	2140	-	-	2866	2269	-	1468	509	1297	1247	-	505	1734	1490	-	-	477	1026	1192	-	-	-	1443
3	3009	-	-	-	2429	1025	2594	2878	3087	2859	2794	1639	3040	2078	997	974	504	-	1785	498	1044	726	2260
4	2816	-	-	-	2382	837	3022	3203	-	2791	2966	1012	1263	2119	-	1330	420	-	1399	890	-	-	2437
5	-	-	-	-	-	-	2557	2745	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2	1819	-	2791	728	1370	-	1861	592	940	-	891	-	704	-	-	-	-	-	-	-	-	-	-
1,3	1200	-	-	-	-	112	2569	2296	3396	719	522	245	1062	-	-	1315	352	-	-	-	-	-	-
1,4	-	-	-	-	-	485	1610	2924	3108	-	1235	-	942	-	-	-	3023	-	1774	-	-	-	252
1,5	-	-	-	-	-	-	3349	2294	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3	2723	1654	1748	2638	1940	1176	1306	737	1307	921	509	404	2106	1297	-	-	-	-	1172	-	-	1024	768
2,4	1636	-	2377	2255	3000	-	-	528	-	1275	1015	-	1770	-	-	-	-	-	1416	-	-	-	2273
2,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4	1377	-	-	-	2018	1060	2773	2949	-	2534	3312	-	1747	359	-	-	-	-	1520	1457	1984	-	2377
3,5	-	-	-	-	-	-	2131	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3	1167	-	2427	622	-	-	1395	494	-	-	857	-	571	-	-	-	-	-	-	-	-	-	-
1,2,4	2606	-	2571	1046	1566	1053	1795	775	3073	-	724	-	631	-	-	-	-	-	-	-	-	-	-
1,2,5	-	-	-	-	-	-	1403	624	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4	-	-	-	-	316	197	1674	1945	3131	-	708	-	542	-	824	-	2645	-	-	-	-	-	-
1,3,5	-	-	-	-	-	-	2609	2288	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,4,5	-	-	-	-	-	-	-	2350	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4	2360	-	-	2384	4336	1203	1619	-	1681	800	1158	-	1723	-	-	-	600	-	1114	-	-	-	910
2,3,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4	1624	-	2060	898	-	560	1722	571	2483	-	1097	-	633	-	-	-	-	-	-	-	-	-	-
1,2,3,5	-	-	-	-	-	-	1779	472	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table B.5: Summary of tail effective sample size (Tail-ESS) among successful fittings across 31 model iterations (MI). Minimum tail-ESS in the referred MI is reported. Column "MI( )" indicates combination of mechanisms deployed in MI.

MI( )	ghhp1	dowb1	ghhp2	ghhp3	ghhp4	ghhp5	brwb1	brwb2	ecpt1	ecpt2	ecwb1	ecwb2	brpb1	brst1	brwb3	brwb4	brst2	brwb7	mywb3	hmwb1	ecpb1	niwb1	niwb2
0	1851	-	-	-	1361	553	2540	2206	1574	2000	2208	1746	-	1185	1746	288	1386	1444	2109	1738	703	-	2290
1	756	-	-	-	103	554	2336	1954	-	313	1017	1130	1423	-	-	199	919	-	-	-	-	-	1536
2	2091	-	-	1334	1695	-	1876	1703	686	1087	-	1154	1312	1217	-	-	1407	1450	1120	-	-	-	1836
3	1687	-	-	-	1532	336	1505	2210	2203	1597	1475	1001	1885	702	1752	710	470	-	1570	1492	1261	828	1944
4	1174	-	-	-	1882	661	1709	1777	-	1529	2013	365	1325	660	-	568	281	-	1775	565	-	-	1930
5	-	-	-	-	-	-	2203	2132	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2	1042	-	1708	810	775	-	1921	1146	1185	-	522	-	1069	-	-	-	-	-	-	-	-	-	-
1,3	853	-	-	-	-	167	1724	2108	1691	214	334	1153	1170	-	-	589	244	-	-	-	-	-	-
1,4	-	-	-	-	-	529	1613	2230	1384	-	1063	-	487	-	-	-	1698	-	1655	-	-	-	234
1,5	-	-	-	-	-	-	2543	1712	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3	2158	1421	1287	1596	800	464	1057	1642	1890	1538	1408	820	1188	865	-	-	-	-	644	-	-	897	1242
2,4	1692	-	1682	1457	1484	-	-	1679	-	1315	1046	-	1353	-	-	-	-	-	1158	-	-	-	2297
2,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4	557	-	-	-	1192	662	1980	1927	-	1257	1709	-	1673	132	-	-	-	-	1501	866	1699	-	1712
3,5	-	-	-	-	-	-	1385	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3	634	-	1706	623	-	-	1613	1394	-	-	1027	-	762	-	-	-	-	-	-	-	-	-	-
1,2,4	1008	-	1295	689	1177	828	1299	1606	1773	-	909	-	777	-	-	-	-	-	-	-	-	-	-
1,2,5	-	-	-	-	-	-	1604	1122	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4	-	-	-	-	1087	414	1776	1182	1836	-	547	-	527	-	687	-	1183	-	-	-	-	-	-
1,3,5	-	-	-	-	-	-	1707	2107	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,4,5	-	-	-	-	-	-	-	1984	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4	2194	-	-	1768	2518	631	1578	-	473	1116	1656	-	1331	-	-	-	1049	-	467	-	-	-	1042
2,3,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4	782	-	1317	1314	-	715	1127	997	1944	-	1084	-	1004	-	-	-	-	-	-	-	-	-	-
1,2,3,5	-	-	-	-	-	-	1703	1198	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,3,4,5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table B.6: Parameter estimates of model iteration (MI) involving mechanisms M1 and M2, MI(2,3) calibrated over different datasets. Means and standard deviations (within parenthesis) of the posterior parameters are shown. All reported values in literature have same units as depicted in table 2.MM with the exception of substrate saturation constants (mg(substrate) mL^{-1}). Green coloured cells indicate parameter estimates that lie within reported ranges. Orange coloured cells indicate parameter estimates that are far from reported range by less than the estimate either divided or multiplied by 10. Red coloured cells indicate parameter estimates that are far from reported range by more than the estimate either divided or multiplied by 10. Parameters of MI(2,3) fitted to datasets *ghhp2*, *ghhp3*, *ghhp4* and *ghhp5* are not shown due to their lack of LA time series.

Parameter	ghhp1	dowb1	brwb1	brwb2	ecpt1	ecpt2	ecwb1	Reported values	Reference
$\mu_{max}^{Y_{Glc}}$	0.406 (0.124)	0.133 (0.11)	0.094 (0.052)	0.381 (0.167)	0.162 (0.103)	0.464 (0.245)	0.141 (0.101)	0.0781 - 0.53	[77–84]
$\mu_{max}^{Y_{Fru}}$	0.233 (0.109)	0.145 (0.118)	0.039 (0.027)	0.551 (0.191)	0.167 (0.101)	0.461 (0.224)	0.141 (0.108)	0.01 - 0.83	[78, 83, 91, 92, 137, 138, 149]
$\mu_{max}^{LAB_{Glc}}$	0.137 (0.079)	0.382 (0.178)	0.122 (0.109)	0.382 (0.169)	0.279 (0.124)	0.561 (0.287)	0.203 (0.144)	0.0072 - 1.41	[85–90]
$\mu_{max}^{LAB_{Fru}}$	0.193 (0.077)	0.272 (0.177)	0.283 (0.103)	0.328 (0.192)	0.265 (0.122)	0.341 (0.269)	0.216 (0.151)	0.0302 - 0.85	[150, 151]
$\mu_{max}^{AAB_{EtOH}}$	0.466 (0.121)	0.424 (0.154)	0.163 (0.051)	0.181 (0.054)	0.309 (0.115)	0.458 (0.145)	0.223 (0.106)	0.0106 - 0.25	[93–97]
$\mu_{max}^{AAB_{LA}}$	0.016 (0.022)	0.05 (0.071)	0.022 (0.015)	0.02 (0.015)	0.062 (0.03)	0.062 (0.047)	0.038 (0.041)		
K_{Glc}^{Y}	38.569 (12.739)	35.079 (15.842)	35.494 (15.024)	28.897 (11.247)	29.164 (13.642)	24.686 (11.317)	23.998 (11.648)	9.73×10^{-4} - 0.5	[82, 84]
K_{Fru}^{Y}	39.902 (15.329)	54.891 (24.524)	30.313 (13.692)	39.954 (17.818)	31.309 (15.428)	23.367 (10.27)	31.981 (14.899)	5.82 - 161.45	[92, 137, 138]
K_{Glc}^{LAB}	34.266 (13.624)	34.838 (15.049)	35.326 (14.934)	23.098 (11.403)	33.794 (13.191)	23.247 (11.143)	27.604 (11.475)	0.790 - 178.0	[88–90]
K_{Fru}^{LAB}	38.586 (15.226)	56.625 (23.97)	31.16 (13.268)	39.39 (18.242)	37.698 (14.198)	23.568 (10.569)	36.596 (14.592)		
K_{EtOH}^{AAB}	15.995 (5.483)	6.16 (2.507)	3.895 (1.652)	4.193 (1.964)	8.448 (3.276)	5.944 (2.46)	11.8 (4.771)		
K_{LA}^{AAB}	2568.5 (1232.144)	30.1 (13.291)	316.919 (154.179)	81.322 (41.956)	194.145 (92.66)	45.999 (22.447)	1063.022 (498.821)		
k_Y	0.0455 (0.0074)	0.0559 (0.0241)	0.0891 (0.0288)	0.0861 (0.0246)	0.0012 (0.0014)	0.0315 (0.019)	0.0111 (0.0081)		
k_{LAB}	0.007 (0.0016)	0.1558 (0.1003)	0.0553 (0.0151)	0.0452 (0.0212)	0.0572 (0.0174)	0.4004 (0.1669)	0.0561 (0.0219)		
k_{AAB}	0.0096 (0.0023)	0.0011 (4×10^{-4})	4×10^{-4} (2×10^{-4})	4×10^{-4} (2×10^{-4})	0.0158 (0.0064)	0.0014 (7×10^{-4})	0.0067 (0.0036)		
$Y_{Glc\|Y}$	33.305 (10.489)	32.153 (15.894)	266.721 (59.451)	136.564 (38.749)	24.931 (11.133)	31.857 (13.823)	34.711 (13.955)	1.56 - 66.67	[78–82, 100, 101]
$Y_{Glc\|LAB}$	46.411 (22.205)	401.212 (150.668)	75.141 (38.771)	13.355 (15.251)	68.313 (18.895)	108.954 (38.638)	46.171 (17.156)	4.50 - 200	[90, 99]
$Y_{Fru\|Y}$	39.129 (13.619)	52.533 (24.239)	194.391 (62.42)	249.378 (65.934)	27.339 (11.936)	29.489 (11.442)	44.019 (18.671)	43.48 - 200	[78, 91]
$Y_{Fru\|LAB}$	51.41 (20.726)	570.03 (245.658)	53.727 (27.002)	41.31 (28.812)	75.755 (21.129)	88.447 (36.51)	59.887 (22.015)		
$Y_{EtOH\|Y}^{Glc}$	4.236 (2.778)	4.477 (2.656)	7.823 (4.279)	7.798 (4.539)	5.581 (2.913)	4.34 (2.776)	8.798 (5.14)	1.39 - 21.49	[81–83, 98]
$Y_{EtOH\|Y}^{Fru}$	6.12 (4.207)	4.422 (2.613)	10.794 (6.626)	6.091 (4.129)	5.436 (2.899)	3.434 (2.495)	8.72 (5.209)	5.679 - 5.896	[83]
$Y_{EtOH\|LAB}^{Glc}$	14.762 (9.467)	62.736 (28.211)	6.547 (4.269)	2.258 (2.457)	6.216 (4.03)	10.483 (6.867)	9.482 (6.375)		
$Y_{EtOH\|LAB}^{Fru}$	9.598 (6.119)	58.662 (28.742)	2.446 (2.025)	4.247 (2.964)	6.158 (3.945)	11.696 (7.572)	9.127 (6.483)		
$Y_{EtOH\|AAB}$	1769.652 (915.5)	126.408 (49.954)	450.646 (163.983)	181.209 (57.503)	332.519 (151.239)	371.945 (115.609)	4330.497 (1646.577)	8.06 - 166.67	[94, 96, 102]
$Y_{LA\|LAB}^{Glc}$	9.021 (3.772)	9.765 (5.536)	5.966 (3.558)	3.114 (1.323)	4.167 (2.089)	3.065 (1.291)	3.282 (1.717)		
$Y_{LA\|LAB}^{Fru}$	9.673 (2.733)	10.845 (5.819)	2.994 (1.81)	3.151 (1.527)	4.045 (2.073)	2.967 (1.378)	3.189 (1.666)		
$Y_{LA\|AAB}$	2012.646 (1187.814)	23.54 (13.299)	297.512 (142.921)	63.715 (38.857)	203.33 (81.7)	41.626 (21.469)	913.388 (483.211)	7.51 - 8.34	[96]
$Y_{Ac\|LAB}^{Glc}$	3.277 (2.365)	43.254 (43.771)	9.423 (7.501)	5.236 (6.142)	0.598 (0.527)	9.834 (10.42)	2.382 (2.464)		
$Y_{Ac\|LAB}^{Fru}$	1.809 (1.395)	49.913 (49.765)	2.144 (2.64)	10.774 (7.852)	0.553 (0.499)	13.229 (12.641)	2.356 (2.394)		
$Y_{Ac\|AAB}^{EtOH}$	68.368 (80.311)	29.516 (37.316)	544.767 (252.44)	292.764 (124.828)	19.977 (18.986)	335.779 (179.607)	1271.525 (697.321)		
$Y_{Ac\|AAB}^{LA}$	1170.869 (828.312)	170.617 (128.778)	677.034 (351.821)	376.785 (181.596)	61.421 (45.193)	406.623 (229.638)	1418.934 (760.394)		
$Y_{Ac\|Y}^{Glc}$	1.038 (0.741)	5.706 (5.528)	4.663 (4.576)	3.422 (3.083)	1.106 (0.911)	5.466 (4.86)	2.688 (2.328)		
$Y_{Ac\|Y}^{Fru}$	1.635 (1.13)	5.251 (5.214)	11.526 (11.225)	3.605 (3.209)	0.981 (0.876)	3.594 (4.094)	2.839 (2.355)		
σ	0.134 (0.009)	0.267 (0.039)	0.165 (0.014)	0.176 (0.015)	0.197 (0.015)	0.25 (0.02)	0.21 (0.016)		

Table B.6 **(cont.)**: Parameter estimates of model iteration (MI) involving mechanisms M1 and M2, MI(2,3) calibrated over different datasets. Means and standard deviations (within parenthesis) of the posterior parameters are shown. All reported values in literature have same units as depicted in table 2.MM with the exception of substrate saturation constants (mg(substrate) mL^{-1}). Green coloured cells indicate parameter estimates that lie within reported ranges. Orange coloured cells indicate parameter estimates that are far from reported range by less than the estimate either divided or multiplied by 10. Red coloured cells indicate parameter estimates that are far from reported range by more than the estimate either divided or multiplied by 10. Parameters of MI(2,3) fitted to datasets *ghhp2*, *ghhp3*, *ghhp4* and *ghhp5* are not shown due to their lack of LA time series.

Parameter	ecwb2	brpb1	brst1	mywb3	niwb1	niwb2	Reported values	Reference
$\mu_{max}^{Y_{Glc}}$	0.533 (0.308)	0.419 (0.243)	0.441 (0.244)	0.05 (0.055)	0.068 (0.145)	0.113 (0.063)	0.0781 - 0.53	[77–84]
$\mu_{max}^{Y_{Fru}}$	0.205 (0.224)	0.428 (0.238)	0.433 (0.248)	0.339 (0.126)	0.052 (0.11)	0.06 (0.059)	0.01 - 0.83	[78, 83, 91, 92, 137, 138, 149]
$\mu_{max}^{LAB_{Glc}}$	0.403 (0.203)	0.494 (0.225)	0.49 (0.237)	0.33 (0.107)	0.324 (0.176)	0.128 (0.094)	0.0072 - 1.41	[85–90]
$\mu_{max}^{LAB_{Fru}}$	0.249 (0.163)	0.487 (0.226)	0.489 (0.242)	0.129 (0.111)	0.311 (0.165)	0.154 (0.093)	0.0302 - 0.85	[150, 151]
$\mu_{max}^{AAB_{EtOH}}$	0.257 (0.133)	0.313 (0.201)	0.076 (0.113)	0.236 (0.104)	0.233 (0.161)	0.157 (0.055)	0.0106 - 0.25	[93–97]
$\mu_{max}^{AAB_{LA}}$	0.057 (0.042)	0.159 (0.101)	0.393 (0.109)	0.03 (0.035)	0.063 (0.076)	0.022 (0.015)		
K_{Glc}^{Y}	19.578 (9.655)	62.116 (26.394)	80.799 (34.188)	23.619 (11.111)	14.641 (6.638)	13.051 (5.875)	9.73×10^{-4} - 0.5	[82, 84]
K_{Fru}^{Y}	20.005 (8.892)	58.289 (24.438)	74.396 (31.561)	34.473 (14.087)	18.624 (8.353)	15.582 (7.053)	5.82 - 161.45	[92, 137, 138]
K_{Glc}^{LAB}	22.453 (9.579)	60.659 (25.981)	79.018 (34.182)	19.758 (10.912)	15.21 (6.583)	13.484 (5.807)	0.790 - 178.0	[88–90]
K_{Fru}^{LAB}	20.887 (9.011)	56.062 (24.172)	72.225 (32.106)	31.599 (14.311)	19.585 (8.357)	16.055 (7.059)		
K_{EtOH}^{AAB}	7.979 (3.375)	41.536 (17.775)	47.863 (21.327)	11.254 (4.676)	10.808 (4.646)	11.616 (5.046)		
K_{LA}^{AAB}	516.058 (253.489)	29557.579 (14464.123)	88156.259 (30255.39)	509.086 (254.786)	645.104 (323.747)	150.711 (77.189)		
k_Y	0.0292 (0.0262)	0.0084 (0.004)	0.0054 (0.0031)	0.0052 (0.0017)	0.0235 (0.0166)	0.0075 (0.0049)		
k_{LAB}	0.019 (0.0083)	0.0083 (0.0075)	0.0019 (0.0011)	0.0095 (0.0035)	0.1128 (0.0314)	0.0075 (0.0061)		
k_{AAB}	7×10^{-4} (3×10^{-4})	2×10^{-4} (1×10^{-4})	0.0028 (5×10^{-4})	0.0028 (8×10^{-4})	0.0014 (0.0013)	0.0018 (0.0011)		
$Y_{Glc\|Y}$	52.08 (19.939)	100.101 (49.917)	83.262 (48.886)	86.243 (41.832)	14.36 (7.361)	10.084 (4.332)	1.56 - 66.67	[78–82, 100, 101]
$Y_{Glc\|LAB}$	46.45 (20.949)	185.227 (70.504)	445.826 (324.583)	23.871 (5.719)	837.761 (367.718)	67.91 (27.827)	4.50 - 200	[90, 99]
$Y_{Fru\|Y}$	34.224 (15.277)	92.125 (45.11)	78.096 (44.916)	184.468 (39.755)	18.344 (9.198)	11.538 (5.614)	43.48 - 200	[78, 91]
$Y_{Fru\|LAB}$	35.514 (13.892)	168.776 (66.785)	420.162 (305.457)	22.443 (10.964)	1090.949 (461.221)	83.376 (29.166)		
$Y_{EtOH\|Y}^{Glc}$	5.967 (4.32)	58.993 (31.485)	54.979 (28.79)	31.918 (17.444)	9.103 (5.12)	5.625 (3.391)	1.39 - 21.49	[81–83, 98]
$Y_{EtOH\|Y}^{Fru}$	7.332 (4.709)	59.377 (30.913)	55.533 (28.766)	32.111 (11.671)	9.128 (5.101)	6.661 (4.043)	5.679 - 5.896	[83]
$Y_{EtOH\|LAB}^{Glc}$	6.901 (4.98)	84.956 (47.129)	375.841 (182.627)	2.971 (2.127)	471.956 (260.117)	37.133 (22.578)		
$Y_{EtOH\|LAB}^{Fru}$	5.697 (3.694)	85.589 (47.144)	373.244 (183.374)	6.046 (3.483)	470.021 (261.764)	30.268 (20.045)		
$Y_{EtOH\|AAB}$	1359.87 (563.799)	65259.148 (29128.894)	185454.191 (102105.223)	300.007 (230.429)	2253.568 (956.26)	402.877 (145.085)	8.06 - 166.67	[94, 96, 102]
$Y_{LA\|LAB}^{Glc}$	4.156 (1.908)	53.487 (23.831)	132.892 (52.066)	5.605 (1.335)	212.055 (92.244)	22.434 (9.313)		
$Y_{LA\|LAB}^{Fru}$	4.683 (1.598)	53.988 (23.427)	132.034 (52.079)	4.972 (2.354)	214.081 (92.846)	22.042 (8.087)		
$Y_{LA\|AAB}$	490.432 (245.904)	29932.608 (13944.798)	13414.789 (19209.268)	294.357 (239.17)	614.49 (314.846)	148.154 (71.871)	7.51 - 8.34	[96]
$Y_{Ac\|LAB}^{Glc}$	10.368 (9.221)	76.167 (46.329)	36.943 (32.826)	1.108 (0.926)	284.313 (221.157)	10.935 (9.642)		
$Y_{Ac\|LAB}^{Fru}$	5.178 (6.049)	76.538 (47.826)	39.531 (34.132)	3.291 (1.975)	279.148 (211.528)	7.524 (7.958)		
$Y_{Ac\|AAB}^{EtOH}$	1722.008 (1032.276)	53537.995 (30498.644)	40471.046 (24408.066)	126.499 (101.311)	1377.261 (843.113)	78.41 (47.761)		
$Y_{Ac\|AAB}^{LA}$	2000.661 (1164.352)	51184.995 (30136.054)	10189.867 (16933.627)	325.67 (203.901)	1605.76 (881.987)	149.246 (79.831)		
$Y_{Ac\|Y}^{Glc}$	7.947 (8.111)	54.115 (32.154)	8.596 (6.597)	15.564 (10.007)	6.982 (4.479)	1.111 (1.28)		
$Y_{Ac\|Y}^{Fru}$	9.826 (8.814)	54.611 (31.819)	8.981 (6.679)	7.59 (4.948)	7.082 (4.488)	1.934 (1.767)		
σ	0.253 (0.02)	0.278 (0.023)	0.234 (0.024)	0.161 (0.013)	0.316 (0.07)	0.182 (0.031)		

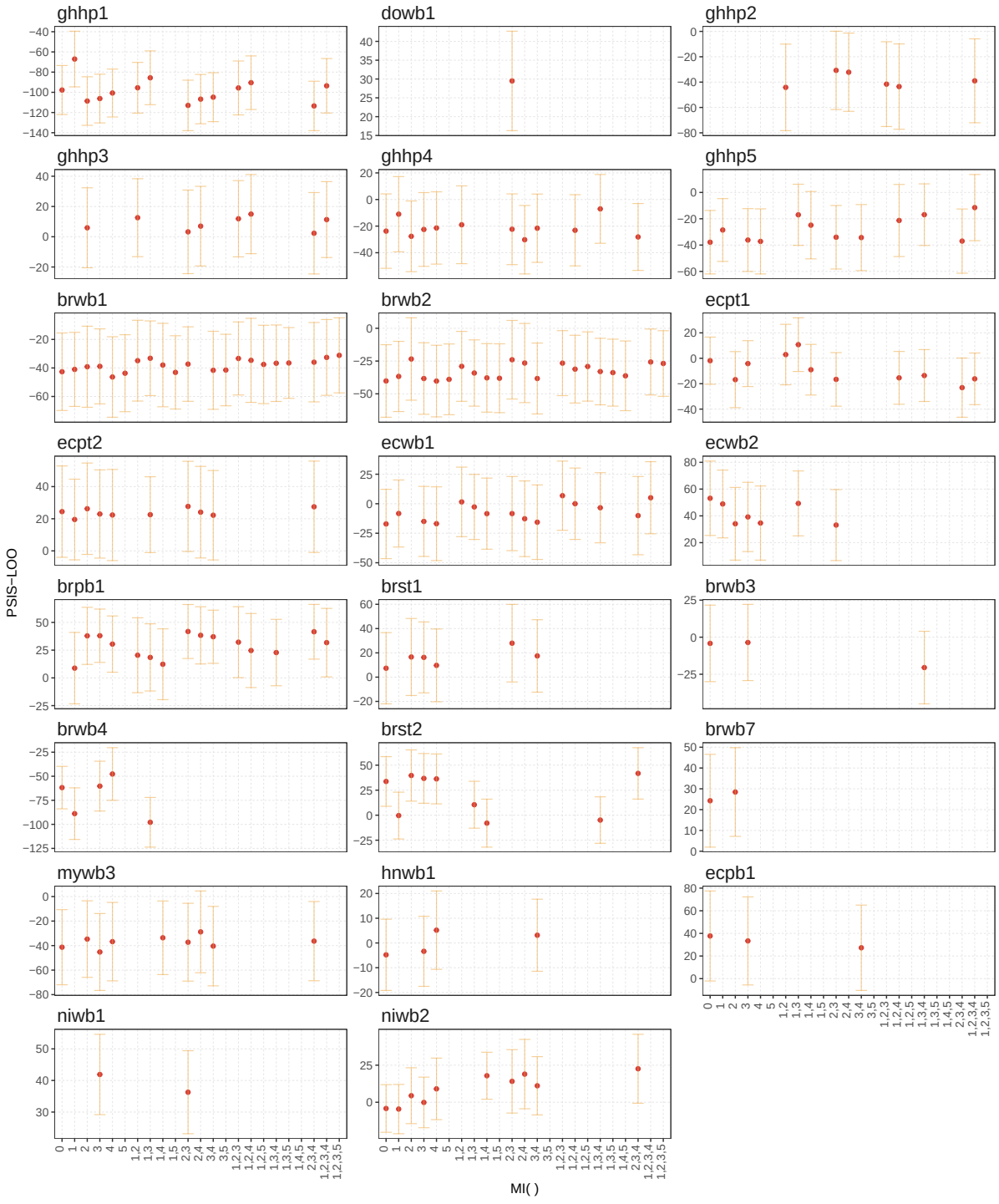


Figure B.1: Pareto-smoothed importance sampling leave-one out cross validation (PSIS-LOO) values for model iterations (MIs) over considered datasets. Error bars correspond to standard error of PSIS-LOO. "MI( )" indicates combination of mechanisms deployed in MI.

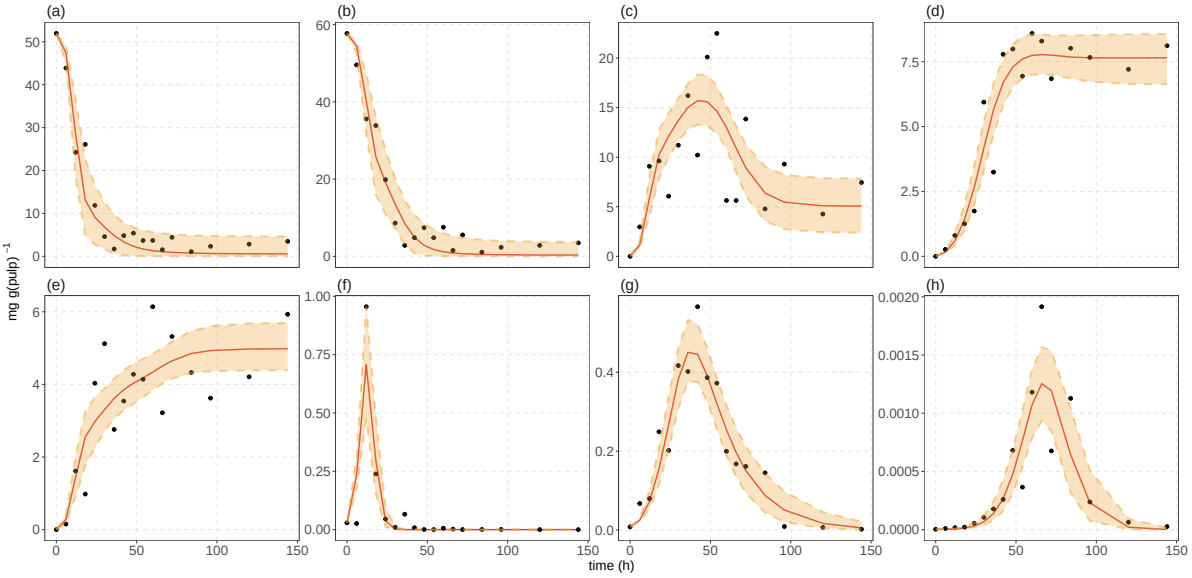


Figure B.2: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ghhp1* reported by Camu *et al.* [12]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

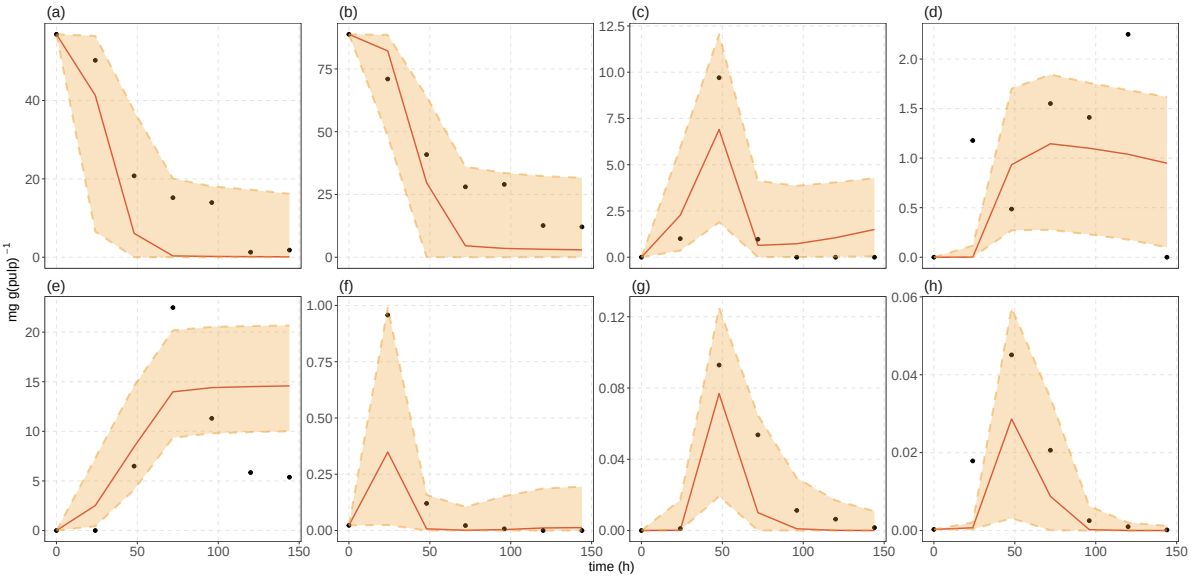


Figure B.3: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *dowb1* reported by Lagunes Gálvez *et al.* [112]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

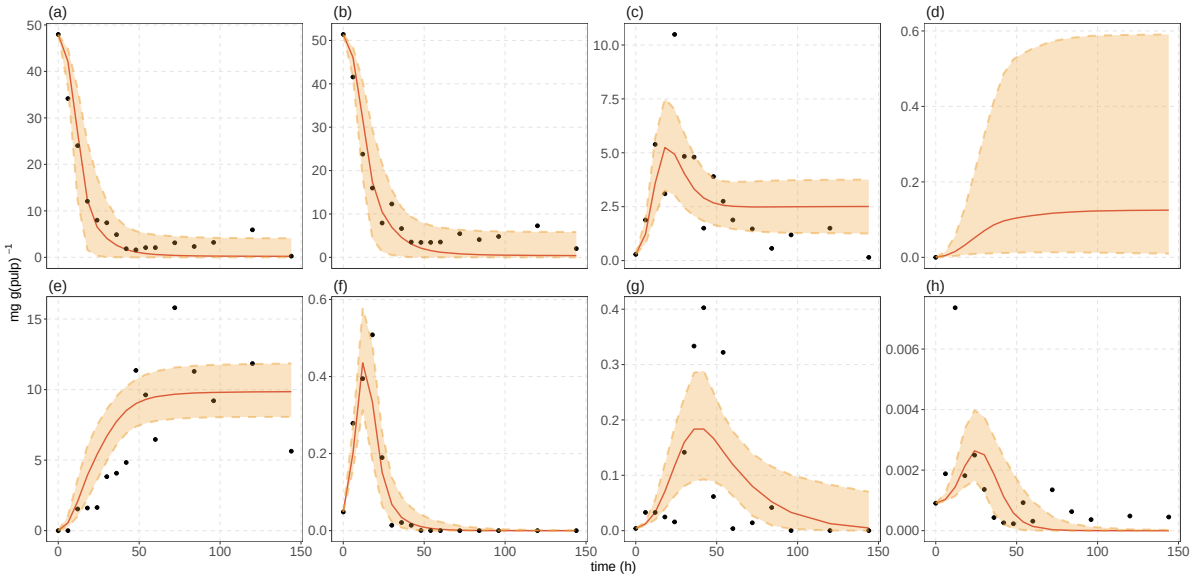


Figure B.4: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ghhp2* reported by Camu *et al.* [17]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid (simulated time series scaled to 1) and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

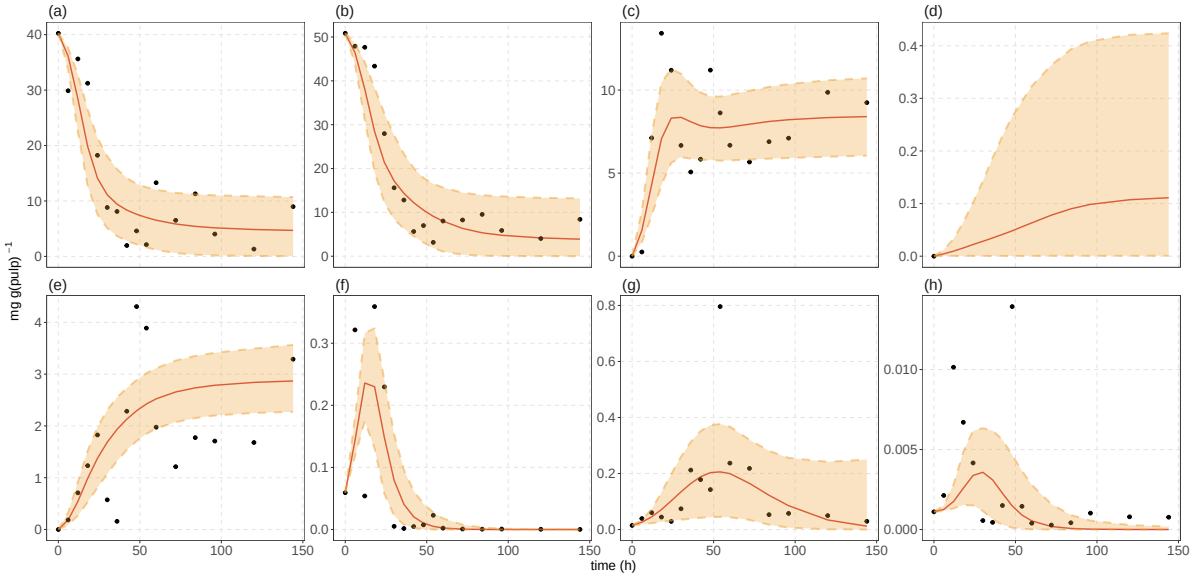


Figure B.5: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ghhp3* reported by Camu *et al.* [17]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid (simulated time series scaled to 1) and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

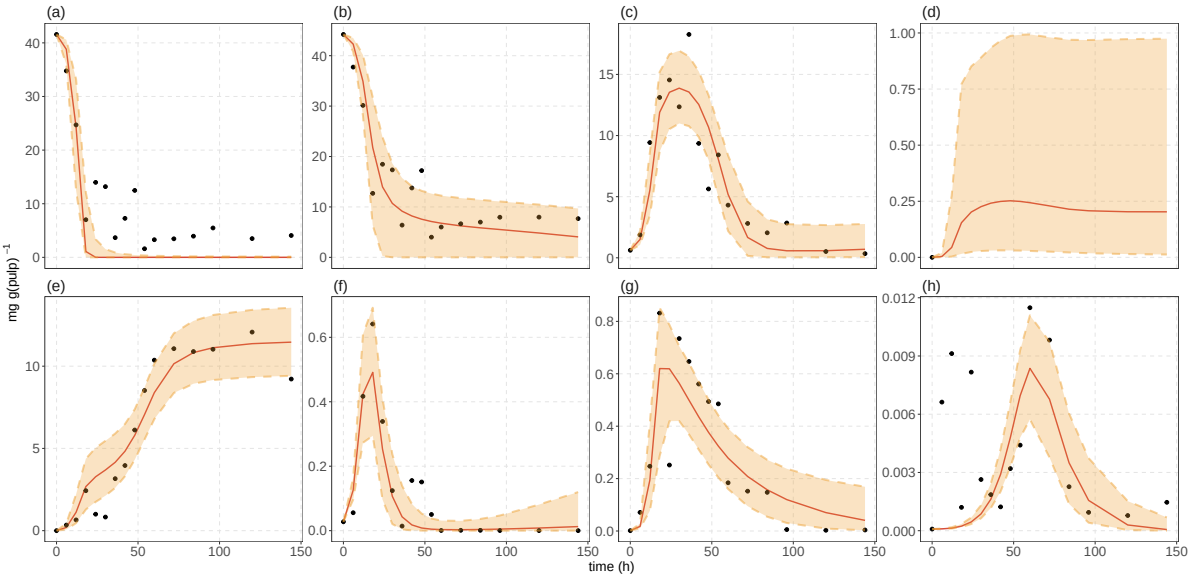


Figure B.6: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ghhp4* reported by Camu *et al.* [17]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid (simulated time series scaled to 1) and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

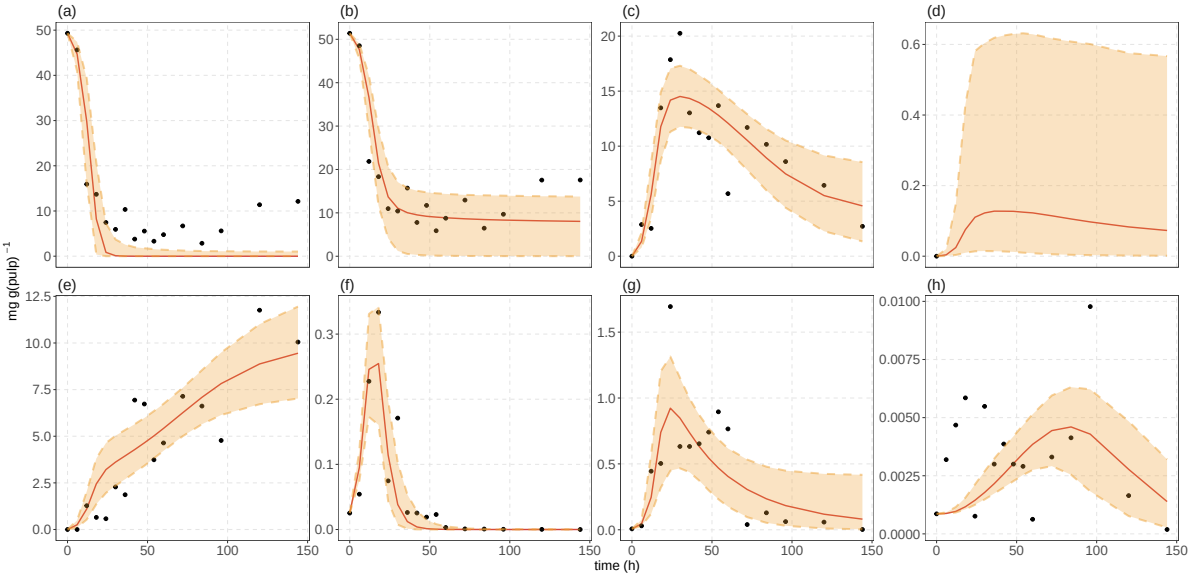


Figure B.7: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ghhp5* reported by Camu *et al.* [17]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid (simulated time series scaled to 1) and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

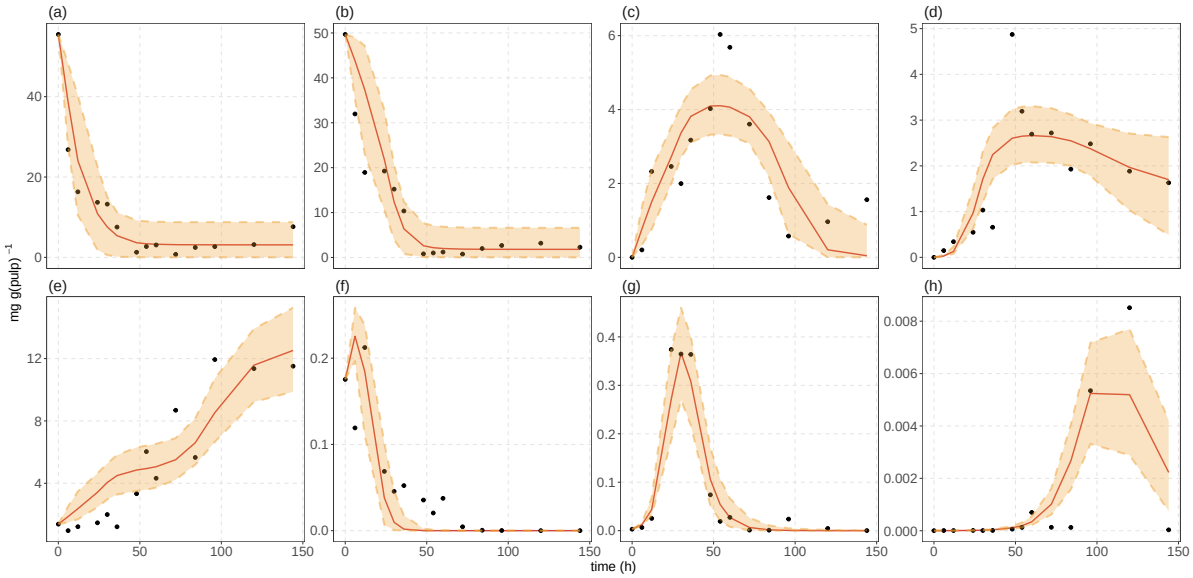


Figure B.8: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *brwb1* reported by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

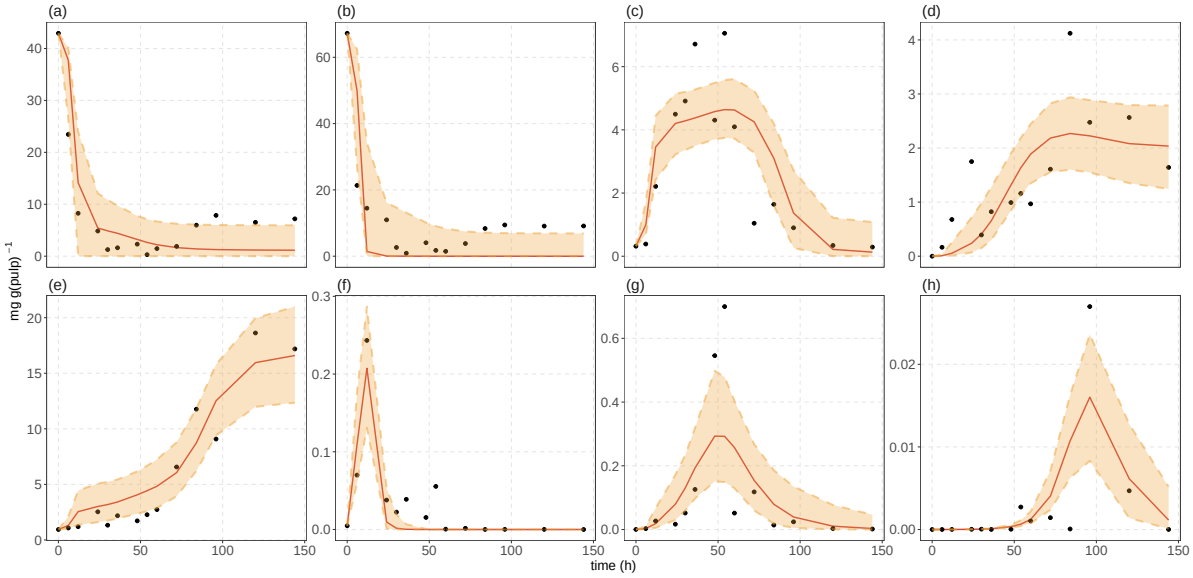


Figure B.9: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *brwb2* reported by Papalexandratou *et al.* [62]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

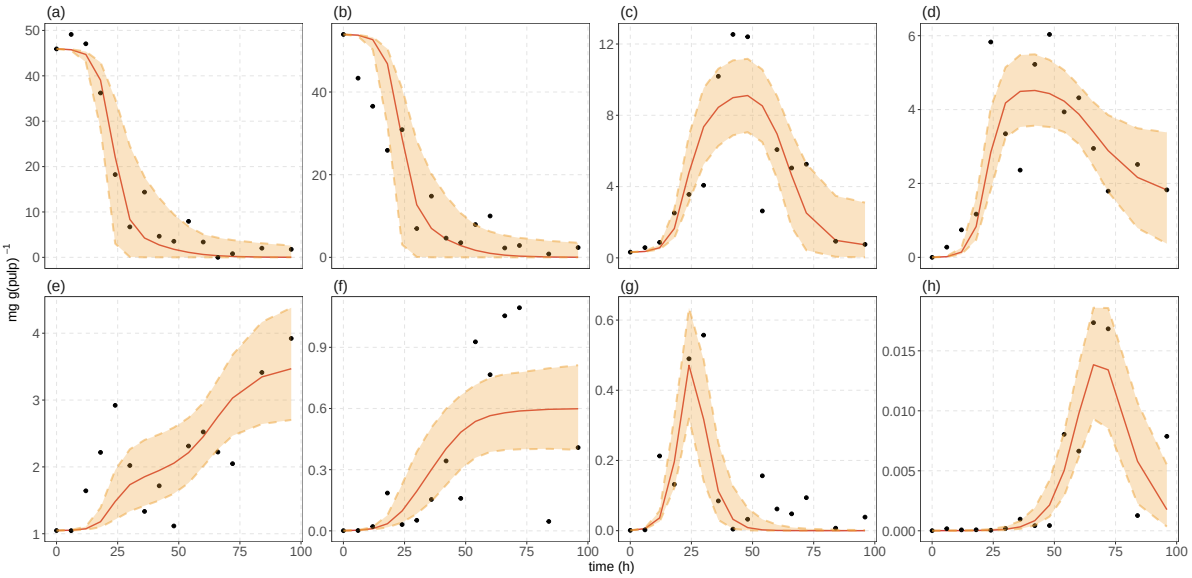


Figure B.10: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ecpt1* reported by Papalexandratou *et al.* [113]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

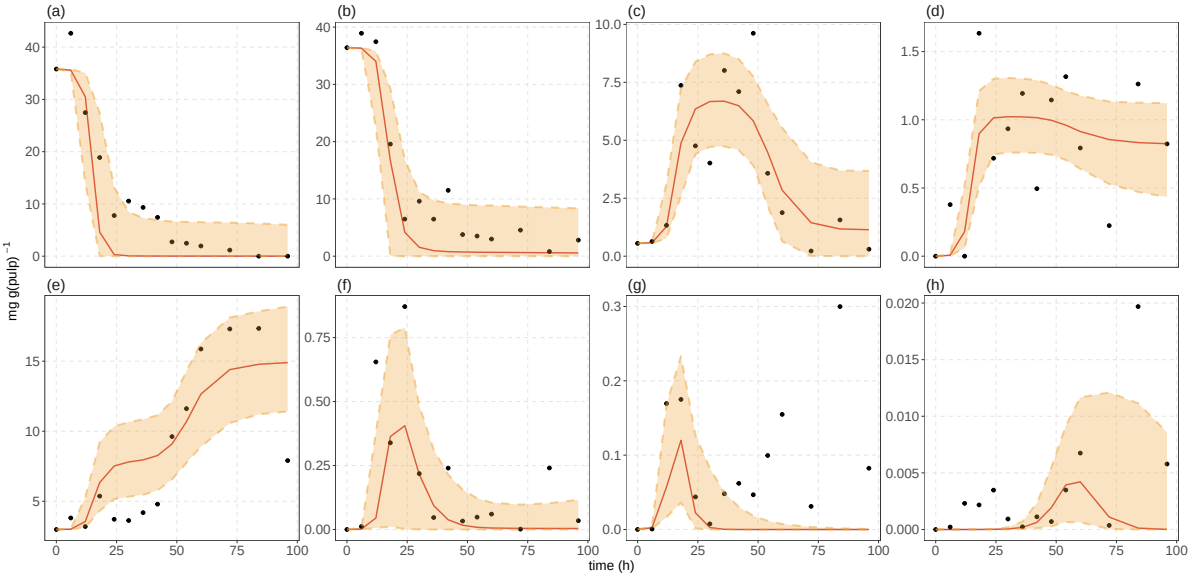


Figure B.11: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ecpt2* reported by Papalexandratou *et al.* [113]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

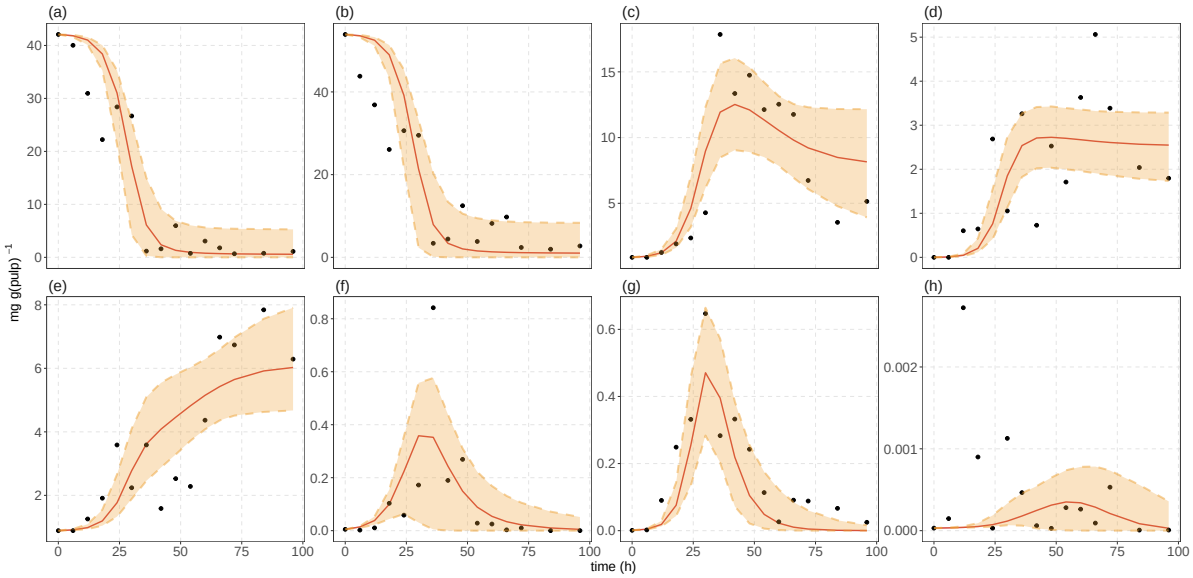


Figure B.12: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ecwb1* reported by Papalexandratou *et al.* [113]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

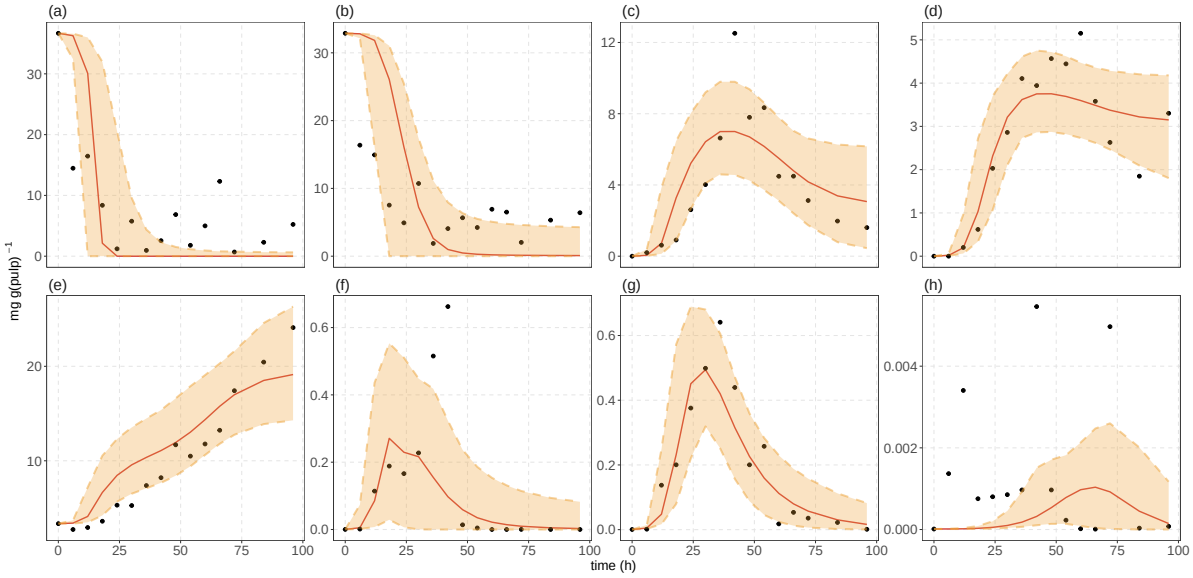


Figure B.13: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *ecwb2* reported by Papalexandratou *et al.* [113]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

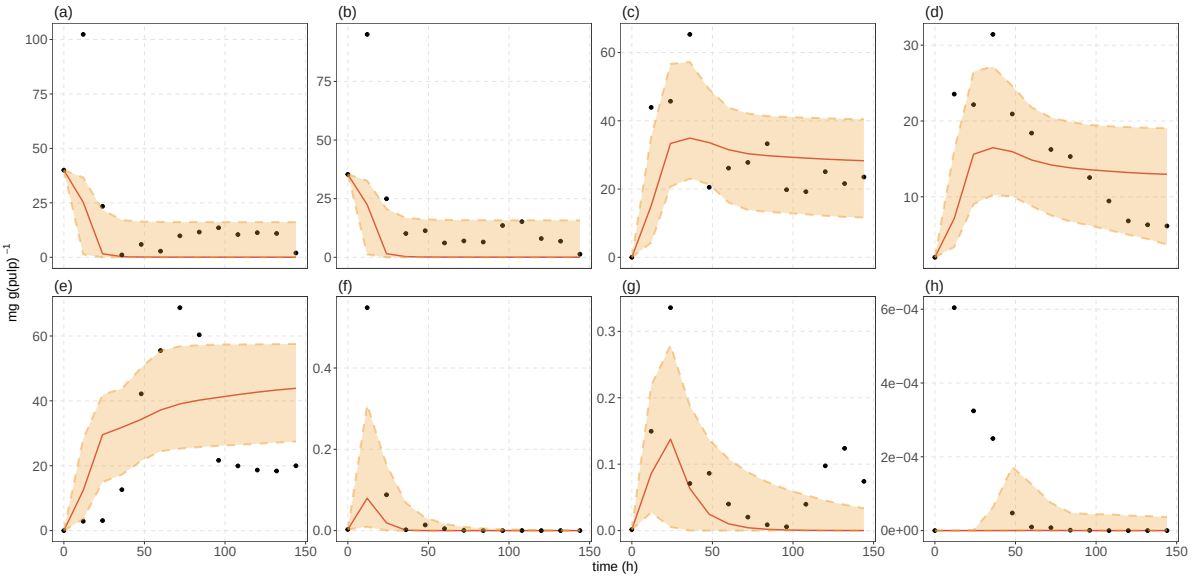


Figure B.14: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *brpb1* reported by Pereira *et al.* [68]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

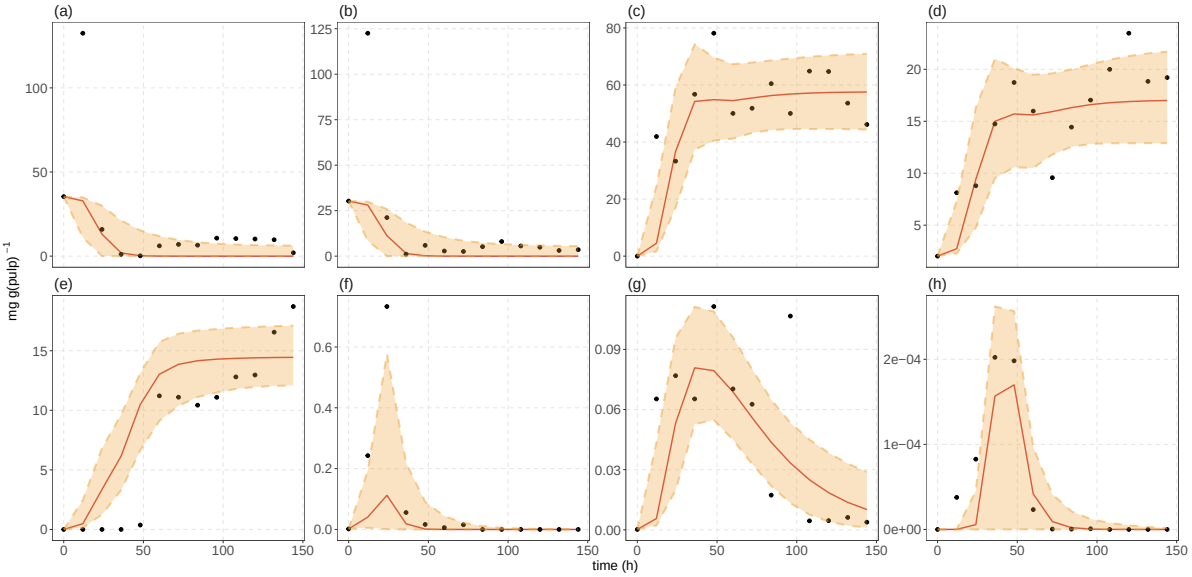


Figure B.15: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *brst1* reported by Pereira *et al.* [68]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

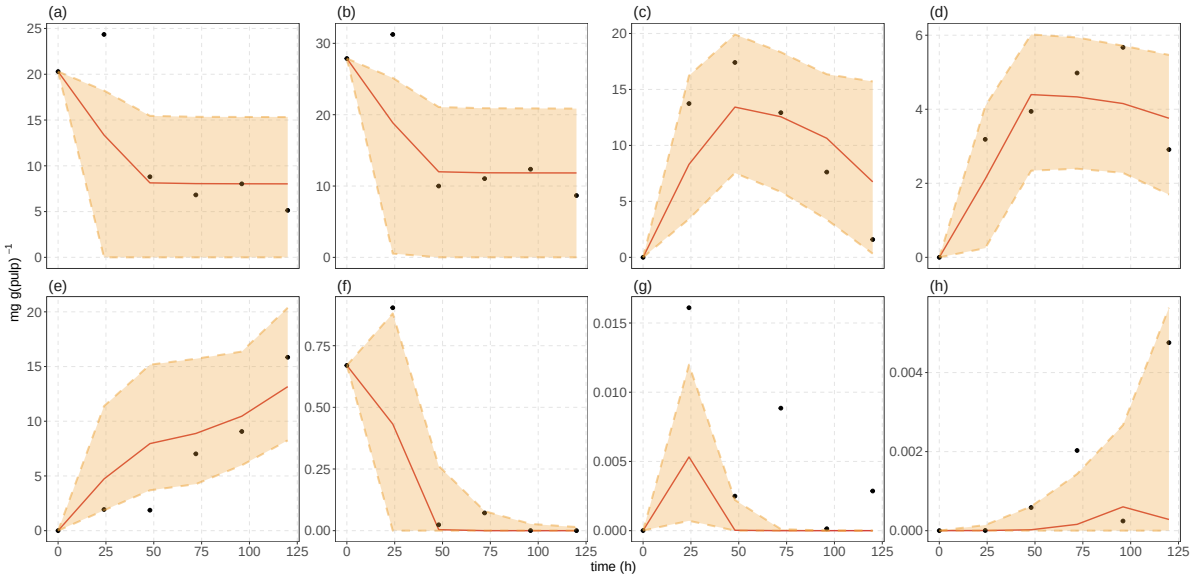


Figure B.16: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *niwb1* reported by Papalexandratou *et al.* [117]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

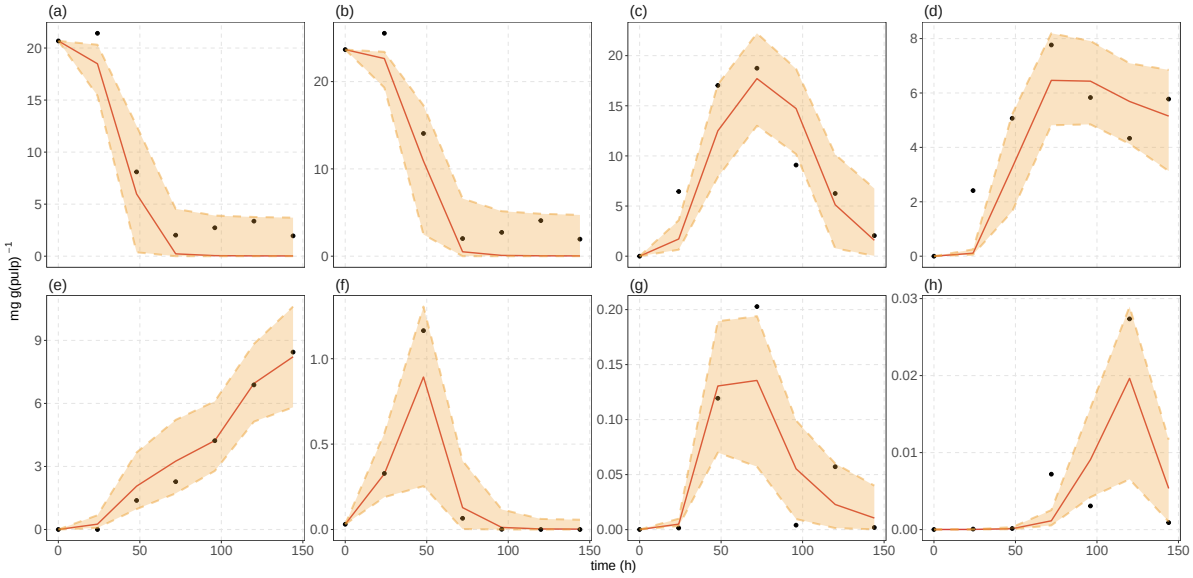


Figure B.17: Posterior predictions of model iteration (MI) corresponding to mechanisms M2 and M3, MI(2,3), fitted to dataset *niwb2* reported by Papalexandratou *et al.* [117]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

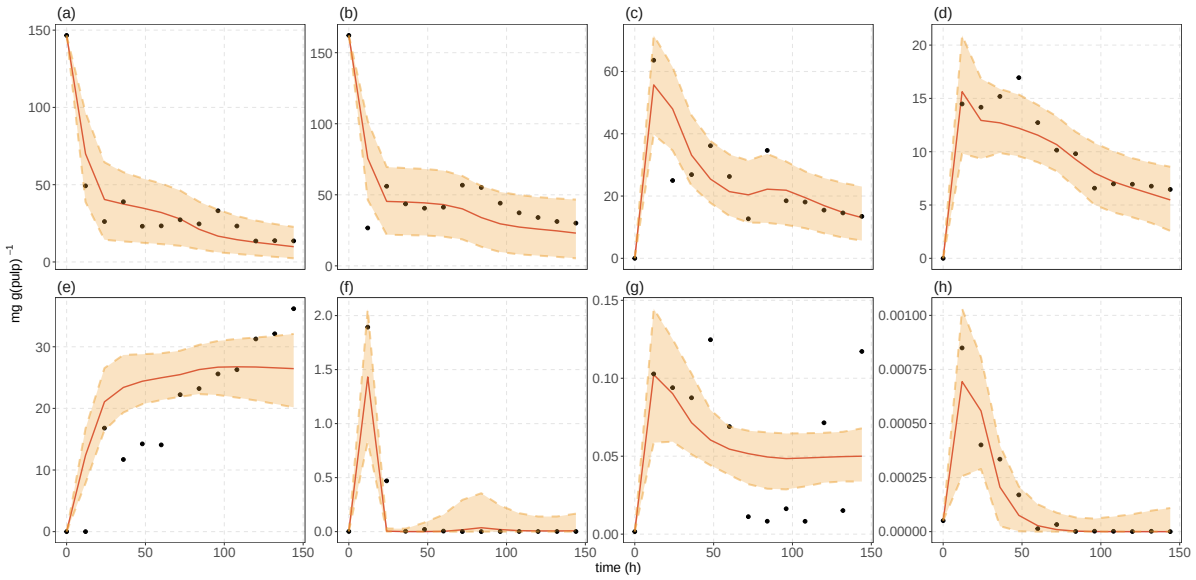


Figure B.18: Posterior predictions of model iteration (MI) corresponding to mechanisms M1, M3 and M4, MI(1,3,4), fitted to dataset *brwb3* reported by Pereira *et al.* [14]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

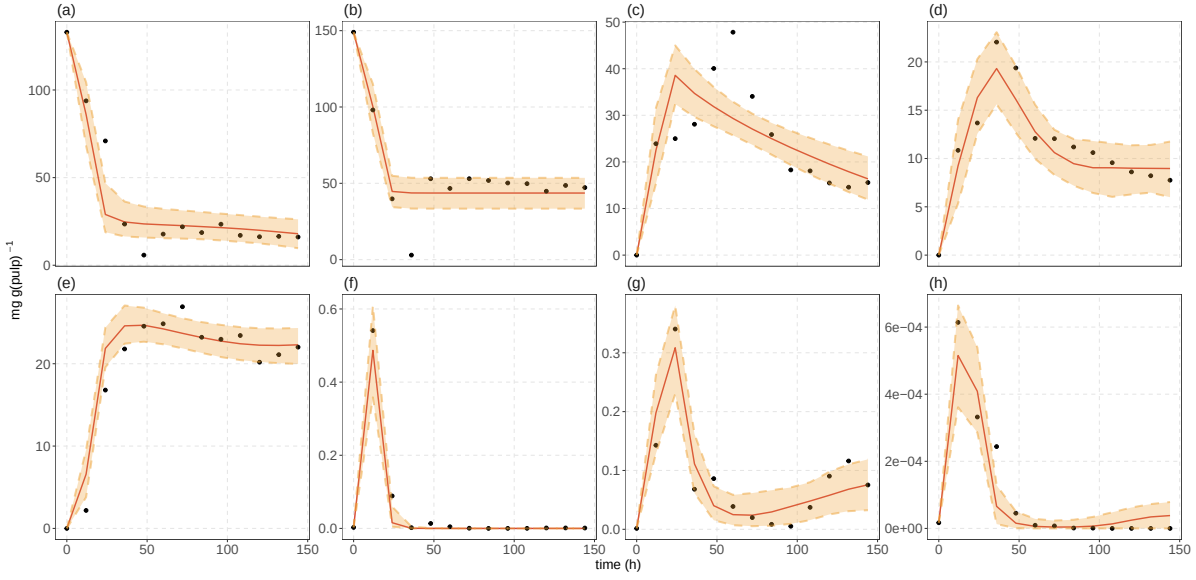


Figure B.19: Posterior predictions of model iteration (MI) corresponding to mechanisms M1 and M3, MI(1,3), fitted to dataset *brwb4* reported by Pereira *et al.* [14]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

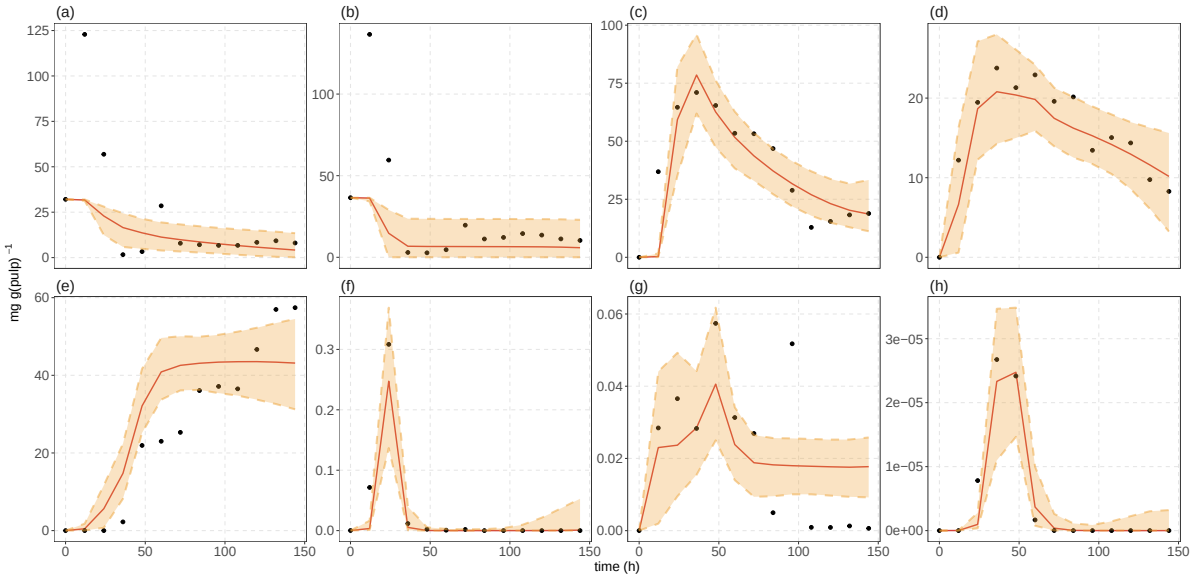


Figure B.20: Posterior predictions of model iteration (MI) corresponding to mechanisms M1 and M4, MI(1,4), fitted to dataset *brst2* reported by Pereira *et al.* [14]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

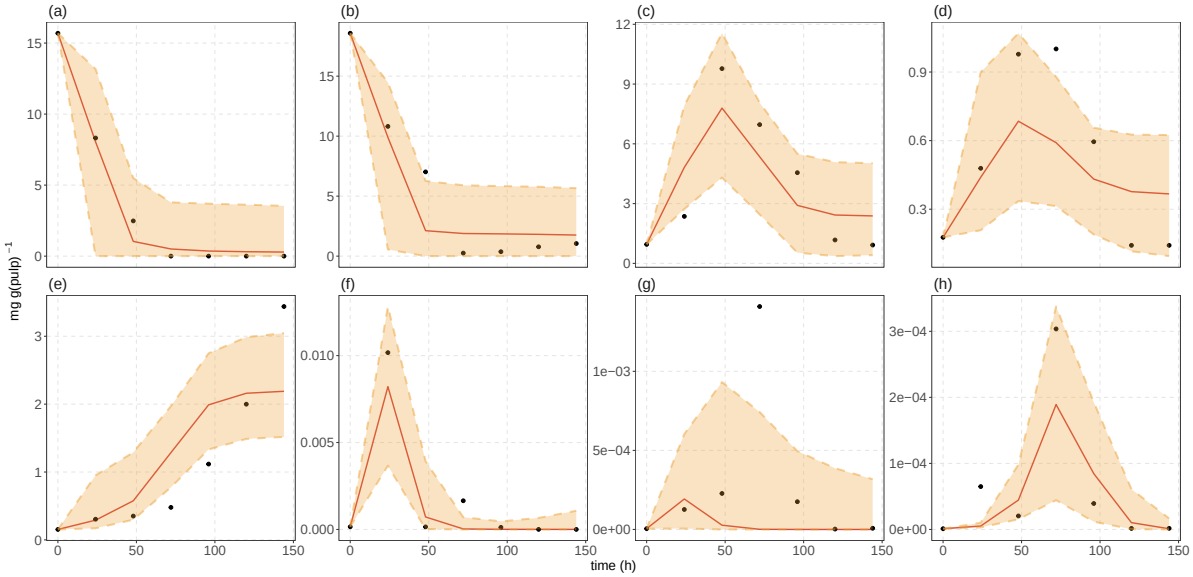


Figure B.21: Posterior predictions of model iteration (MI) corresponding to baseline model, MI(0), fitted to dataset *brwb7* reported by Moreira *et al.* [15]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

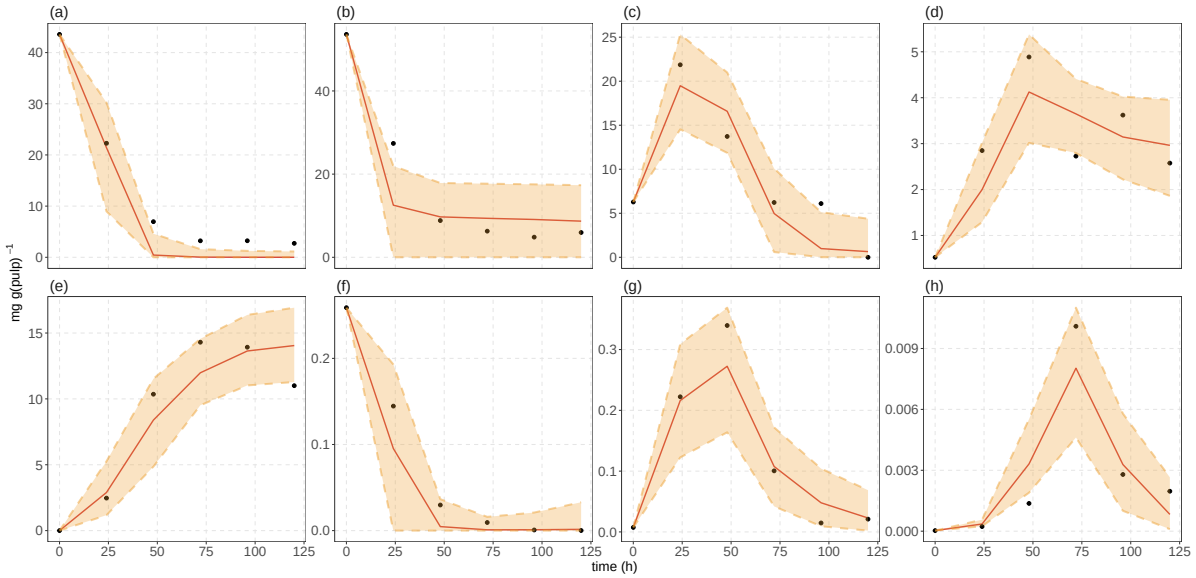


Figure B.22: Posterior predictions of model iteration (MI) corresponding to baseline model, MI(0), fitted to dataset *hnwb1* reported by Romanens *et al.* [115]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

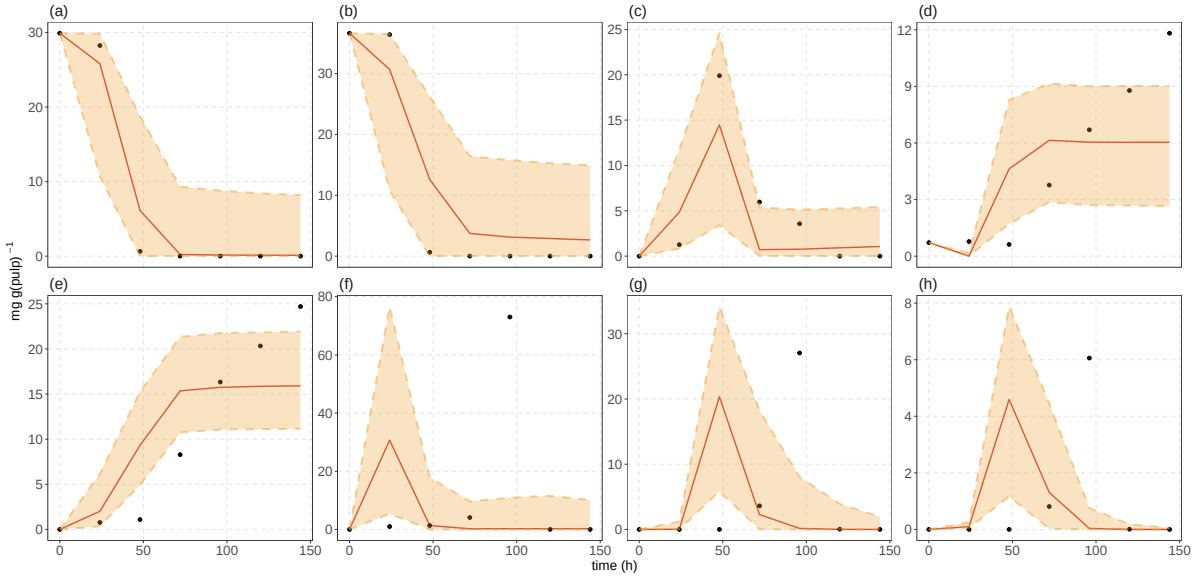


Figure B.23: Posterior predictions of model iteration (MI) corresponding to mechanisms M3 and M4, MI(3,4), fitted to dataset *ecpb1* reported by Lee *et al.* [116]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

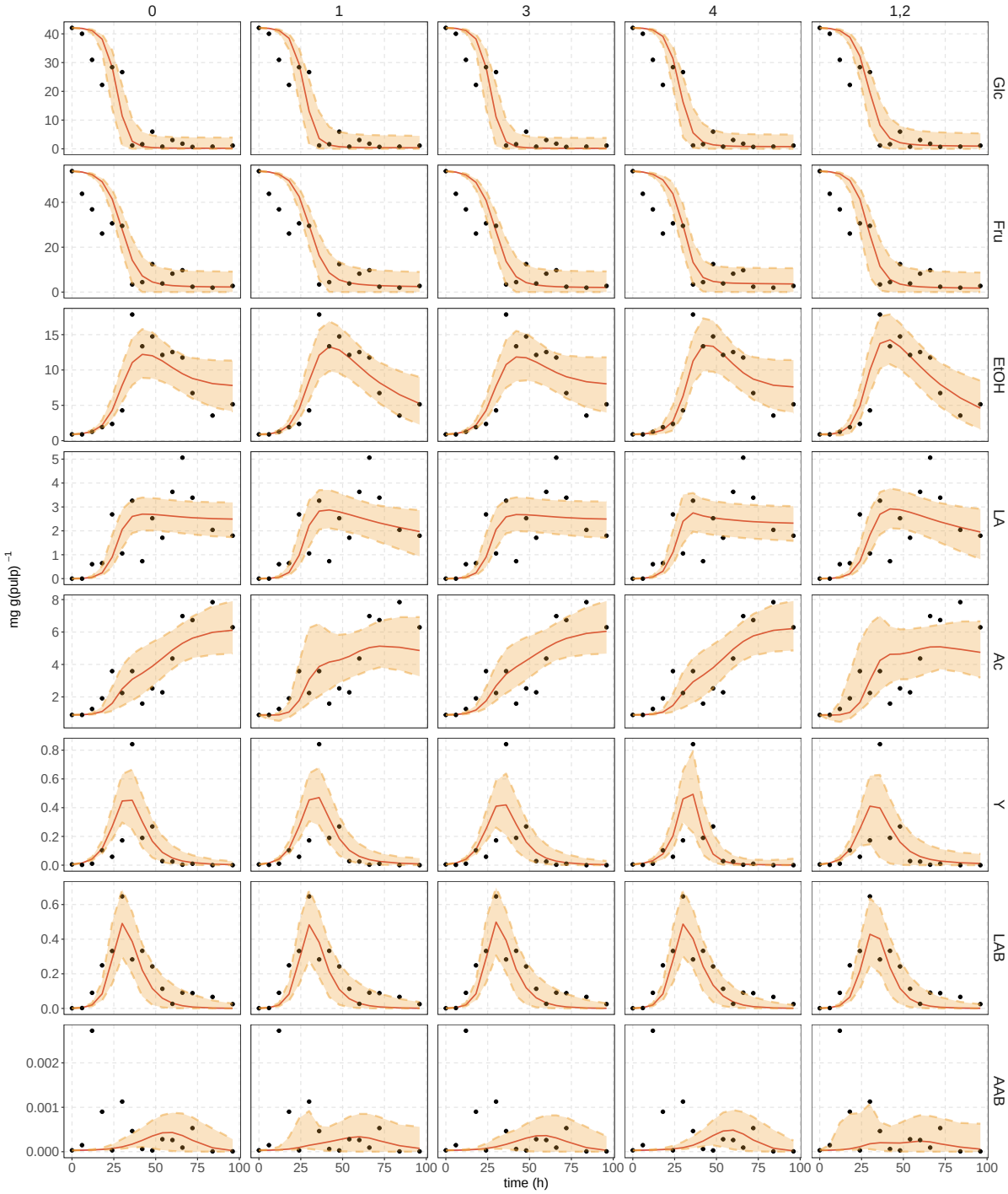


Figure B.24: Posterior predictions across model iterations (MIs) fitted on dataset *ecwb1* reported by Papalexandratou *et al.* [113]. Columns names refer to mechanisms involved in each model iteration (MI). Row names refer to metabolites and microbial groups: glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA), acetic acid (Ac), yeasts (Y), lactic acid bacteria (LAB), and acetic acid bacteria (AAB). Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

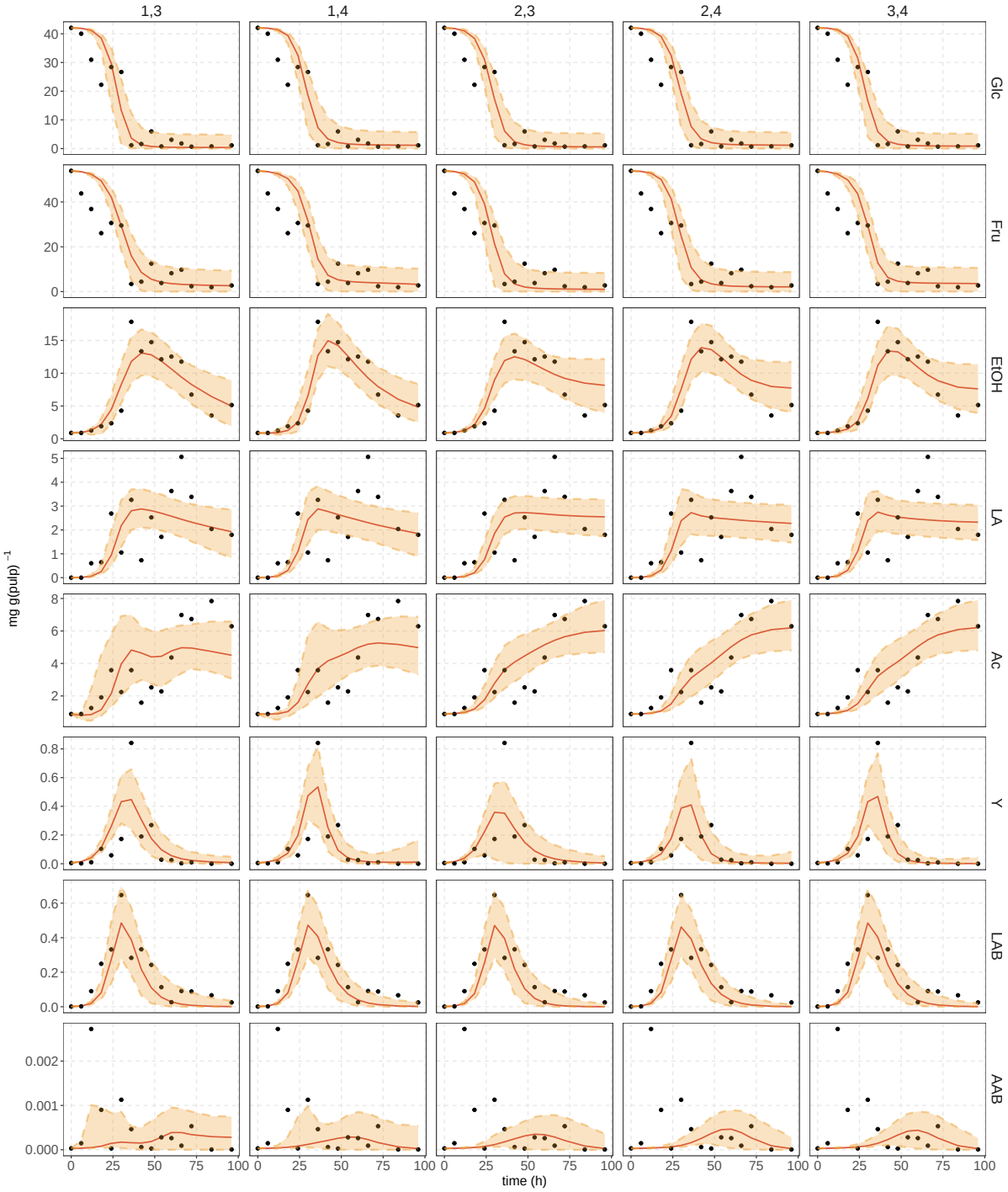


Figure B.24 **(cont.)**: Posterior predictions across model iterations (MIs) fitted on dataset *ecwb1* reported by Papalexandratou *et al.* [113]. Columns names refer to mechanisms involved in each model iteration (MI). Row names refer to metabolites and microbial groups: glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA), acetic acid (Ac), yeasts (Y), lactic acid bacteria (LAB), and acetic acid bacteria (AAB). Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

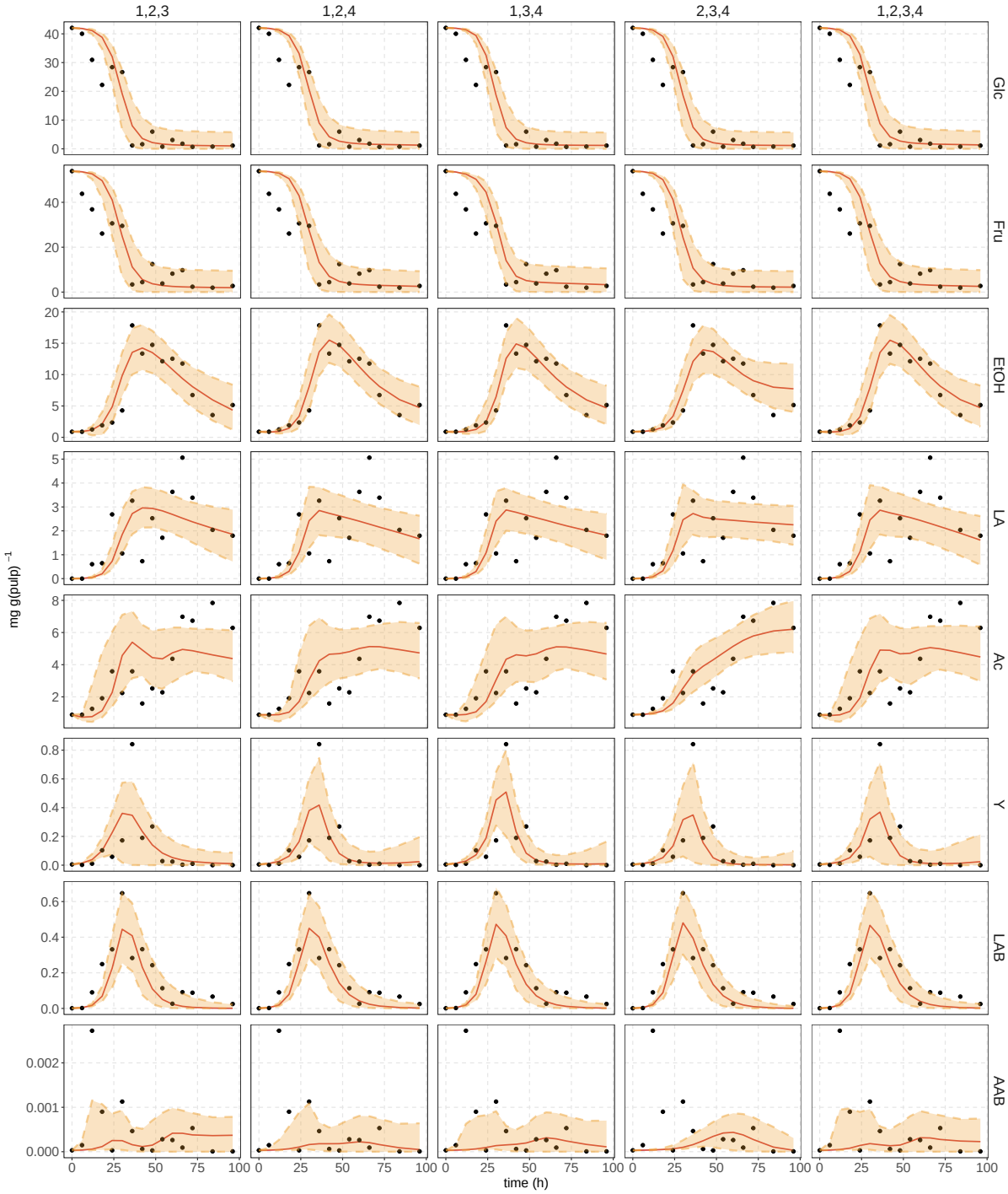


Figure B.24 **(cont.)**: Posterior predictions across model iterations (MIs) fitted on dataset *ecwb1* reported by Papalexandratou *et al.* [113]. Columns names refer to mechanisms involved in each model iteration (MI). Row names refer to metabolites and microbial groups: glucose (Glc), fructose (Fru), ethanol (EtOH), lactic acid (LA), acetic acid (Ac), yeasts (Y), lactic acid bacteria (LAB), and acetic acid bacteria (AAB). Solid red lines represent medians of the posterior predictions, solid black points denote experimental data and orange ribbons describe 95% credible intervals of posterior predictions.

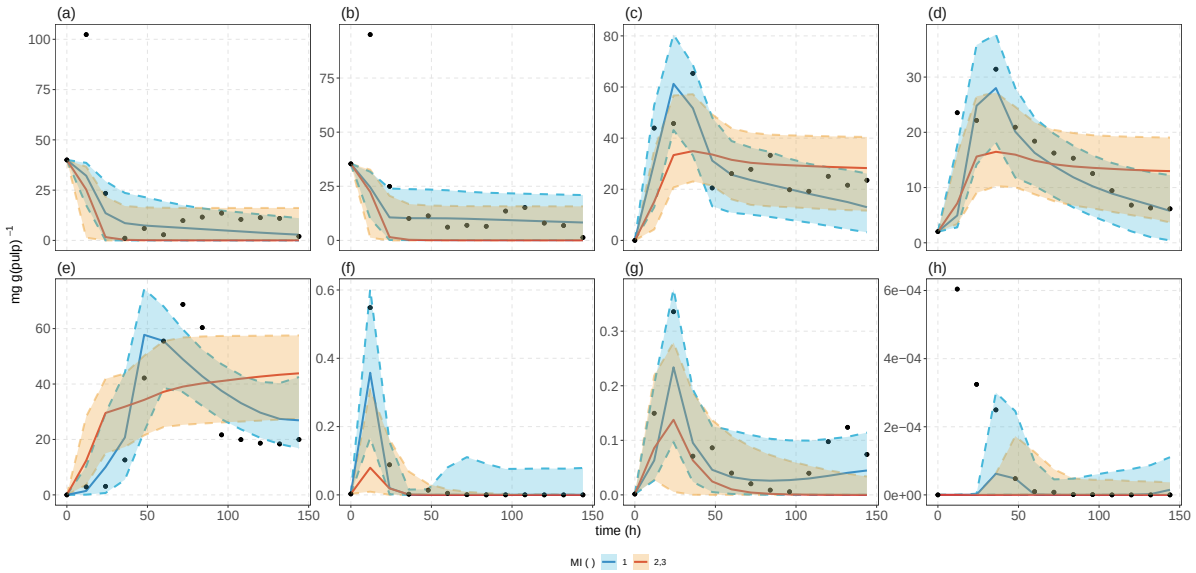


Figure B.25: Comparison of posterior predictions for model iterations (MIs) MI(1) and MI(2,3), fitted to dataset *brpb1* reported by Pereira *et al.* [68]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. For MI(1), solid blue lines represent medians of the posterior predictions and sky-blue ribbons describe their 95% credible intervals. For MI(2,3), solid red lines represent medians of the posterior predictions and orange ribbons describe their 95% credible intervals. Solid black points denote experimental data.

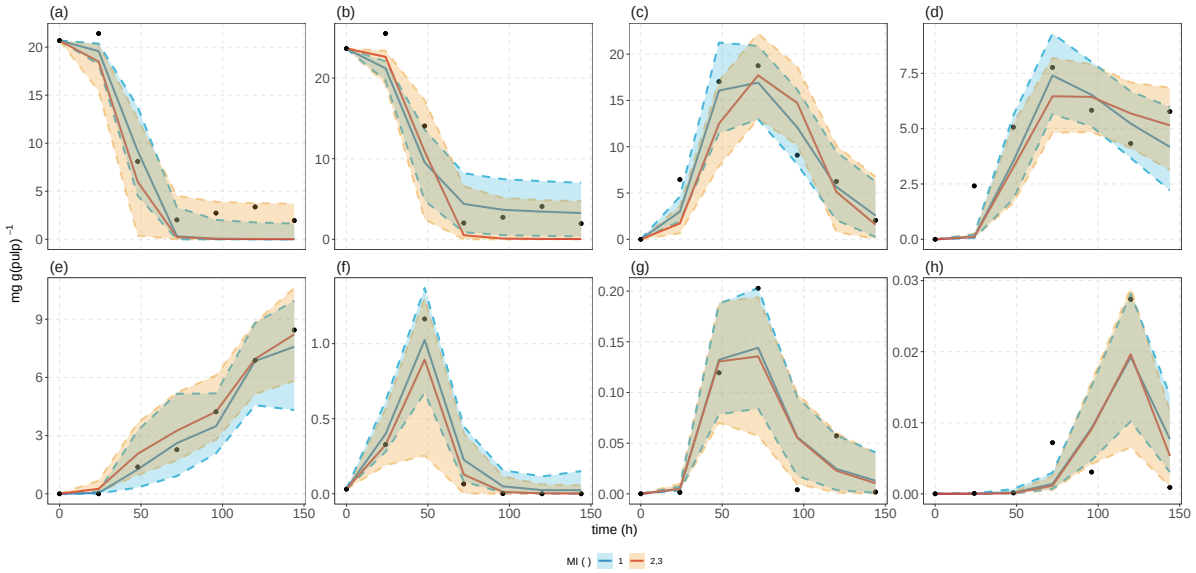


Figure B.26: Comparison of posterior predictions for model iterations (MIs) MI(1) and MI(2,3), fitted to dataset *niwb2* reported by Papalexandratou *et al.* [117]. Metabolites: (a) glucose, (b) fructose, (c) ethanol, (d) lactic acid and (e) acetic acid. Microbial groups: (f) yeast, (g) lactic acid bacteria and (h) acetic acid bacteria. For MI(1), solid blue lines represent medians of the posterior predictions and sky-blue ribbons describe their 95% credible intervals. For MI(2,3), solid red lines represent medians of the posterior predictions and orange ribbons describe their 95% credible intervals. Solid black points denote experimental data.

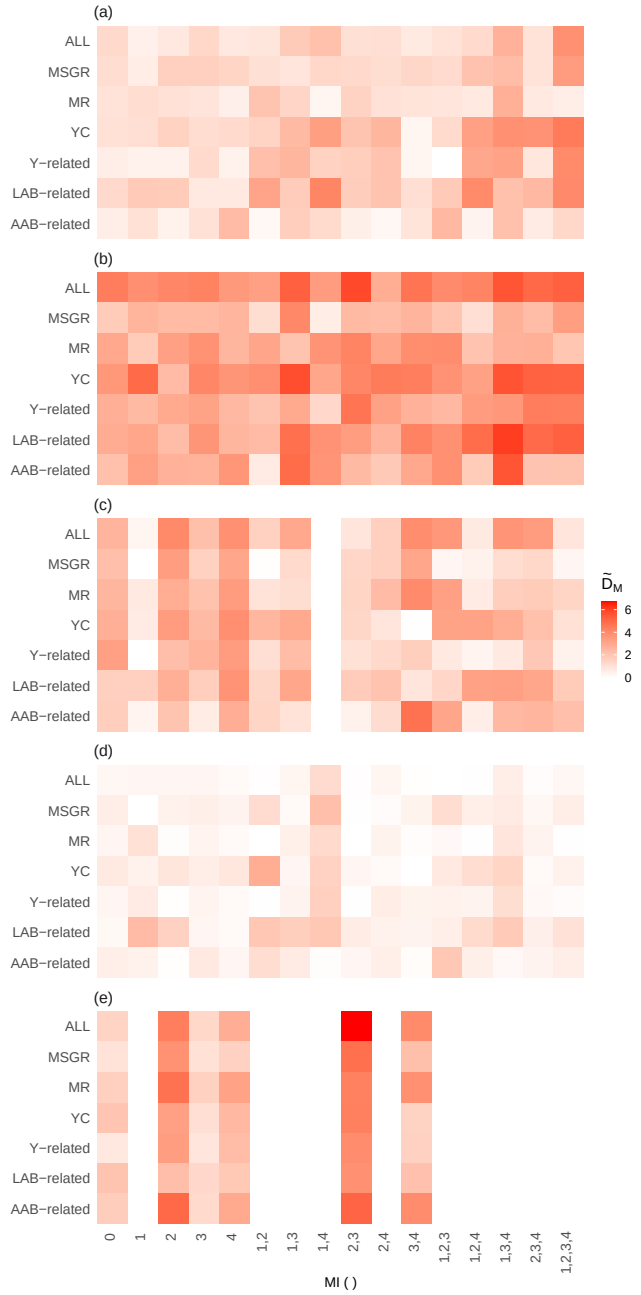


Figure B.27: Heat map of medians of pairwise squared Malahanobis distances ($\tilde{D}_M$) between centroids of grouping classes scores from Principal Component Analysis (PCA) per model iteration (MI) and subgroups of parameter estimates. (a) Country, (b) cocoa cultivar, (c) fermentation method, (d) turning of fermenting mass, and (e) use of controlled temperature. White rows in (b), (c) and (e) correspond to model iterations (MIs) where only one group class was available and PCA groupings could not be performed. Subgroups ALL, MSGR, MR, YC, Y-related, LAB-related and AAB-related correspond to all common MIs parameter estimates, maximum specific growth rates, yield coefficients, yeast-related parameters, lactic acid bacteria-related parameters, and acetic acid bacteria-related paremeters, respectively.

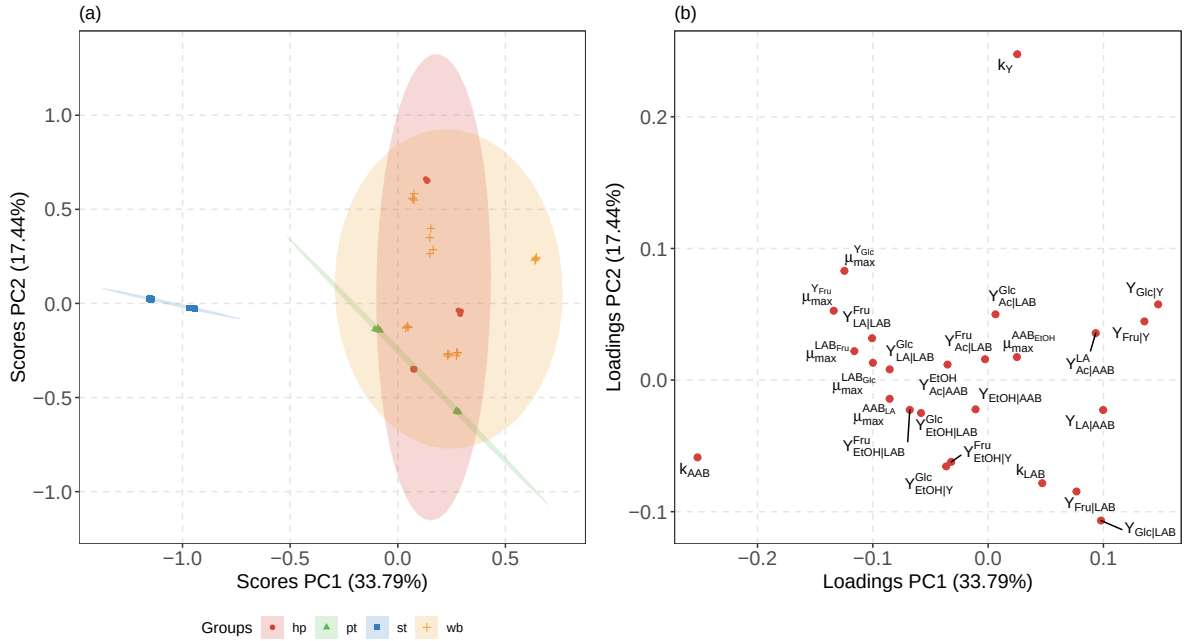


Figure B.28: PCA score (a) and loading plot (b) from all parameters of model iteration (MI) MI(2). Heap (hp), platform (pt), stainless-steel tank (st) and wooden box (wb) fermentation methods are shown. Parameters located on the left and right with respect to 0 in the horizontal axis in the loading plot determine differentiation between stainless-steel tank and rest of fermentation methods, respectively.

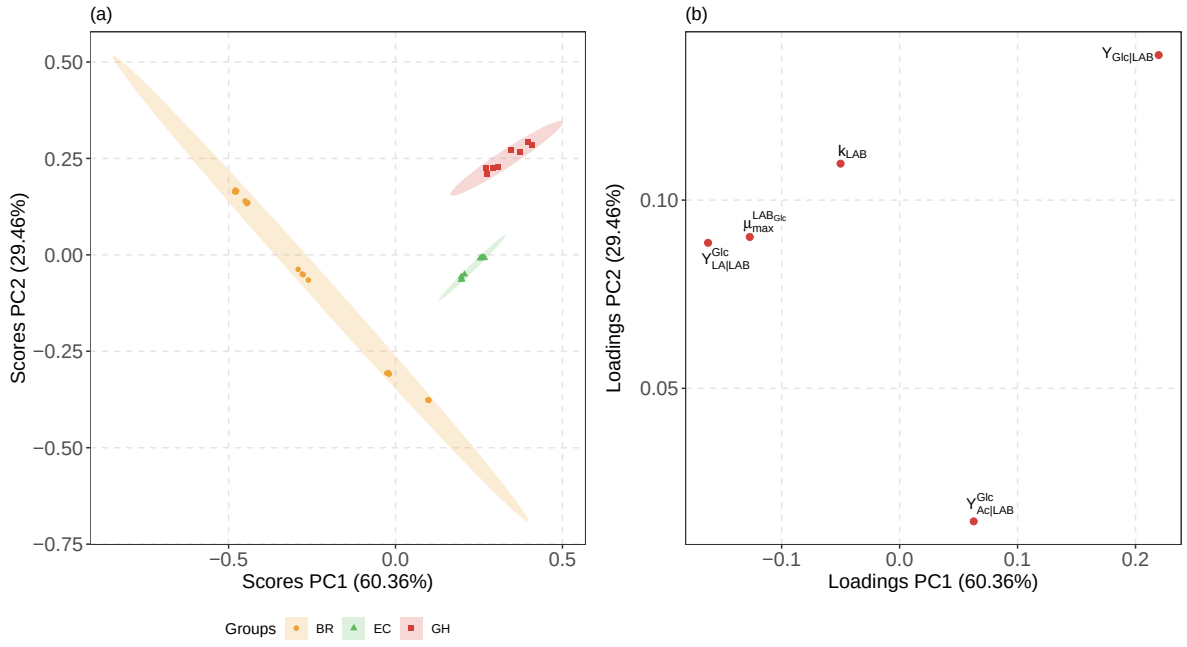


Figure B.29: PCA score (a) and loading plot (b) from lactic acid bacteria-related parameters of model iteration (MI) MI(1,3,4). Brazil (BR), Ecuador (EC) and Ghana (GH) are clearly separated. Parameters located on the left and right with respect to 0 in the horizontal axis in the loading plot determine differentiation between all classes.

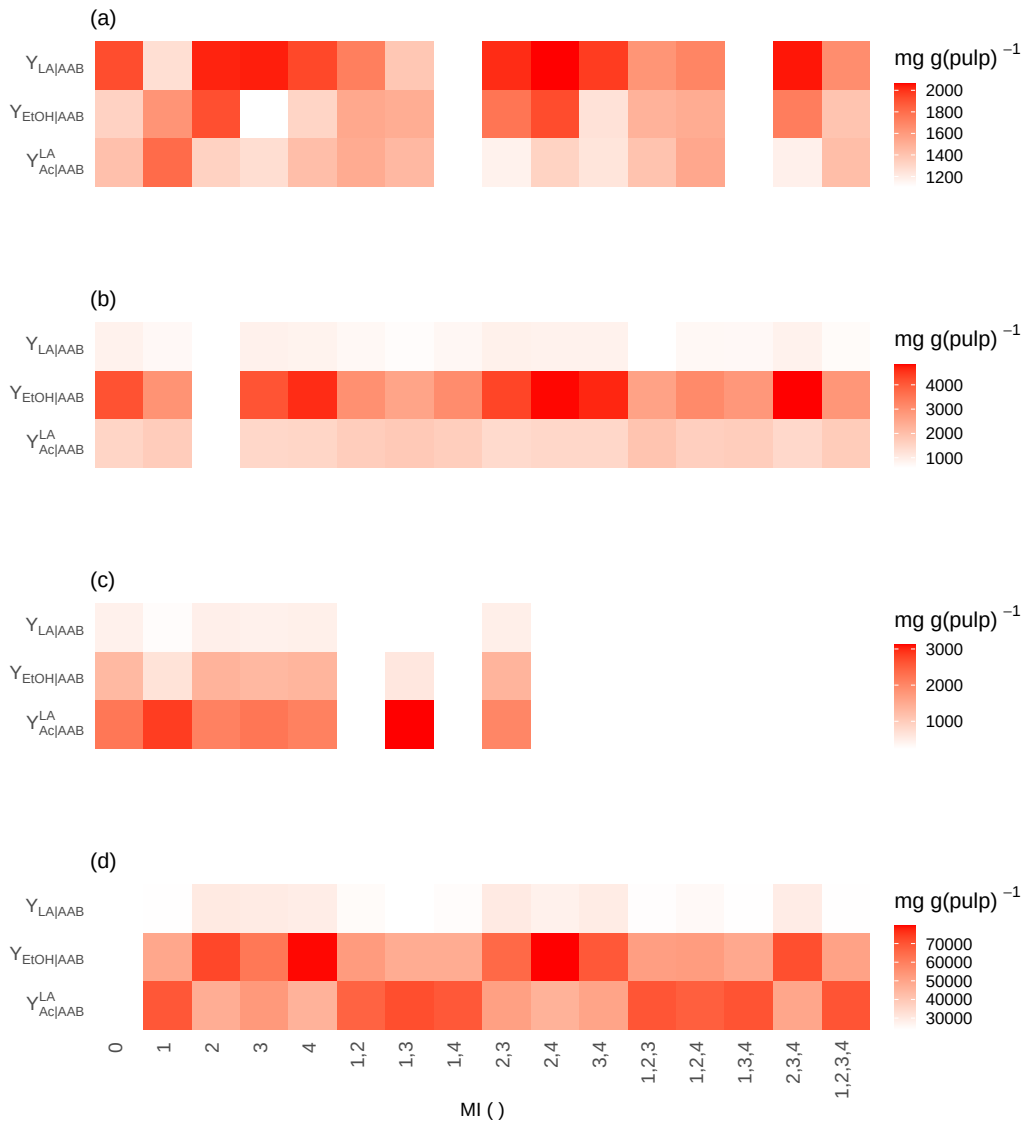


Figure B.30: Heat map of posterior distributions means of $Y_{\mathrm{LA|AAB}}$, $Y_{\mathrm{EtOH|AAB}}$, and $Y^{\mathrm{LA}}_{\mathrm{Ac|AAB}}$. Dataset : (a) *ghhp1*, (b) *ecwb1*, (c) *ecwb2* and (d) *brpb1*.

References

[1] Moreno-Zambrano, M., Grimbs, S., Ullrich, M. S. & Hütt, M.-T. (2018) A mathematical model of cocoa bean fermentation. *Royal Society Open Science*, **5**(10), 180 964. (doi:10.1098/rsos.180964)

[2] Moreno-Zambrano, M., Ullrich, M. S. & Hütt, M.-T. (2021) Exploring cocoa bean fermentation mechanisms by kinetic modelling. (Submitted to Royal Society Open Science).

[3] Lopez, A. S. & Dimick, P. S. (1995) Cocoa Fermentation. In *Biotechnology* (eds G. Reed & T. W. Nagodawithana), vol. 9, chap. 14, pp. 561–577. Weinheim: VCH Verlagsgesellschaft mbH, 2nd edn.

[4] Schwan, R. F., Pereira, G. V. d. M. & Fleet, G. H. (2014) Microbial Activities During Cocoa Fermentation. In *Cocoa and Coffee Fermentations* (eds R. F. Schwan & G. H. Fleet), chap. 4, pp. 129–192. Boca Raton: CRC Press.

[5] Pereira, G. V. d. M., Soccol, V. T. & Soccol, C. R. (2016) Current state of research on cocoa and coffee fermentations. *Current Opinion in Food Science*, **7**, 50–57. (doi:10.1016/j.cofs.2015.11.001)

[6] 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, R. R. Raya & G. M. Vignolo), chap. 16, pp. 248–278. Chichester: Wiley Blackwell, 2nd edn. (doi:10.1002/9781118868386.ch16)

[7] Moss, S. & Badenoch, A. (2009) *Chocolate. A Global History.* London: Reaktion Books.

[8] Thompson, S. S., Miller, K. B. & Lopez, A. S. (2007) Cocoa and coffee. In *Food microbiology: Fundamentals and frontiers* (eds M. P. Doyle & L. R. Beuchat), chap. 39, pp. 837–850. Washington: ASM Press, 3rd edn.

[9] Varnam, A. H. & Sutherland, J. P. (1994) *Beverages, Technology, Chemistry and Microbiology.* London: Chapman &. Hall, 1st edn.

[10] Jespersen, L., Nielsen, D. S., Hønholt, S. & Jakobsen, M. (2005) Occurrence and diversity of yeasts involved in fermentation of West African cocoa beans. *FEMS Yeast Research*, **5**(4-5), 441–453. (doi:10.1016/j.femsyr.2004.11.002)

[11] Crafack, M., Mikkelsen, M. B., Saerens, S., Knudsen, M., Blennow, A., Lowor, S., Takrama, J., Swiegers, J. H., Petersen, G. B. *et al.* (2013) Influencing cocoa flavour using *Pichia kluyveri* and *Kluyveromyces marxianus* in a defined mixed starter culture for cocoa fermentation. *International Journal of Food Microbiology*, **167**(1), 103–116. (doi:10.1016/j.ijfoodmicro.2013.06.024)

[12] Camu, N., De Winter, T., Verbrugghe, K., Cleenwerck, I., Vandamme, P., Takrama, J. S., Vancanneyt, M. & De Vuyst, L. (2007) Dynamics and biodiversity of populations of lactic acid bacteria and acetic acid bacteria involved in spontaneous heap fermentation of cocoa beans in Ghana. *Applied and Environmental Microbiology*, **73**(6), 1809–1824. (doi:10.1128/AEM.02189-06)

[13] Lefeber, T., Gobert, W., Vrancken, G., Camu, N. & De Vuyst, L. (2011) Dynamics and species diversity of communities of lactic acid bacteria and acetic acid bacteria during spontaneous cocoa bean fermentation in vessels. *Food Microbiology*, **28**(3), 457–464. (doi:10.1016/j.fm.2010.10.010)

[14] Pereira, G. V. d. M., Magalhães, K. T., de Almeida, E. G., da Silva Coelho, I. & Schwan, R. F. (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. *International Journal of Food Microbiology*, **161**(2), 121–133. (doi:10.1016/j.ijfoodmicro.2012.11.018)

[15] Moreira, I. M. d. V., Miguel, M. G. d. C. P., Duarte, W. F., Dias, D. R. & Schwan, R. F. (2013) Microbial succession and the dynamics of metabolites and sugars during the fermentation of three different cocoa (*Theobroma cacao* L.) hybrids. *Food Research International*, **54**(1), 9–17. (doi:10.1016/j.foodres.2013.06.001)

[16] Schwan, R. F. (1998) Cocoa fermentations conducted with a defined microbial cocktail inoculum. *Applied and Environmental Microbiology*, **64**(4), 1477–1483. (doi:10.1128/AEM.64.4.1477-1483.1998)

[17] Camu, N., González, Á., De Winter, T., Van Schoor, A., De Bruyne, K., Vandamme, P., Takrama, J. S., Addo, S. K. & De Vuyst, L. (2008) Influence of Turning and Environmental Contamination on the Dynamics of Populations of Lactic Acid and Acetic Acid Bacteria Involved in Spontaneous Cocoa Bean Heap Fermentation in Ghana. *Applied and Environmental Microbiology*, **74**(1), 86–98. (doi:10.1128/AEM.01512-07)

[18] De Vuyst, L. & Weckx, S. (2016) The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *Journal of Applied Microbiology*, **121**(1), 5–17. (doi:10.1111/jam.13045)

[19] Adler, P., Bolten, C. J., Dohnt, K., Hansen, C. E. & Wittmann, C. (2013) Core Fluxome and Metafluxome of Lactic Acid Bacteria under Simulated Cocoa Pulp Fermentation Conditions. *Applied and Environmental Microbiology*, **79**(18), 5670–5681. (doi:10.1128/AEM.01483-13)

[20] Nielsen, D., Teniola, O., Ban-Koffi, L., Owusu, M., Andersson, T. & Holzapfel, W. (2007) The microbiology of Ghanaian cocoa fermentations analysed using culture-dependent and culture-independent methods. *International Journal of Food Microbiology*, **114**(2), 168–186. (doi:10.1016/j.ijfoodmicro.2006.09.010)

[21] Moens, F., Lefeber, T. & De Vuyst, L. (2014) Oxidation of Metabolites Highlights the Microbial Interactions and Role of *Acetobacter pasteurianus* during Cocoa Bean Fermentation. *Applied and Environmental Microbiology*, **80**(6), 1848–1857. (doi:10.1128/AEM.03344-13)

[22] Kresnowati, M. T. A. P., Gunawan, A. Y. & Muliyadini, W. (2015) Kinetics model development of cocoa bean fermentation. *AIP Conference Proceedings*, **1699**(1), 030 004–1–9. (doi:10.1063/1.4938289)

[23] López-Pérez, P. A., Cuervo-Parra, J. A., Robles-Olvera, V. J., Jimenes, G. D. C. R., España, V. H. P. & Romero-Cortes, T. (2018) Development of a Novel Kinetic Model for Cocoa Fermentation Applying the Evolutionary Optimization Approach. *International Journal of Food Engineering*, **14**(5-6). (doi:10.1515/ijfe-2017-0206)

[24] Torres, N. V. & Santos, G. (2015) The (Mathematical) Modeling Process in Biosciences. *Frontiers in Genetics*, **6**, 354. (doi:10.3389/fgene.2015.00354)

[25] Motta, S. & Pappalardo, F. (2012) Mathematical modeling of biological systems. *Briefings in Bioinformatics*, **14**(4), 411–422. (doi:10.1093/bib/bbs061)

[26] Voit, E. O. (1992) Optimization in integrated biochemical systems. *Biotechnology and Bioengineering*, **40**(5), 572–582. (doi:10.1002/bit.260400504)

[27] Monod, J. (1949) The Growth of Bacterial Cultures. *Annual Review of Microbiology*, **3**(1), 371–394. (doi:10.1146/annurev.mi.03.100149.002103)

[28] Owens, J. & Legan, J. (1987) Determination of the Monod substrate saturation constant for microbial growth. *FEMS Microbiology Letters*, **46**(4), 419–432.

[29] Contois, D. E. (1959) Kinetics of Bacterial Growth: Relationship between Population Density and Specific Growth Rate of Continuous Cultures. *Microbiology*, **21**(1), 40–50. (doi:10.1099/00221287-21-1-40)

[30] Peleg, M. & Corradini, M. G. (2011) Microbial Growth Curves: What the Models Tell Us and What They Cannot. *Critical Reviews in Food Science and Nutrition*, **51**(10), 917–945. (doi:10.1080/10408398.2011.570463)

[31] Klipp, E., Herwig, R., Kowald, A., Wierling, C. & Lehrach, H. (2005) *Systems biology in practice: Concepts, implementation and application*. Weinheim: Wiley-Blackwell.

[32] Watson, H. E. (1908) A Note on the Variation of the Rate of Disinfection with Change in the Concentration of the Disinfectant. *The Journal of Hygiene*, **8**(4), 536–542.

[33] Coulson, J., Richardson, J. & Peacock, D. (1994) *Chemical and Biochemical Reactors & Process Control*. Chemical Engineering. Oxford: Butterworth-Heinemann, 3rd edn.

[34] Lefeber, T., Janssens, M., Camu, N. & De Vuyst, L. (2010) Kinetic Analysis of Strains of Lactic Acid Bacteria and Acetic Acid Bacteria in Cocoa Pulp Simulation Media toward Development of a Starter Culture for Cocoa Bean Fermentation. *Applied and Environmental Microbiology*, **76**(23), 7708–7716. (doi:10.1128/AEM.01206-10)

[35] Boersch-Supan, P. H., Ryan, S. J. & Johnson, L. R. (2017) deBInfer: Bayesian inference for dynamical models of biological systems in R. *Methods in Ecology and Evolution*, **8**(4), 511–518. (doi:10.1111/2041-210X.12679)

[36] Gelman, A., Simpson, D. & Betancourt, M. (2017) The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, **19**(10). (doi:10.3390/e19100555)

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

[38] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014) *Bayesian Data Analysis*. Texts in Statistical Science. London: Chapman & Hall/CRC, 3rd edn.

[39] Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109. (doi:10.1093/biomet/57.1.97)

[40] Hoffman, M. D. & Gelman, A. (2014) The No-U-turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**(1), 1593–1623.

[41] Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987) Hybrid Monte Carlo. *Physics Letters B*, **195**(2), 216–222. (doi:10.1016/0370-2693(87)91197-X)

[42] Neal, R. M. (1994) An Improved Acceptance Procedure for the Hybrid Monte Carlo Algorithm. *Journal of Computational Physics*, **111**(1), 194–203. (doi:10.1006/jcph.1994.1054)

[43] Betancourt, M. & Girolami, M. (2015) Hamiltonian Monte Carlo for Hierarchical Models. In *Current Trends in Bayesian Methodology with Applications* (eds S. K. Upadhyay, U. Singh, D. K. Dey & A. Loganathan), chap. 4, pp. 79–95. New York: Chapman & Hall/CRC, 1st edn.

[44] Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. (2020) Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC. *Bayesian Analysis*, pp. 1 – 28. (doi:10.1214/20-ba1221)

[45] Gelman, A. & Rubin, D. B. (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**(4), 457–472.

[46] Brooks, S. P. & Gelman, A. (1998) General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, **7**(4), 434–455. (doi:10.2307/1390675)

[47] Watanabe, S. (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, **11**, 3571–3594.

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

[49] Petrov, M. (2019) Modelling and Multi-Criteria Decision Making for Selection of Specific Growth Rate Models of Batch Cultivation by *Saccharomyces cerevisiae* Yeast for Ethanol Production. *Fermentation*, **5**(3). (doi:10.3390/fermentation5030061)

[50] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. *et al.* (2017) Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*, **76**(1), 1–32. (doi:10.18637/jss.v076.i01)

[51] Dormand, J. & Prince, P. (1980) A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, **6**(1), 19–26. (doi:10.1016/0771-050X(80)90013-3)

[52] Ahnert, K. & Mulansky, M. (2011) Odeint – Solving Ordinary Differential Equations in C++. *AIP Conference Proceedings*, **1389**(1), 1586–1589. (doi:10.1063/1.3637934)

[53] Rosenbaum, B., Raatz, M., Weithoff, G., Fussmann, G. F. & Gaedke, U. (2019) Estimating parameters from multiple time series of population dynamics using bayesian inference. *Frontiers in Ecology and Evolution*, **6**, 234. (doi:10.3389/fevo.2018.00234)

[54] Schwan, R. F. & Wheals, A. E. (2004) The Microbiology of Cocoa Fermentation and its Role in Chocolate Quality. *Critical Reviews in Food Science and Nutrition*, **44**(4), 205–221. (doi:10.1080/10408690490464104)

[55] Lima, L. J. R., Almeida, M. H., Nout, M. J. R. & Zwietering, M. H. (2011) *Theobroma cacao* L., "The Food of the Gods": Quality Determinants of Commercial Cocoa Beans, with Particular Reference to the Impact of Fermentation. *Critical Reviews in Food Science and Nutrition*, **51**(8), 731–761. (doi:10.1080/10408391003799913)

[56] Ho, V. T. T., Zhao, J. & Fleet, G. (2014) Yeasts are essential for cocoa bean fermentation. *International Journal of Food Microbiology*, **174**, 72–87. (doi:10.1016/j.ijfoodmicro.2013.12.014)

[57] Ho, V. T. T., Zhao, J. & Fleet, G. (2015) The effect of lactic acid bacteria on cocoa bean fermentation. *International Journal of Food Microbiology*, **205**, 54–67. (doi:10.1016/j.ijfoodmicro.2015.03.031)

[58] MacManus Chinenye, N., Ogunlowo, A. & Olukunle, O. (2010) Cocoa Bean (*Theobroma cacao* L.) Drying Kinetics. *Chilean journal of agricultural research*, **70**, 633–639. (doi:10.4067/S0718-58392010000400014)

[59] Hii, C., Law, C. & Suzannah, S. (2012) Drying kinetics of the individual layer of cocoa beans during heat pump drying. *Journal of Food Engineering*, **108**(2), 276–282. (doi:10.1016/j.jfoodeng.2011.08.017)

[60] Adler, P., Frey, L. J., Berger, A., Bolten, C. J., Hansen, C. E. & Wittmann, C. (2014) The Key to Acetate: Metabolic Fluxes of Acetic Acid Bacteria under Cocoa Pulp Fermentation-Simulating Conditions. *Applied and Environmental Microbiology*, **80**(15), 4702–4716. (doi:10.1128/AEM.01048-14)

[61] Ghaly, A. & El-Taweel, A. (1994) Kinetics of batch production of ethanol from cheese whey. *Biomass and Bioenergy*, **6**(6), 465–478. (doi:https://doi.org/10.1016/0961-9534(94)00079-9)

[62] Papalexandratou, Z., Vrancken, G., Bruyne, K. D., Vandamme, P. & Vuyst, L. D. (2011) Spontaneous organic cocoa bean box fermentations in Brazil are characterized by a restricted species diversity of lactic acid bacteria and acetic acid bacteria. *Food Microbiology*, **28**(7), 1326–1338. (doi:10.1016/j.fm.2011.06.003)

[63] Schwabe, A. & Bruggeman, F. J. (2014) Single yeast cells vary in transcription activity not in delay time after a metabolic shift. *Nature Communications*, **5**, 4798. (doi:10.1038/ncomms5798)

[64] Brenner, D. J., Krieg, N. R., Staley, J. T. & Garrity, G. M. (eds) (2009) *Bergey's Manual of Systematic Bacteriology: Volume 2: The proteobacteria. Part C.* New York: Springer.

[65] Vos, P., Garrity, G., Jones, D., Krieg, N., Ludwig, W., Rainey, F., Schleifer, K. & Whitman, W. (eds) (2009) *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes.* New York: Springer.

[66] Neidhardt, F. & Curtiss, R. (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1. Washington: ASM Press, 2nd edn.

[67] Phillips, R., Kondev, J., Theriot, J. & Garcia, H. (2012) *Physical Biology of the Cell.* New York: Taylor & Francis Group, 2nd edn.

[68] Pereira, G. V. d. M., Miguel, M. G. d. C. P., Ramos, C. L. & Schwan, R. F. (2012) Microbiological and physicochemical characterization of small-scale cocoa fermentations and screening of yeast and bacterial strains to develop a defined starter culture. *Applied and Environmental Microbiology*, **78**(15), 5395–5405. (doi:10.1128/AEM.01144-12)

[69] Maier, R. M. (2009) Bacterial Growth. In *Environmental Microbiology* (eds R. M. Maier, I. L. Pepper & C. P. Gerba), chap. 3, pp. 37–54. San Diego: Academic Press, 2nd edn. (doi:10.1016/B978-0-12-370519-8.00003-1)

[70] Reilly, P. J. (1974) Stability of commensalistic systems. *Biotechnology and Bioengineering*, **16**(10), 1373–1392. (doi:10.1002/bit.260161006)

[71] Fredrickson, A. G. (1977) Behavior of Mixed Cultures of Microorganisms. *Annual Review of Microbiology*, **31**(1), 63–88. (doi:10.1146/annurev.mi.31.100177.000431)

[72] Fredrickson, A. & Stephanopoulos, G. (1981) Microbial competition. *Science*, **213**(4511), 972–979. (doi:10.1126/science.7268409)

[73] R Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[74] Stan Development Team (2016) RStan: the R interface to Stan. R package version 2.14.1.

[75] Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. New York: Taylor & Francis, 2nd edn.

[76] Sawilowsky, S. (2009) New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, **8**(2), 467–474. (doi:10.22237/jmasm/1257035100)

[77] Kaspar von Meyenburg, H. (1969) Energetics of the budding cycle of *Saccharomyces cerevisiae* during glucose limited aerobic growth. *Archiv für Mikrobiologie*, **66**(4), 289–303. (doi:10.1007/BF00414585)

[78] Lee, K. J., Skotnicki, M. L., Tribe, D. E. & Rogers, P. L. (1981) The kinetics of ethanol production by *Zymomonas mobilis* on fructose and sucrose media. *Biotechnology Letters*, **3**(5), 207–212. (doi:10.1007/BF00154646)

[79] Bideaux, C., Alfenore, S., Cameleyre, X., Molina-jouve, C., Uribelarrea, J.-l. & Guillouet, E. (2006) Minimization of Glycerol Production during the High-Performance Fed-Batch Ethanolic Fermentation Process in *Saccharomyces cerevisiae*, Using a Metabolic Model as a Prediction Tool. *Applied and Environmental Microbiology*, **72**(3), 2134–2140. (doi:10.1128/AEM.72.3.2134)

[80] Alfenore, S., Cameleyre, X., Benbadis, L., Bideaux, C., Uribelarrea, J.-L., Goma, G., Molina-Jouve, C. & Guillouet, S. E. (2004) Aeration strategy: a need for very high ethanol performance in *Saccharomyces cerevisiae* fed-batch process. *Applied Microbiology and Biotechnology*, **63**(5), 537–542. (doi:10.1007/s00253-003-1393-5)

[81] Aguilar-Uscanga, M., Garcia-Alvarado, Y., Gomez-Rodriguez, J., Phister, T., Delia, M. & Strehaiano, P. (2011) Modelling the growth and ethanol production of *Brettanomyces bruxellensis* at different glucose concentrations. *Letters in Applied Microbiology*, **53**(2), 141–149. (doi:10.1111/j.1472-765X.2011.03081.x)

[82] Sonnleitner, B. & Käppeli, O. (1986) Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis. *Biotechnology and Bioengineering*, **28**(6), 927–937. (doi:10.1002/bit.260280620)

[83] Wang, D., Xu, Y., Hu, J. & Zhao, G. (2004) Fermentation Kinetics of Different Sugars by Apple Wine Yeast *Saccharomyces cerevisiae*. *Journal of the Institute of Brewing*, **110**(4), 340–346. (doi:10.1002/j.2050-0416.2004.tb00630.x)

[84] Snoep, J. L., Mrwebi, M., Schuurmans, J. M., Rohwer, J. M. & Teixeira de Mattos, M. J. (2009) Control of specific growth rate in *Saccharomyces cerevisiae*. *Microbiology*, **155**(5), 1699–1707.

[85] Berry, A. R., Franco, C. M., Zhang, W. & Middelberg, A. P. (1999) Growth and lactic acid production in batch culture of *Lactobacillus rhamnosus* in a defined medium. *Biotechnology Letters*, **21**(2), 163–167. (doi:10.1023/A:1005483609065)

[86] 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*. *Journal of Microbiology and Biotechnology*, **6**(6), 482–486.

[87] Siebold, M., Frieling, P., Joppien, R., Rindfleisch, D., Schügerl, K. & Röper, H. (1995) Comparison of the production of lactic acid by three different lactobacilli and its recovery by extraction and electrodialysis. *Process Biochemistry*, **30**(1), 81–95. (doi:10.1016/0032-9592(95)87011-3)

[88] Corman, A. & Pave, A. (1983) On parameter estimation of Monod's bacterial growth model from batch culture data. *The Journal of General and Applied Microbiology*, **29**, 91–101.

[89] Åkerberg, 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* ssp. *lactis* ATCC 19435 in whole-wheat flour. *Applied Microbiology and Biotechnology*, **49**(6), 682–690. (doi:10.1007/s002530051232)

[90] Fu, W. & Mathews, A. (1999) Lactic acid production from lactose by *Lactobacillus plantarum*: kinetic model and effects of pH, substrate, and oxygen. *Biochemical Engineering Journal*, **3**(3), 163–170. (doi:10.1016/S1369-703X(99)00014-5)

[91] Toran-Diaz, I., Delezon, C. & Baratti, J. (1983) The kinetics of ethanol production by *Zymomonas mobilis* on fructose medium. *Biotechnology Letters*, **5**(6), 409–412. (doi:10.1007/BF00131282)

[92] Arroyo-López, F. N., Querol, A. & Barrio, E. (2009) Application of a substrate inhibition model to estimate the effect of fructose concentration on the growth of diverse *Saccharomyces cerevisiae* strains. *Journal of Industrial Microbiology & Biotechnology*, **36**(5), 663–669. (doi:10.1007/s10295-009-0535-x)

[93] Park, Y. S., Ohtake, H., Fukaya, M., Okumura, H., Kawamura, Y. & Toda, K. (1989) Enhancement of acetic acid production in a high cell-density culture of *Acetobacter aceti*. *Journal of Fermentation and Bioengineering*, **68**(5), 315–319. (doi:10.1016/0922-338X(89)90004-4)

[94] Bar, R., Gainer, J. L. & Kirwan, D. J. (1987) An unusual pattern of product inhibition: Batch acetic acid fermentation. *Biotechnology and Bioengineering*, **29**(6), 796–798. (doi:10.1002/bit.260290625)

[95] Park, Y. S., 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. *Applied Microbiology and Biotechnology*, **35**(2), 149–153. (doi:10.1007/BF00184678)

[96] Salgueiro Machado, S., Luttik, M. A. H., van Dijken, J. P., Jongejan, J. A. & Pronk, J. T. (1995) Regulation of alcohol-oxidizing capacity in chemostat cultures of *Acetobacter pasteurianus*. *Applied Microbiology and Biotechnology*, **43**(6), 1061–1066. (doi:10.1007/BF00166926)

[97] Luttik, M. A. H., van Spanning, R., Schipper, D., van Dijken, J. & Pronk, J. T. (1997) The Low Biomass Yields of the Acetic Acid Bacterium *Acetobacter pasteurianus* Are Due to a Low Stoichiometry of Respiration-Coupled Proton Translocation. *Applied and Environmental Microbiology*, **63**(9), 3345–3351.

[98] Cheng, N. G., Hasan, M., Chahyo Kumoro, A., Ling, C. F. & Tham, M. (2009) Production of Ethanol by Fed-Batch Fermentation. *Pertanika Journal of Science and Technology*, **17**(2), 399–408.

[99] Moustafa, H. H. & Collins, E. B. (1968) Molar Growth Yields of Certain Lactic Acid Bacteria as Influenced by Autolysis. *Journal of Bacteriology*, **96**(I), 117–125.

[100] Babel, W., Müller, R. H. & Markuske, K. D. (1983) Improvement of growth yield of yeast on glucose to the maximum by using an additional energy source. *Archives of Microbiology*, **136**(3), 203–208. (doi:10.1007/BF00409845)

[101] Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & van Dijken, J. P. (1991) A theoretical evaluation of growth yields of yeasts. *Antonie van Leeuwenhoek*, **59**(1), 49–63. (doi:10.1007/BF00582119)

[102] Gómez, J. M. & Cantero, D. (1998) Kinetics of substrate consumption and product formation in closed acetic fermentation systems. *Bioprocess Engineering*, **18**(6), 439–444. (doi:10.1007/s004490050468)

[103] Hines, K. E., Middendorf, T. R. & Aldrich, R. W. (2014) Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *Journal of General Physiology*, **143**(3), 401–416. (doi:10.1085/jgp.201311116)

[104] Hesseltine, C. W. (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), chap. 6, pp. 52–57. National Academies Press.

[105] Price, N. D., Reed, J. L. & Palsson, B. Ø. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, **2**, 886. (doi:10.1038/nrmicro1023)

[106] Levy, R. & Borenstein, E. (2013) Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, **110**(31), 12 804–12 809. (doi:10.1073/pnas.1300926110)

[107] Magnúsdottir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J. *et al.* (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, **35**(1), 81–89. (doi:10.1038/nbt.3703)

[108] Ponomarova, O., Gabrielli, N., Sévin, D. C., Mülleder, M., Zirngibl, K., Bulyha, K., Andrejev, S., Kafkia, E., Typas, A. *et al.* (2017) Yeast Creates a Niche for Symbiotic Lactic Acid Bacteria through Nitrogen Overflow. *Cell Systems*, **5**(4), 345–357.e6. (doi:10.1016/j.cels.2017.09.002)

[109] De Vuyst, L. & Leroy, F. (2020) Functional role of yeasts, lactic acid bacteria, and acetic acid bacteria in cocoa fermentation processes. *FEMS Microbiology Reviews*, pp. 432–453. (doi:10.1093/femsre/fuaa014)

[110] Figueroa-Hernández, C., Mota-Gutierrez, J., Ferrocino, I., Hernández-Estrada, Z. J., González-Ríos, O., Cocolin, L. & Suárez-Quiroz, M. L. (2019) The challenges and perspectives of the selection of starter cultures for fermented cocoa beans. *International Journal of Food Microbiology*, **301**, 41–50. (doi:10.1016/j.ijfoodmicro.2019.05.002)

[111] John, W. A., Böttcher, N. L., Behrends, B., Corno, M., D'souza, R. N., Kuhnert, N. & Ullrich, M. S. (2020) Experimentally modelling cocoa bean fermentation reveals key factors and their influences. *Food Chemistry*, **302**, 125 335. (doi:10.1016/j.foodchem.2019.125335)

[112] Lagunes Gálvez, S., Loiseau, G., Paredes, J. L., Barel, M. & Guiraud, J.-P. (2007) Study on the microflora and biochemistry of cocoa fermentation in the Dominican Republic. *International Journal of Food Microbiology*, **114**(1), 124–130. (doi:10.1016/j.ijfoodmicro.2006.10.041)

[113] Papalexandratou, Z., Falony, G., Romanens, E., Jimenez, J. C., Amores, F., Daniel, H.-M. & De Vuyst, L. (2011) Species diversity, community dynamics, and metabolite kinetics of the microbiota associated with traditional ecuadorian spontaneous cocoa bean fermentations. *Applied and Environmental Microbiology*, **77**(21), 7698–7714. (doi:10.1128/AEM.05523-11)

[114] Papalexandratou, Z., Lefeber, T., Bahrim, B., Lee, O. S., Daniel, H.-M. & De Vuyst, L. (2013) *Hanseniaspora opuntiae*, *Saccharomyces cerevisiae*, *Lactobacillus fermentum*, and *Acetobacter pasteurianus* predominate during well-performed Malaysian cocoa bean box fermentations, underlining the importance of these microbial species for a successful cocoa bean fermentation process. *Food Microbiology*, **35**(2), 73–85. (doi:10.1016/j.fm.2013.02.015)

[115] Romanens, E., Näf, R., Lobmaier, T., Pedan, V., Leischtfeld, S. F., Meile, L. & Schwenninger, S. M. (2018) A lab-scale model system for cocoa bean fermentation. *Applied Microbiology and Biotechnology*, **102**(7), 3349–3362. (doi:10.1007/s00253-018-8835-6)

[116] Lee, A. H., Neilson, A. P., O'Keefe, S. F., Ogejo, J. A., Huang, H., Ponder, M., Chu, H. S. S., Jin, Q., Pilot, G. *et al.* (2019) A laboratory-scale model cocoa fermentation using dried, unfermented beans and artificial pulp can simulate the microbial and chemical changes of on-farm cocoa fermentation. *European Food Research and Technology*, **245**(2), 511–519. (doi:10.1007/s00217-018-3171-8)

[117] Papalexandratou, Z., Kaasik, K., Villagra Kauffmann, L., Skorstengaard, A., Bouillon, G., Espensen, J. L., Hansen, L. H., Jakobsen, R. R., Blennow, A. *et al.* (2019) Linking cocoa varietals and microbial diversity of Nicaraguan fine cocoa bean fermentations and their impact on final cocoa quality appreciation. *International Journal of Food Microbiology*, **304**, 106–118. (doi:10.1016/j.ijfoodmicro.2019.05.012)

[118] De Vuyst, L., Lefeber, T., Papalexandratou, Z. & Camu, N. (2010) The Functional Role of Lactic Acid Bacteria in Cocoa Bean Fermentation. In *Biotechnology of Lactic Acid Bacteria* (eds F. Mozzi, R. R. Raya & G. M. Vignolo), chap. 17, pp. 301–325. Chichester: Wiley Blackwell. (doi:10.1002/9780813820866.ch17)

[119] Viesser, J. A., de Melo Pereira, G. V., de Carvalho Neto, D. P., de S. Vandenberghe, L. P., Azevedo, V., Brenig, B., Rogez, H., Góes-Neto, A. & Soccol, C. R. (2020) Exploring the contribution of fructophilic lactic acid bacteria to cocoa beans fermentation: Isolation, selection and evaluation. *Food Research International*, **136**, 109 478. (doi:10.1016/j.foodres.2020.109478)

[120] Papalexandratou, Z., Camu, N., Falony, G. & De Vuyst, L. (2011) Comparison of the bacterial species diversity of spontaneous cocoa bean fermentations carried out at selected farms in Ivory Coast and Brazil. *Food Microbiology*, **28**(5), 964–973. (doi:10.1016/j.fm.2011.01.010)

[121] R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[122] Stan Development Team (2020) RStan: the R interface to Stan. R package version 2.21.1.

[123] Goodpaster, A. M. & Kennedy, M. A. (2011) Quantification and statistical significance analysis of group separation in NMR-based metabonomics studies. *Chemometrics and Intelligent Laboratory Systems*, **109**(2), 162–170. (doi:10.1016/j.chemolab.2011.08.009)

[124] Mahalanobis, P. C. (1936) On the Generalised Distance in Statistics. *Proceedings of the National Institute of Sciences of India*, pp. 49–55.

[125] McFerrin, L. (2013) HDMD: Statistical Analysis Tools for High Dimension Molecular Data (HDMD). R package version 1.2.

[126] Lefeber, T., Papalexandratou, Z., Gobert, W., Camu, N. & De Vuyst, L. (2012) On-farm implementation of a starter culture for improved cocoa bean fermentation and its influence on the flavour of chocolates produced thereof. *Food Microbiology*, **30**(2), 379–392. (doi:10.1016/j.fm.2011.12.021)

[127] Bastos, V. S., Santos, M. F., Gomes, L. P., Leite, A. M., Flosi Paschoalin, V. M. & Del Aguila, E. M. (2018) Analysis of the cocobiota and metabolites of Moniliophthora perniciosa-resistant *Theobroma cacao* beans during spontaneous fermentation in southern Brazil. *Journal of the Science of Food and Agriculture*, **98**(13), 4963–4970. (doi:10.1002/jsfa.9029)

[128] Racine, K. C., Lee, A. H., Wiersema, B. D., Huang, H., Lambert, J. D., Stewart, A. C. & Neilson, A. P. (2019) Development and Characterization of a Pilot-Scale Model Cocoa Fermentation System Suitable for Studying the Impact of Fermentation on Putative Bioactive Compounds and Bioactivity of Cocoa. *Foods*, **8**(3). (doi:10.3390/foods8030102)

[129] Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C. *et al.* (2020) Bayesian Workflow. Preprint.

[130] Caligiani, A., Palla, L., Acquotti, D., Marseglia, A. & Palla, G. (2014) Application of 1H NMR for the characterisation of cocoa beans of different geographical origins and fermentation levels. *Food Chemistry*, **157**, 94–99. (doi:10.1016/j.foodchem.2014.01.116)

[131] Megías-Pérez, R., Grimbs, S., D'Souza, R. N., Bernaert, H. & Kuhnert, N. (2018) Profiling, quantification and classification of cocoa beans based on chemometric analysis of carbohydrates using hydrophilic interaction liquid chromatography coupled to mass spectrometry. *Food Chemistry*, **258**, 284–294. (doi:10.1016/j.foodchem.2018.03.026)

[132] Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N. & Ullrich, M. S. (2018) Origin and varietal based proteomic and peptidomic fingerprinting of theobroma cacao in non-fermented and fermented cocoa beans. *Food Research International*, **111**, 137–147. (doi:10.1016/j.foodres.2018.05.010)

[133] Biehl, B., Passern, D. & Sagemann, W. (1982) Effect of acetic acid on subcellular structures of cocoa bean cotyledons. *Journal of the Science of Food and Agriculture*, **33**(11), 1101–1109. (doi:10.1002/jsfa.2740331107)

[134] Rosso, L., Lobry, J. R., Bajard, S. & Flandrois, J. P. (1995) Convenient Model To Describe the Combined Effects of Temperature and pH on Microbial Growth. *Applied and environmental microbiology*, **61**(2), 610–616. (doi:10.1128/AEM.61.2.610-616.1995)

[135] Boulton, R. (1980) The Prediction of Fermentation Behavior by a Kinetic Model. *American Journal of Enology and Viticulture*, **31**(1), 40–45.

[136] Peleg, M., Normand, M. D. & Corradini, M. G. (2012) The Arrhenius Equation Revisited. *Critical Reviews in Food Science and Nutrition*, **52**(9), 830–851. (doi:10.1080/10408398.2012.667460)

[137] Yalçin, S. K. & Özbaş, Z. Y. (2005) Determination of Growth and Glycerol Production Kinetics of a Wine Yeast Strain *Saccharomyces cerevisiae* Kalecik 1 in Different Substrate Media. *World Journal of Microbiology and Biotechnology*, **21**(6), 1303–1310. (doi:10.1007/s11274-005-2634-9)

[138] Yalçin, S. K. & Özbaş, Z. Y. (2004) Effects of different substrates on growth and glycerol production kinetics of a wine yeast strain *Saccharomyces cerevisiae* Narince 3. *Process Biochemistry*, **39**(10), 1285–1291. (doi:10.1016/S0032-9592(03)00252-8)

[139] de Oliveira, G. H. H., Corrêa, P. C., de Souza Santos, E., Treto, P. C. & Diniz, M. D. M. S. (2011) Evaluation of thermodynamic properties using GAB model to describe the desorption process of cocoa beans. *International Journal of Food Science & Technology*, **46**(10), 2077–2084. (doi:10.1111/j.1365-2621.2011.02719.x)

[140] Teh, Q. T. M., Tan, G. L. Y., Loo, S. M., Azhar, F. Z., Menon, A. S. & Hii, C. L. (2016) The Drying Kinetics and Polyphenol Degradation of Cocoa Beans. *Journal of Food Process Engineering*, **39**(5), 484–491. (doi:10.1111/jfpe.12239)

[141] Barreiro, J. A. & Sandoval, A. J. (2020) Kinetics of moisture adsorption during simulated storage of whole dry cocoa beans at various relative humidities. *Journal of Food Engineering*, **273**, 109 869. (doi:10.1016/j.jfoodeng.2019.109869)

[142] Pelicaen, R., Gonze, D., De Vuyst, L. & Weckx, S. (2020) Genome-scale metabolic modeling of *Acetobacter pasteurianus* 386B reveals its metabolic adaptation to cocoa fermentation conditions. *Food Microbiology*, **92**, 103 597. (doi:10.1016/j.fm.2020.103597)

[143] Miller, K. V. & Block, D. E. (2020) A review of wine fermentation process modeling. *Journal of Food Engineering*, **273**, 109 783. (doi:10.1016/j.jfoodeng.2019.109783)

[144] van Boekel, M. (2020) On the pros and cons of Bayesian kinetic modeling in food science. *Trends in Food Science & Technology*, **99**, 181–193. (doi:10.1016/j.tifs.2020.02.027)

[145] Pillonetto, G., Sparacino, G. & Cobelli, C. (2003) Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of Bayesian estimation. *Mathematical Biosciences*, **184**(1), 53–67. (doi:10.1016/S0025-5564(03)00044-0)

[146] Hines, K. (2015) A Primer on Bayesian Inference for Biophysical Systems. *Biophysical Journal*, **108**(9), 2103–2113. (doi:10.1016/j.bpj.2015.03.042)

[147] Kouamé, C., Loiseau, G., Grabulos, J., Boulanger, R. & Mestres, C. (2021) Development of a model for the alcoholic fermentation of cocoa beans by a *Saccharomyces cerevisiae* strain. *International Journal of Food Microbiology*, **337**, 108 917. (doi:10.1016/j.ijfoodmicro.2020.108917)

[148] Weber, S. (2020) New Within-Chain Parallelisation in Stan 2.23: This One's Easy for Everyone! [Online]. Available from: `https://statmodeling.stat.columbia.edu/2020/05/05/easy-within-chain-parallelisation-in-stan/`. [Accessed 15th April 2021].

[149] de Andrade Silva, C. A., Oka, M. L. & Fonseca, G. G. (2019) Physiology of yeast strains isolated from Brazilian biomes in a minimal medium using fructose as the sole carbon source reveals potential biotechnological applications. *3 Biotech*, **9**(5), 191. (doi:10.1007/s13205-019-1721-9)

[150] Vrancken, G., Rimaux, T., De Vuyst, L. & Leroy, F. (2008) Kinetic analysis of growth and sugar consumption by *Lactobacillus fermentum* IMDO 130101 reveals adaptation to the acidic sourdough ecosystem. *International Journal of Food Microbiology*, **128**(1), 58–66. (doi:10.1016/j.ijfoodmicro.2008.08.001)

[151] Gustaw, K., Michalak, M., Polak-Berecka, M. & Waśko, A. (2018) Isolation and characterization of a new fructophilic *Lactobacillus plantarum* FPL strain from honeydew. *Annals of Microbiology*, **68**(7), 459–470. (doi:10.1007/s13213-018-1350-2)