Modelling and Analysis of Ecological Interactions in Microbial Community based Bioreactors



Sihao Di

Supervisor: Professor Aidong Yang

Department of Engineering Science

St Anne's College

University of Oxford

Acknowledgements

The last four years of my life have been a great journey of learning, understanding, making mistakes and self-recognizing, at the Department of Engineering Science, University of Oxford. It has been a great period in which I explored modelling and scientific computing in chemical engineering and grasped skills and knowledge in this area. The more I learned, the more I understood that there is so much I still don't know. Thankfully I was so lucky in having around me great people who have advised me, guided me and provided me tremendous help.

First of all, I would like to express my sincere gratitude and special appreciation to my supervisor, Professor Aidong Yang, for his enthusiasm, commitment, patience and great support, motivating me growing and developing my research skills and guiding me in my most difficult times. I still remember the first time when we communicated by email, 'Thanks for your interest in our group and I would encourage you to submit an application'. The first encouragement from him, which thoroughly changed my life and enabled me to know such a great person.

Besides my supervisor, I would like to thank my family for all the financial and emotional support. I wish to thank my college and college members, especially Hang Yuan and Zicong Hu who helped me on my programming skills during my first year. Finally, I wish to thank all my fellow researchers and friends whose friendship I truly value.

Abstract

Microbial communities refer to a number of microorganisms living together. To-date, a number of naturally occurring microbial communities are used for engineering purposes, e.g. anaerobic digestion, and the construction of artificial communities opens the door to wider applications. In microbial communities, the ecological interactions between species (e.g. competition, predation and mutualism) play important roles. The purpose of this thesis is to explore (1) the implications of such interactions in a bioreactor in terms of productivity and stability, and (2) the inference of such interactions particularly in a complex microbial community in order to form a basis for further manipulation such as optimisation and control.

This work firstly studies the cases involving different microbial species at the same trophic level, with a particular focus on the potential of overyielding (the collective yield of multiple microalgae species, in terms of overall biomass abundance or a specific biomass component, is greater than that of any single species grown in the same system) in the poly-culture of microalgae arising from the complementary use of a key resource, namely light over the spectrum range of photosynthetically active radiation. Using a mathematical model, it is demonstrated that complementarity of light absorption alone does not guarantee overyielding; other factors such as the operational settings (e.g. loss rate of the bioreactor) and species traits (complementarity of light absorption spectra) can also affect the possibility and the degree of overyielding and even the co-existence of multiple species.

Moving onto microbial communities with species at different trophic levels which therefore involve cross-feeding links, this work models the co-culture of *C. cellulovorans* and *R.palustris* for hydrogen production from cellulose as a representative case, followed by a more generalised study on theoretically constructed 2-species and 3-species systems with cross-feeding and multiple types of inhibition links. Both dynamic simulation and bifurcation analysis are used in the investigation. These studies reveals potential trade-off between productivity and stability of such systems, the occurrence of oscillatory behaviour due to feedback loops, and the negative impact of inhibition on both productivity and stability. Furthermore, the possibility of predicting the dynamics of a complex microbial communities from that of a simpler community is also analysed.

The final part of the work focuses on microbial communities with a large number of species, where a key challenge is the inference of the inter-species interactions from abundance data. Recognizing the fact that most microbial interactions are mediated by chemicals, this work systematically evaluates an existing network inference model (generalized Lotka Volterra, gLV) in comparison with a new model proposed in this study (chemical mediated interactions, CMI). The results show that CMI has the potential to infer multiple interactions in a microbial community, which however suffers from the accumulation of errors when systems become complex. In comparison, gLV demonstrates a stable performance in most cases.

Overall, this work has offered insights about the productivity and stability of microbial communities which are potentially useful in guiding the design and control

of bioreactors, and produced evidence of the effectiveness of existing and new models for the inference of structures of complex microbial communities which provides a basis for future work on chemical-mediated microbial networks.

Table of Contents

Chapter 1. Introduction 15

1.1. Background 15

1.2 Project aim and thesis structure 18

Chapter 2. Literature review 24

2.1. Studies on the dynamics and performance of microbial communities 24

2.1.1. Competition and complementarity resources utilisation 25

2.1.1.1 Competition 25

2.1.1.2. Complementary light use for microalgal overyielding potential 26

2.1.2. Trophic links based microbial communities 31

2.1.2.1. Anaerobic digestion 31

2.1.2.2 H_2 production by dark and photo fermentative bacteria 33

2.1.2.3 Other cases 35

2.1.3. Generalized studies on ecological interactions in microbial communities 36

2.2. Networks and network construction for microbial communities 39

2.2.1 Empirical studies of networks 39

2.2.1.1 Social networks 39

2.2.1.2 Technological networks 40

2.2.1.3 Biological networks 40

2.2.2. Structural analysis of networks 41

2.2.2.1 Centrality 41

2.2.2.2 Detection of communities 42

2.2.2.3 Employing structural information for control 43

2.2.3. Microbial network construction from species abundance data 44

2.2.3.1 Undirected networks 45

2.2.3.2 Directed networks 47

2.3. Bioreactor design 53

2.4. Research Gaps 55

Chapter 3. Overyielding potential of microalgal polyculture with complementary light absorption spectra: A model-based analysis 72

3.1. Introduction 75

3.2. Methods 77

3.2.1 Modelling approach 77

3.2.2. Investigating a two-species community with simplified light absorption spectra 79

3.2.3 Investigating a community of real species 84

3.3 Results and discussion 86

3.3.1 Impact of operation settings 86

3.3.2 Impact of species traits 92

3.3.3 Simulation of a community of two species with real spectra 95

Chapter 4. Modelling and analysis of microbial community based bioreactors for hydrogen production 102

4.1. Introduction 106

4.2. Modelling kinetics of individual species 109

4.3. Model parameterisation and batch bioreactor simulation 113

4.3.1. The two-species system 113

4.3.2. The three-species system 116

4.4. Simulation and analysis of CSTR 121

4.4.1. Simulation and analysis of the two-species system 122

4.4.2. Simulation and analysis of the three-species system 128

4.5. Discussion 133

4.6. Conclusion 136

Chapter 5. Analysis of productivity and stability of synthetic microbial communities 144

5.1. Introduction 147

5.2. Methods 151

5.2.1. Construction of representative systems 151

5.2.2. Mathematical models 157

5.2.3. Quantification of stability and productivity 159

5.3. Results 161

5.3.1. The base system: two species with no inhibition 161

5.3.2 Layer-1 cases: two species with one inhibition 164

5.3.3. Layer-2 cases: two-species with double inhibitions 167
5.3.4 Layer-3 cases: addition of a third species 173
5.4. Discussion 177
5.5. Conclusions 181
Chapter 6. Ecological network inference in microbial community with multiple inter-species interactions through chemicals 200
6.1. Introduction 204
6.2. Methodology 210
6.2.1. Inference models 210
6.2.1.1. Generalized Lotka Volterra (gLV) model 210
6.2.1.2. Chemical-mediated interaction (CMI) model 211
6.2.1.2.1. The model equations 211
6.2.1.2.2. Microbial species interactions (M-M) from interactions between microbial and chemical species (M-C) 213
6.2.2. Sparse Solution and Lasso regression 215
6.2.3. Data generation from numerical simulation 216
6.2.3.1. Anaerobic digestion: inference of a food web 216
6.2.3.2. Artificial microbial communities 217
6.2.3.2.1. The mathematical model 218
6.2.3.2.2. Generated networks 219
6.2.3.3. Data generation: perturbation strategy and measurement noise 223
6.2.4. Evaluation of the inference results 224
6.3. Results and discussion 226
6.3.1. The AD System 226
6.3.1.1 Performance of the gLV model 227
6.3.1.2 Performance of CMI 230
6.3.1.3 Impact of the selection of chemicals for perturbation 234
6.3.1.4. Key results summary 236
6.3.2. Artificially generated networks 237
6.3.2.1. 2-layer network 238

6.3.2.1.1 Results of the gLV model 238

6.3.2.1.2 Results of the CMI model 241

6.3.2.2. 3-layer network 246

6.3.2.2.1. Performance of gLV with rolling perturbation on inlet substrates 246

6.3.2.2.2. Analysis of the inference performance when dealing with a "deeper" network 247

6.3.2.2.3. Inference of multiple M-M interactions 252

6.3.2.3. Large scale networks and impact of sparsity 254

6.3.2.4 Summary of key results 258

6.4. Conclusion 260

Chapter 7. Conclusions and future work 273

7.1. Summary of research and key conclusions 273

7.2. Limitations 276

7.3. Future work 278

7.3.1 Flux balance analysis based models to predict phenotypes of microorganisms ... 278

7.3.2 Linking data richness with statistical learning models for inferring microbial community networks 280

7.3.3 Applications of inferred microbial community networks 280

List of Figures:

Figure 1. 1. Some common interactions between microbial species mediated by substrates/metabolites. *16*
Figure 2. 1. Spurious Correlation. *46*
Figure 3. 1. Typical relationship between photosynthetic rate and light intensity. *78*
Figure 3. 2. Illustration of simplified light absorption spectra of species a (upper plot) and b (lower plot). *83*
Figure 3. 3. Light absorption spectra of two species [12]. BS4 and BS5 represent (blue) picocyanobacteria and (red) picocyanobacteria, respectively. *85*
Figure 3. 4. Predicted overyielding level as a function of light absorption complementarity at different loss rates. On the top left graph, the loss rate is, from the bottom to the top, 0.02, 0.05, 0.1 and 0.2 day^{-1}. On the other two graphs, the loss rate is, from the bottom to the top, 0.2, 0.4, 0.6 and 0.8 day^{-1}. *88*
Figure 3. 5. Light factor predicted at different cells concentrations, affected by light absorption complementarity, loss rate and cultural type. *89*
Figure 3. 6. Predicted changes of light intensity (a), light absorption (b), saturation/inhibition effect (c) and light factor (d) at different depths of the reactor. Dashed and solid lines represent monoculture and co-culture, respectively. *91*
Figure 3. 7. Predicted overyielding level changes in co-culture of species with different light use efficiencies. The incident light intensities are 40, 400, 800 μmol photons s^{-1}/m^2 s for the 1st, 2nd and 3rd rows, respectively. ϕ_a for the 1st, 2nd and 3rd columns are 4.2×10^{10}, 4.5×10^{10} and 5×10^{10} respectively. The loss rates are 0.02, 0.05, 0.1 and 0.2 day^{-1} for curves in each plot of the first row, from bottom to top, respectively, and 0.2, 0.4, 0.6 and 0.8 day^{-1} for the other two rows. *94*
Figure 3. 8. Minimum level of light absorption complementarity required for co-existence. (a) assumes loss rate = 0.2 day^{-1}. (b) assumes $\phi_a = 4.2 \times 1010$. *95*
Figure 3. 9. Simulation results of the cultivation of BS4 and BS5. ..... *97*

Figure 4. 1. Interactions between C. cellulovorans and R. palustris. *114*
Figure 4. 2. Comparison of experimental data (Lu and Lee, 2015) and numerical simulation results. Solid line and "" represent simulation results and experimental data, respectively. a, b: mono-culture; c, d: co-culture.* *116*
Figure 4. 3. Interactions between C. cellulovorans, R. palustris and HEC. *117*
Figure 4. 4. Comparison of hydrogen production, VFAs concentration and cells concentration for two-species and three-species system at initial cellulose concentration 3g/l. Ac, Bu, Fc and La represent acetate, butyrate, formate and lactate, respectively. Cc and Rp represent C. cellulovorans and R. palustris, respectively. *120*
Figure 4. 5. Bifurcation diagrams for the two-species CSTR system. Solid line: stable branch; dash line: unstable branch; '': static limit point (marked as "LP").* *123*

Figure 4. 6. Evolvement of the two-species CSTR system to two different stable steady states. Cc and Rp represent C. cellulovorans and R. palustris. The upper two plots show the transition to a C. cellulovorans only state; the lower two plots show the transition to a co-existence state. HRT=60 hours; ICC = 2.2 g/l. *124*
Figure 4. 7. pH inhibition curves of R. palustris and C. cellulovorans in the original model, variation 1 and variation 2. ..... *125*
Figure 4. 8. Continuity diagram for the two-species CSTR system in variation 1; HRT = 60hrs. *126*
Figure 4. 9. Continuity diagram for the two-species CSTR system in variation 2. Solid line: stable branch; dash line: unstable branch; '': static limit point; 'o': Hopf bifurcation (marked as "HB").* *127*
Figure 4. 10. Oscillatory behaviour for the two-species CSTR system, variation 2. HRT = 39 hrs; ICC = 9 g/l. *127*
Figure 4. 11. Continuity diagram for the three-species CSTR system in the original model. Solid line: stable branch; dash line: unstable branch; '': static limit point (marked as "LP"). HRT = 60 hrs.* . *129*
Figure 4. 12. Continuity diagram for the three-species CSTR system with improved complementarity in VFAs utilization. Solid line: stable branch; dash line: unstable branch; '': static limit point (marked as "LP"), HRT = 60 hrs.* *130*
Figure 4. 13. Continuity diagram for the three-species CSTR system with altered pH inhibition effect. Solid line: stable branch; dash line: unstable branch; '': static limit point (marked as "LP"); 'o': Hopf bifurcation (marked as "HB"). HRT = 39 hrs.* *132*
Figure 4. 14. Oscillatory behaviour of the three-species CSTR system, with altered pH inhibition effect. HRT = 39 hrs; ICC = 9 g/l. *132*

Figure 5. 1. The base system. *151*
Figure 5. 2. Two-species systems with a single inhibition (Layer-1 cases). *152*
Figure 5. 3. Two-species systems with double inhibitions (Layer-2 cases). *154*
Figure 5 4. Construction of the two cases of three-species system (layer-3 cases). *155*
Figure 5. 5. Productivity (a) and stability (b) of the base system under different parameter settings. *163*
Figure 5. 6. Results for the layer-1 cases (θ_1=3, θ_2=3, α_1=1). *164*
Figure 5. 7. Oscillation in case 1-2. ϕ = 5, θ_1 = 3, θ_2 = 3 and α_1 = 1. a. Productivity, solid line represents stable steady state; dashed line represents oscillation region; '' represents onset of stable oscillation. b. Stability at stable steady state. c. Numerical simulation of the stable oscillation, $s2, in = 6$.* *167*
Figure 5. 8 Results of layer-2 cases in comparison with corresponding layer-1 cases. *170*
Figure 5. 9. Oscillation in case 2-5. ϕ_2 = 2, ϕ_4 = 0.2. (a) Productivity, solid line represents stable steady state, dashed line represents oscillation region, '' represents onset of stable oscillation. (b) Stability at stable steady state. (c) Numerical simulation of the stable oscillation, $s2, in$=8.* *172*
Figure 5. 10. Three-species system, case 3-1. Solid line represents three species systems with different parameter settings; dashed line represents the base system. The dotted line corresponds to a dimensionless strength of 0.2 of the new inhibition; the solid line corresponds to a strength of 1. ..... *174*

Figure 5. 11. Occurrence of stable oscillation in three-species system, case 3-1. The dimensionless strength of the new inhibition is 6. (a) Productivity, solid line represents stable steady state, dashed line represents oscillation region, '' represents onset of stable oscillation. (b) Stability at stable steady state. (c) Numerical simulation of the stable oscillation, $s2, in = 6$.* *175*
Figure 5. 12. Three-species system, case 3-2. (a) Productivity; (b) Stability. *175*

Figure 6. 1. Illustration of deriving M-M interactions from M-C interactions. *215*
Figure 6. 2. The AD networks (food web only) adapted from ADM1. ..... *217*
Figure 6. 3. Examples of the generated networks. *222*
Figure 6. 4. Illustration of perturbation strategy *224*
Figure 6. 5. An illustrative ROC curve. *225*
Figure 6. 6. ROC curves for network inference using the gLV model. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance between different runs is not significant, thus the error bar is not shown. *229*
Figure 6. 7. ROC curves for M-M network inference using the CMI model. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance of different runs is not significant, thus the error bar is not shown. *230*
Figure 6. 8. ROC curves for CMI model M-C network inference. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance of different runs is not significant, thus the error bar is not shown. In the left-hand-side figures, dash line represents production and solid line represents consumption; in the right-hand-side figures, dash line with circle represents production and solid line represents consumption. *233*
Figure 6. 9. How consumption inference is better than production. *234*
Figure 6. 10. ROC curve for inference of the 2-layer network by the gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation". *238*
Figure 6. 11. Impact of λ on inferring inhibition by gLV (data: network instance 1a; Noise level = 0.05, Inhibition level = 0.01, perturbation level = 50). Left: ROC curve for food web and inhibition inference; Right: relations between λ and inferred frequency of positive and negative elements. *240*
Figure 6. 12. ROC curve for inference of the 2-layer network by the CMI model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation". *242*
Figure 6. 13. ROC curve for inference of the 2-layer network M-C interactions by the CMI model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation". *243*
Figure 6. 14. Errors new model makes in consumption inference. *244*
Figure 6. 15. Potential for CMI to infer multiple M-M interactions in a 2-layer network. Results are from testing network 1a and averaged with 5 runs with different actual perturbations generated from the same perturbation level. The error bars represent "mean +/- standard deviation". *245*

Figure 6. 16. ROC curve for gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation". 247
Figure 6. 17. Comparison between inferred and true 3-layer networks by gLV (with rolling perturbation on only inlet substrates and inlet and intermediate substrates). Network instance 2a was used as an example.....250
Figure 6. 18. Network inference performance with network 2a using gLV and different perturbation strategies. Each result is averaged with 5 runs with different actual perturbations generated from the same perturbation level distribution. The error bars represent "mean +/- standard deviation". 251
Figure 6. 19. Potential for CMI to infer multiple M-M interactions in a 3-layer network. Results are from testing network 2e and averaged with 5 runs with different actual perturbations generated from same perturbation level. The error bars represent "mean +/- standard deviation".....253
Figure 6. 20. Network inference of large-scale dense and sparse networks by the gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation".....255
Figure 6. 21. Comparison between inferred and true 3-layer networks by gLV (with rolling perturbation on only inlet substrates and inlet and intermediate substrates). Network instance 3e is used as the example.....257
Figure 6. 22. Network inference performance with network 3e using gLV and different perturbation strategies. Results are from testing network 3e and averaged with 5 runs with different perturbations generated from same perturbation level. The error bars represent "mean +/- standard deviation". .258

List of Tables:

Table 3. 1. Default parameters used in simplified absorption spectra. *82*
Table 3. 2. Parameter values adopted in simulation of real species. *96*

Table 4. 1. Model parameters for C. cellulovorans and R. palustris. *115*
Table 4. 2. Model parameters for the third species (HEC). *118*

Table 5. 1. Literature examples of engineered cross-feeding microbial consortia *150*
Table 5. 2. Relations between layer 1 and layer 2 models. *156*
Table 5. 3. Sets of parameters tested for the base system. *163*
Table 5. 4 Parameter values adopted in the layer-2 cases. *170*

Table 6. 1. Summary of existing network inference models or tools. *208*
Table 6. 2. Parameter values for the artificial microbial community model. *219*
Table 6. 3. Properties of artificially generated networks. *220*
Table 6. 4. Perturbation strategy and noise level for the AD case. *227*
Table 6. 5. Key results summary. Performance is marked from A to E, representing network inference accuracy from the best to the worst. *236*
Table 6. 6. Parameters used in the 2-layer and the 3-layer networks. *237*
Table 6 7 Key results summary. F: foodweb, I: inhibition. Performance is marked from A to E representing network inference accuracy from the best to the worst. *259*

Chapter 1. Introduction

1.1. Background

Communities of microorganisms have been used for engineering purposes both historically and more recently, for example in anaerobic digestion for treating waste organic streams [1] [2], bio-hydrogen production using synthetic microbial communities [3] [4] and cultivation of multiple microalgal species for bio-fuel production [5] [6]. Such systems received special attentions for higher yield and better stability compared to monoculture based systems.

To-date, microbial community based applications have been developed in different areas, including bioprocess engineering, environmental remediation and biomedicine [5] [6] [7]. In a microbial community, the ecological interactions between species (e.g. competition, predation and mutualism) and between species and environmental conditions play a very important role [8]. Therefore, an in-depth understanding and carefully designed manipulation of these interactions have the potential to contribute to further improvement of both the productivity and the stability of community-based bioreactors. However, work on systematic exploration of this potential has been rather limited.

Different types of relationship exist between microbial species. Figure 1.1 shows some commonly occurring patterns for a pair of microbial species to interact with each other, mediated by a chemical species which could be a substrate or a metabolite. A

food chain is formed by microbial species B living off the “waste” of microbial species A. Competition arises when two microbial species consume the same substrate. On the other hand, cooperation can emerge when each of the microbial species gains from the metabolic reactions of the other. When a microbial species produces a metabolite that inhibits another species, leading to inhibition. Two microbial species may interact with one or multiple types of relationship. As an example of multiple types of interactions, one microbial species may produce a product that feeds another species, which however inhibits the latter at the same time [9].

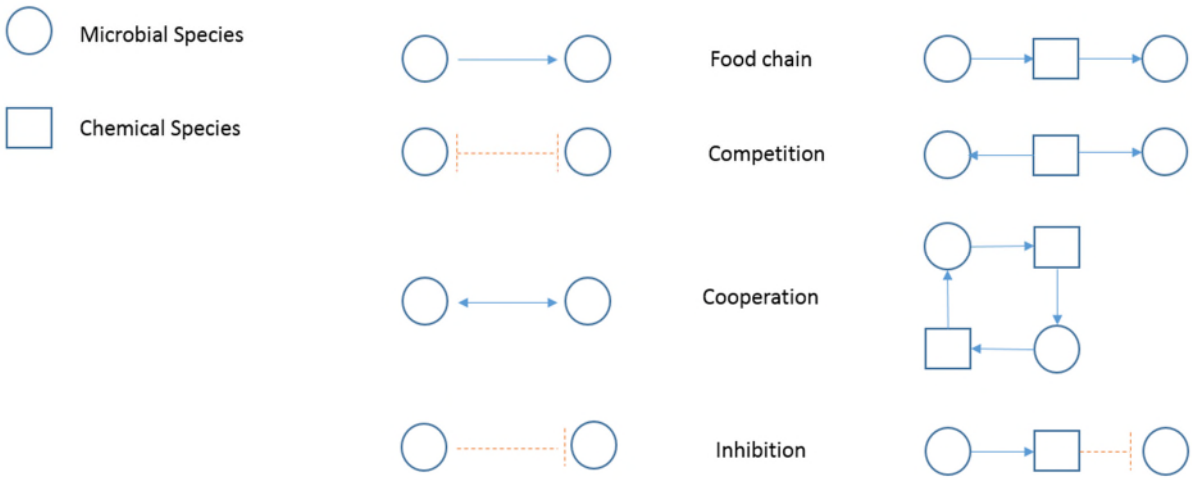


Figure 1. 1. Some common interactions between microbial species mediated by substrates/metabolites.

In general, an engineering system is expected to produce desired products efficiently in a stable manner, i.e. to meet the requirements on both productivity and stability. Productivity represents the output of the desired product, often measured by the amount of product obtained within a given time period. Stability, on the other

hand, represents the ability for a system to maintain or recover to an acceptable level of productivity when exposed to disturbance. One would expect that productivity and stability are inter-connected as both are the results of the structure and the dynamics of the underlying system. In a microbial community, an important aspect of stability is the co-existence of its member species, which in turn could significantly affect productivity. Therefore, understanding the interplay between productivity and stability is an important issue for studying microbial community based bioreactors.

Microbial communities can be either naturally occurring or synthetic. In a synthetic microbial community, there are normally a small number of microbial species which are well defined so that the traits of individual species and hence the nature of their interactions are known. In terms of modelling tools, this work makes use of dynamical simulation [10] and bifurcation analysis [11] to study such synthetic communities. Dynamic simulation is able to obtain temporal profiles of a change of the abundance of microbial species. Bifurcation analysis is another type of tool to gain insights in steady state properties (or more generally, properties of attractors) of a microbial community, as a function of a system parameter, e.g. inlet substrate concentration.

In a naturally occurring microbial community, the number of species tends to be much larger, and it is difficult for all species to be clearly defined; typical cases are anaerobic digestion and human gut communities. For modelling and analysis, networks are always used to represent and visualize such a microbial community, with nodes representing the microbial or chemical species, and edges representing pairwise

interactions [12]. Using the notion of networks, the development of tools that are able to capture the community structure becomes key to gaining insights on a complex microbial community. With an accurately constructed network, one would be able to learn about the community by e.g. identifying sub-communities and important species [14]. Furthermore, several techniques have been recently developed to control a complex network only based on its topological structure, which has the potential to be applied to microbial community based engineering systems [15] [16].

1.2 Project aim and thesis structure

The overall purpose of the thesis is to develop and apply mathematical modelling approaches for systematic design and analysis of microbial community based engineering systems. More specifically, the project is aimed to explore the implication of important and common inter-species relations on productivity and stability of bioreactors based on synthetic microbial communities, and to develop and evaluate approaches for inferring network structures of large-scale (naturally occurring) microbial communities with potential engineering or medical applications. The project aim is achieved by a combination of specific case studies and investigation of generalised systems. The outline of the thesis is as follows:

- Chapter 1 describes briefly the background and motivation of the thesis.
- Chapter 2 presents a literature review. The review is divided into 2 parts, the first part is relevant research on model based analysis of dynamics and

performance of microbial communities. The second part involves research on large scale microbial communities, particularly based on constructing and analysing networks.

- Chapter 3 deals with a poly-culture microalgal system, in which we focus on the complementary utilization of resources (particularly light segments) between different species, as our first case study. The hypothesis that complementary use of light segments of different wavelength ranges leads to overyielding is tested by mathematical modelling, which reveals key factors that affect the validity of the hypothesis.
- Chapter 4 reports another case study on microbial community systems for producing hydrogen from cellulose, involving multiple microbial interactions, including facilitation, inhibition and competition. With dynamic modelling and bifurcation analysis, it offers a first attempt to understand the interplay between productivity and stability under the influence of these inter-species relationships.
- Chapter 5 offers a generalization of the specific case study by chapter 4 by numerically modelling and analysing a set of two or three species communities with chosen structures that involve various combinations of typical inter-species relationships, to derive a set of general observations and guidance for designing synthetic communities.
- Chapter 6 moves on to investigate systems that involve a large number of species which are not necessarily clearly identified individually, as in the cases of naturally occurring communities. It uses data generalised by numerical

simulation of the adapted anaerobic digestion system and of a set of artificially composed microbial networks. A systematic evaluation of an existing model for inferring microbial networks is carried out, and a new model that explicitly captures microbe-chemical connections is proposed and systematically assessed against the existing model.

- Chapter 7 provides the overall conclusion and describes the possible future research directions based on what has been achieved by this thesis.

It should be noted that each of the research chapters (3-6) comprises a published article or a manuscript to be submitted for publication. The numbering of sections, figures, tables and equations in the article or manuscript has been re-arranged to be consistent with the chapter structures of the thesis.

References:

[1]. Wade, M. J., Pattinson, R. W., Parker, N. G., & Dolfing, J. (2016). Emergent behaviour in a chlorophenol-mineralising three-tiered microbial 'food web'. Journal of theoretical biology, 389, 171-186.

[2]. Weedermann, M., Wolkowicz, G. S., & Sasara, J. (2015). Optimal biogas production in a model for anaerobic digestion. Nonlinear Dynamics, 81(3), 1097-1112.

[3]. Lu, H., & Lee, P. K. (2015). Effects of cellulose concentrations on the syntrophic interactions between *Clostridium cellulovorans* 743B and *Rhodopseudomonas palustris* CGA009 in coculture fermentation for biohydrogen production. International Journal of Hydrogen Energy, 40(35), 11800-11808.

[4]. Argun, H., & Kargi, F. (2011). Bio-hydrogen production by different operational modes of dark and photo-fermentation: an overview. International Journal of Hydrogen Energy, 36(13), 7443-7459.

[5]. Pulz, O., & Gross, W. (2004). Valuable products from biotechnology of microalgae. Applied microbiology and biotechnology, 65(6), 635-648.

[6]. Zhou, W., Chen, P., Min, M., Ma, X., Wang, J., Griffith, R., ... & Shi, J. (2014). Environment-enhancing algal biofuel production using wastewaters. Renewable and sustainable energy reviews, 36, 256-269.

[7]. Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C. H., Ahn, J. H., Foster, C. E., ... & Lin, X. N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proceedings of the National Academy of Sciences, 110(36), 14592-14597.

[8]. Dolinšek, J., Goldschmidt, F., & Johnson, D. R. (2016). Synthetic microbial ecology and the dynamic interplay between microbial genotypes. FEMS microbiology reviews, 40(6), 961-979.

[9]. Großkopf, T., & Soyer, O. S. (2014). Synthetic microbial communities. Current opinion in microbiology, 18, 72-77.

[10]. Kambam, P. K. R., Henson, M. A., & Sun, L. (2008). Design and mathematical modelling of a synthetic symbiotic ecosystem. IET systems biology, 2(1), 33-38.

[11]. Ajbar, A., & Alhumazi, K. (2000). Microbial competition: study of global branching phenomena. AIChE journal, 46(2), 321-334.

[12]. Newman, M. (2010). Networks: an introduction. Oxford university press.

[13]. Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582.

[14]. Newman, M. E. (2016). Mathematics of networks. The new Palgrave dictionary of economics, 1-8.

[15]. Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2011). Controllability of complex networks. Nature, 473(7346), 167.

[16]. Angulo, M. T., Moog, C. H., & Liu, Y. Y. (2019). A theoretical framework for controlling complex microbial communities. Nature communications, 10.

Chapter 2. Literature review

2.1. Studies on the dynamics and performance of microbial communities

Microbial communities considered for engineering purposes can be either naturally occurring (e.g. anaerobic digestion [1] [2]), or artificially constructed which are referred to as synthetic microbial communities [3]. Compared to the systems based on the monoculture of wild or genetically-engineered microorganisms, industrial applications of microbial communities could benefit from improved robustness [4] and/or efficiency [5]. Robustness is the ability to survive perturbations, which may refer to stability or resilience, such as the improvement demonstrated by microbial communities in biofilms [6]. Efficiency, on the other hand, is associated to the design purpose, which could be, for example, hydrogen yield in bio-hydrogen production [7]. To design or improve microbial community systems for engineering purposes, an in-depth understanding of the ecological interactions between species (e.g. competition, predation and mutualism) and between species and environmental conditions (e.g. inhibition) is essential [8]. The following review will first consider microbial communities with species within the same trophic level, i.e. competition or complementary resources utilisation [9] [10] [11], followed by an introduction to microbial communities within different trophic levels, i.e. some species feeding others [1] [2]. Finally, a review on generalized studies will be presented.

2.1.1. Competition and complementarity resources utilisation

2.1.1.1 Competition

Existing literature has reported a number of investigations on the interactions of species in microbial communities, among which, competition has become an important topic especially when co-existence becomes a problem. A theoretical analysis was carried out by Tilman, 1980 [11]. In this work, it was demonstrated for multiple species competing for multiple types of resources, the major criterion for stable co-existence under steady-state operation is that each species consume relatively more of the resource that more limits its own growth rate, compared to the other resources. When two or more microbial species compete for a single rate-limiting substrate, they are not able to co-exist under steady-state operation. However, there are other specific operational settings which are able to make co-existence possible for such systems.

In another work [12], a numerical study of two microbial species competing for a single substrate with a periodically varying dilution rate was performed, and showed that co-existence of the competing microbial species under this operating regime was possible, usually in the form of limit cycle in accordance with the externally imposed variation. Similarly, a system with two microbial species competing for a single pollutant (phenol) was studied in a sequencing fed-batch reactor (SFBR) [13], with different operation regimes. It was found that the long-term behaviour could be no species surviving, only one species surviving, or both species surviving in steady cycles.

Model predictions in this study were tested experimentally which showed good agreement between predicted and measured concentrations of the substrate and the biomass of each species.

Besides operating with variation of the inlet concentration profile, coexistence of two microbial species competing for a single substrate can sometimes also be possible when other biological mechanisms exist. For example, consortium involving plasmid-bearing and plasmid-free cell populations competing for a single resource were shown to lead to co-existence by the loss of plasmid-bearing organisms or the modification of the recombinant DNA during reproduction [14]. Existing literature also investigated a three-species system competing on two resources [7]. It demonstrated spatial heterogeneity applied to the system, by considering configurations of interconnected ideal chemostats, made it possible for all species co-existence in such a system.

Competition occurs usually when different microbial species uptake same resources. In an engineering application, introducing multiple species to consume different resources, which are available and which otherwise cannot be fully utilised by a single species, can potentially lead to a higher resources consumption ratio, and hence higher productivity. Complementary light utilisation by microalgal co-culture is a typical case.

2.1.1.2. Complementary light use for microalgal overyielding potential

Experimental evidence for overyielding potential from complementary light utilisation

Microalgae are unicellular algae that have been considered for a range of industrial applications. Up to now, production of high value products from microalgae has been already commercially applied in industrial scale, for example, *Spirulina* biomass can be processed for many food products [15]. Microalgae can also be used as animal feed for aquaculture and biofertilizer. Besides, their pigments such as chlorophyll (which works as primary photosynthetic pigments) and others like β-Carotene can serve as food colouring agents and health products [15]. One of the promising applications for microalgal biotechnology is biofuel production [16] [17]. Acting like corn, soybean, wood and other plants, microalgae can convert solar energy to chemical energy that is stored in the form of oils, proteins and carbohydrates. Unlike other plants, microalgae grow much faster and exceed the yield of other oil production crops by 10-100 times. They have been considered as the only biofuel producer which has the potential to totally replace traditional fossil fuels [18].

Experiments were carried out and showed with an increase in microalgal species richness, the yield, measured by micolalgal biomass or specific bioproduct such as lipid, increased [19] [20]. Such increased productivity has been referred to as overyieldingThis effect is frequently observed in many ecosystems, not only in poly-culture of microalgae [10]. There are often two mechanisms proposed in literature to explain why an increase in species diversity can lead to an increase in productivity. The first mechanism is ‘sampling effect’, sometimes also called ‘selection effect’, which refers to the fact that a highly productive species in a number of possible species has a

higher likelihood of being present in a more diverse community, and hence increases the overall yield. The second mechanism is called ‘complementarity effect’, referring to the mechanism where a resource partitioning and/or facilitation among species leads to increased resource used and hence a higher productivity in more diverse communities [5]. In poly-culture of microalgae, an important mechanism for a ‘complementary effect’ is the complementary use of light at different wavelengths by different algal species [10].

Striebel et al. demonstrates an increase in microalgal taxon richness can lead to an increase of assembled communities, photosynthetically active radiation (PAR) absorbance, specific rate of net primary production and overall biomass [20]. The results suggested that the observed positive relationship between microalgal diversity and productivity was at least in part of a consequence of complementary use of light, while the selection effect was shown to be unrelated to the microalgal taxon richness from the experiment. Stockenreiter et al. carried out similar experiments. They claimed that there was no increase in microalgal lipid yield with species richness in polycultures consisting of members within a single functional group [19]. However, a clear increase in lipid production was observed in communities consisting of different functional groups. Similarly, light usage increased significantly with species richness in all poly-cultures that consisted of species from multiple functional groups. These findings further highlighted the importance of the diversity-productivity relationship in microalgal poly-culture. Behl S et al. conducted laboratory experiments with 85

assembled phytoplankton communities consisting of species from 4 functional groups to investigate the importance of species diversity on carbon uptake [21]. The analysis of similarities (ANOSIM) suggested that there were significant differences in pigment composition among algal groups, implying the possible existence of the complementarity effect.

Light effect on microalgal growth

Analysing the complementary light use by microalgae requires an understanding of the light effect on microalgal growth in the first place. The typical relationship between light intensity and photosynthetic rate for individual cells is often represented by a P-I curve ('P' represents photosynthetic rate and 'I' represents light intensity) that has conventionally been divided to three regimes according to the level of light intensity [22].

1) Linear regime: when light intensity is low, photosynthesis is primarily limited by the capture of light photons, which is defined as light–limiting step. In this case, photosynthetic rate is proportional to light intensity.
2) Light saturation regime: as more light is absorbed, algal photosynthetic rate tends to be saturated. This is because photosynthetic rate is limited by the rate of reactions following the capture of photons.

3) Light inhibition regime: when light intensity increases further, the photosynthetic rate starts to decrease because of the damage of key proteins in photosynthetic units [23].

A tremendous amount of modelling and experimental work has been done to study the relationship between light intensity and photosynthesis. *Spirulina sp.* was studied in photoautotrophic, heterotrophic and mixotrophic cultures [24]. In photoautotrophic culture, the 'P-I' relationship was investigated. The photosynthetic rate was saturated when light intensity reached 30 W/m^2. When the light intensity exceeded 50 W/m^2, light inhibition was observed. Algae samples collected in Como Lake, Minnesota were studied under different temperature and light intensity, and the 'P-I' relationship was constructed under different temperatures [25].

In order to better understand the relationship between light intensity and photosynthesis, mathematical models can be established for further analysis. Some models are named 'empirical models' and fitted to experimental data, other models are called 'mechanistic models' as they are based on the internal mechanisms of cells. Lee et al. and Bechet et al. have summarized many empirical models and indicated the application range (in terms of light intensity) of each model [26] [27]; most of the parameters cited come from fitting to the experimental data. However, in order to better understand the mechanism of photosynthesis at different levels of light intensity, mechanistic models have been used. For example, Han et al. proposed the term 'photosynthetic units (PSU)' with reactive and activated states in order to represent

detailed dynamic processes within individual cells relating to light effect [28]. Photoinhibition caused by the damage of key proteins was also addressed in a mechanistic model by the interactions between three states: open, closed and inhibited [29]. However, it should be noted that assuming the consequence of photosynthesis is only affected by light has a prerequisite that other nutrients supply is sufficient. Also, although mechanistic models can represent photosynthesis of individual cells in greater detail, for the purpose of capturing the interaction of multiple algal species, empirical models have presented the advantage of simplicity.

2.1.2. Trophic links based microbial communities

Above studies showed cases with microbial species at the same trophic level, which means there is no species feeding other species in the community. However, trophic links in naturally occurring or synthetic microbial communities play a very important role to deliver desired products, improve communities' efficiency and stability. Here, the widely studied anaerobic digestion system for trophic interactions is first reviewed, followed by a review on synthetic communities for biohydogen production, which is the target of a case study in this project. A brief summary of other cases is finally provided.

2.1.2.1. Anaerobic digestion

Anaerobic digestion (AD) is a process that involves multiple types of micro-organisms to decompose and convert organic compounds such as carbohydrates, proteins and

liquids, through a multiple-step process. Sari and Harmand studied a model of a two-step biological reaction with feedback inhibition for propionate degradation and hydrogenotrophic methanogenesis, which was a simplified version of the well-known anaerobic digestion model ADM1. Studying the stability of this two-tiered food chain, the authors demonstrated the preservation of stability over a wide range of kinetics parameters [30]. Similarly, Wade et al. analysed the stability of AD based microbial communities with two-tiered and three-tiered food web, as a function of the inlet substrate concentration and dilution rate. The steady states under sensible operation conditions were shown to be always stable and non-oscillatory. On productivity, this work showed a low inlet substrate concentration (chlorophenol), still sufficient to maintain the degradation populations in the upstream (chlorophenol- and phenol-degradation), could result in a poor yield in the downstream conversion (hydrogenotrophic methanogenesis). This type of comprehensive analysis identified important properties for such systems and also provided a guide for manipulating or controlling such systems for potential bioprocess improvements [1].

Weedermann et al. investigated a model for AD and showed that a productive biogas yield could be achieved in the region where operating parameters “pushed” the system into multiple steady states [2]. This result demonstrates that good productivity is sometimes achieved with a reduced stability as a compromise. It was also shown that in the regions where multiple steady states co-existed, biogas yield could be significantly different between different steady states. In all anaerobic digestion studies

discussed above, inhibition plays a very important role not only in affecting productivity, but also in bringing the systems to regions with coexistence of multiple steady states and/or emergence of oscillatory behaviours.

2.1.2.2 H_2 production by dark and photo fermentative bacteria

Among artificially designed engineering microbial systems, bio-hydrogen production from waste materials by means of the integration of dark and photo fermentation is a representative example. This conversion process consists of two parts, a dark-fermentation process and a photo-fermentation process. In dark fermentation, hydrogen is produced from sugars or organic wastes as substrates with volatile fatty acid (VFA) as by-product under strictly anaerobic condition. In photo fermentation, photo heterotrophic purple non-sulfur (PNS) bacteria consumes volatile fatty acid as substrates to produce additive hydrogen [31]. Compared with monoculture of dark-fermentation bacteria only, the introduction of photo fermentation bacteria is able to reduce VFAs concentration and hence moderates the drop of pH, which is able to ease pH inhibition to dark fermentation bacteria resulting in more sugars or organic wastes conversion to hydrogen [32] [33] [34].

In the co-culture experiment by Ding et al. [35], glucose was used as the sole substrate to produce acetic acid and butyric acid for the dark-fermentation bacteria (*C. butyricum* in their experiment), and the utilization rate of these intermediate products by the photo-fermentation bacteria was found to be far slower, which therefore became

the rate limiting process [35]. This finding holds in many other publications [36] [37]. Instead of glucose as the carbon source, Zagrodnik and Łaniecki [38] used starch as the substrate to carry out the co-culture experiments under the fed-batch condition. It was found that pH played an important role in hydrogen production. pH below 7 was suitable for the dark fermentation bacteria and led to their rapid growth. However, this condition was found to inhibit hydrogen production by the photo-fermentation bacteria whose optimal pH was above 7.

Jiao et al. and Lu and Lee used cellulose as the carbon source for bio-hydrogen production with a dark and photo fermentation bacterial consortium. Both work observed significant increase in H_2 production compared with the monoculture of dark fermentation bacteria [38]. In Jiao et al.'s work [37], a co-culture of dark and photo fermentation bacteria resulted in twice the rate of growth compared with the dark fermentation monoculture, with the total H_2 production by the co-culture being 1.6 times higher. In Lu and Lee's work, different levels of initial cellulose concentration were tested. The hydrogen productivity and cellulose degradation of the co-culture system were all significantly enhanced in comparison to monocultures.

Although the existing studies have demonstrated that the co-culture of dark and photo fermentation bacteria can significantly increase hydrogen production and substrates consumption, there is still potential for further improvement. First of all, some intermediates are not converted to hydrogen, which remain in the system and decrease pH, such as formate [39]. Secondly, the growth rate and pH tolerance of dark

fermentation bacteria are always higher than photo fermentation bacteria, which can hinder the working of the overall process when the pH level drops below the tolerance of photo fermentation bacteria. These gaps deserve attention from future research.

2.1.2.3 Other cases

Minty et al. developed a synthetic consortium consisting of fungus *Trichoderma reesei* and bacterium *E.coli* for efficient biosynthesis of isobutanol from lignocellulosic feedstock [40]. In this consortium, *T. reesei* secretes cellulases to hydrolyze cellulose to form soluble saccharides. The oligosaccharides are further hydrolysed to glucose by *T. reesei*. Soluble cellobiose and glucose serve as growth substrates for *T. reesei*, and glucose is then metabolized by *E. coli* to produce isobutabol. Similarly, a symbiotic co-culture ecosystem was developed to produce ethanol [41]. In this consortium, the cellulose is hydrolysed by *C. phytofermentans* to form cellodextrins, which is further converted into ethanol by both Yeast and *C. phytofermentans*. The growth of *C. phytogermentans* is inhibited by oxygen, however, yeast in the consortium can consume oxygen which provide an anaerobic environment to further increase the ethanol yield. Zhou et al. reported the use of two microbes, *E. coli* and *S. cerevisiae* in a consortium to produce precursors of the anti-cancer drug paclitaxel [42]. The product of this bioprocess is oxygenated taxadiene, however either of the microbes is unable to produce it solely. In this synthetic consortium, an engineered *E. coli* was used to produce taxadine, with *S. cerevisiae* oxygenating it to oxygenated taxadiene. This

experiment demonstrated that a complex product was able to be produced by using synthetic bioprocesses constructed from engineered microbial consortia.

2.1.3. Generalized studies on ecological interactions in microbial communities

Although there are many specific species-species and species-environmental interactions, generalised, analytical studies on these interactions are possible. Jan et al. [43] proposed two major criteria to distinguish different types of interactions: whether the interaction is passive or active in nature and whether the interaction has positive or negative effects on each of involved microbial species. A passive interaction refers to interactions where the growth of one species is affected by the inadvertent activities of a second species, for example, one species grows on a waste product that is produced by another species. An active interaction refers to a case where one species actively invests resources into metabolic processes that affect other species, in return, this species needs to grow based on the resources or activities provided by the other species, for example, a cooperation relation between two species. Although such interactions represent a number of real applications, they might not be readily identified when an ecosystem becomes very complicated, as the authors claimed.

Liu and Sumpter [44] developed a generalized bottom-up model of consumer-resource interactions, which reflects basic properties of many bacterial and other community ecologies by a set of well-defined ecological rules. The model demonstrated key properties between microbial species and evolved resources. For

example, the model successfully predicted that all available resources are fully consumed and the ecosystem can collapse even without external perturbations. Friedman et al. [45] proposed an assembly rule that predicts whether a group of species could co-exist or extinct from the outcomes of pairwise competition. The assembly rule is: in a multispecies competition, species that co-exist within pairs with all other species will survive in the multi-species community, whereas species that are excluded by any of the surviving species will go extinct. In predicting the outcome of a consortium with 3 species, the assembly rule predicted species survival across all the three-way competitions (prediction from outcomes of pairwise competitions) with an accuracy of 89.5%. When predicting the co-existence of more species (7-8) competition, by incorporating information regarding the trio outcomes, the prediction accuracy reached 86%. This work shed light on predicting the outcome of multi-species community from the outcomes of the subset microbial communities.

Allesina and Tang [46] conducted a theoretical work on the stability criteria for microbial species connected in the form of unstructured networks in which species interact through randomly generated pairwise relations, in predator-prey, mutualistic, and competitive pairs. Results demonstrated that stability correlates with the number of species, strength of interactions, types of interactions and sparsity of the interactions. To stabilize an ecosystem, one may consider to decrease its complexity or increase the self-regulation. For competition, mutualism and mixtures of the two, besides reducing the complexity of an ecosystem, decreasing the average interaction strength can also

stabilize the ecosystem. However, predator-prey ecosystems can be stabilized by increasing the strength of interactions. Coyte conducted a similar work [47]. They claimed that an increasing of cooperative links would destabilize the microbial communities. The reason for the destabilizing effect of cooperation in the model is that cooperation causes coupling between species and positive feedbacks. Consequently, cooperation can create dependency and hence the potential for mutual downfall. They also conducted a numerical experiment to introduce competing species into a co-operative network. Competitors were shown to initially stabilize the system because they would reduce the proportion of cooperative interactions. Once the number of competing species went beyond a threshold, adding more competing species began to reduce the stability due to the increased interaction complexity.

Theoretical analysis discussed above do not consider specific strains or applications by using microbial communities. However, conclusions from these analyses do hold for a wide range of real applications.

2.2. Networks and network construction for microbial communities

A network is a collection of vertices or nodes joined together in pairs by edges. Many systems of interest in biological, technological and social sciences can be thought of as networks. The structure of such a network, particularly patterns of interactions between nodes, can have a major effect on the behaviour of the underlying system. Scientists in a wide variety of fields have developed an extensive set of tools for analysing, modelling and understanding networks [48]. This project is particularly interested in the networks that represent microbial communities, with a reference to the approaches and tools applied in other fields which more recently have started to find applications in analysing microbial systems.

2.2.1 Empirical studies of networks

2.2.1.1 Social networks

Social networks are networks in which the vertices are people, or sometimes groups of people, and the edges represent some form of social interactions between people, such as friendship. In sociology, the vertices, i.e. the people are referred to as actors and the edges as ties. Techniques for analysing social networks have been used to a great variety of different communities. For example, Bernard et al. used social networks to analyse friendship and acquaintance patterns in local communities [49]. Davis and Greve analysed contacts between business people and other professionals, and boards of

directors of companies in Mexico city [50]. In all these studies, the way to collect data and information is by means of direct questioning and use of archival records.

2.2.1.2 Technological networks

There are a number of examples of technological networks, such as power grids, transportation networks, delivery and distribution networks and telephone networks. A power grid is the network of electric transmission lines that provide long-distance transport of electric power. Grid networks play an important role from geographic, social, and economic points of view; with some network statistics, insights in growth, shape, and unusual behaviours can be gained [51]. Transportation networks include airline, road, and rail routes, which are often constructed by using timetables and maps data [52]. Both power grids and transportation networks are distribution networks where the spatial distribution plays an important role. For example, unusual results such as the prediction of cascading failures in power grids are outcomes of applying network statistics [53]. In order to better demonstrate the importance of distance in a transportation network, a bipartite network was proposed, where two types of vertex were used, one representing the locations and the other representing train routes [54].

2.2.1.3 Biological networks

Networks are widely used in many branches of biology as a convenient representation of patterns of interactions between biological elements. Interactions between metabolites can be represented using a metabolic network, where the nodes represent

biological molecules and edges represent metabolic pathways by which a complex chemical breaks down to multiple simple compounds [55]. Similarly, interactions between molecular regulators that interact with each other and with other substance in the cell to govern the gene expression levels of mRNA and proteins can be constructed by a genetic regulatory network [56]. Biological networks introduced above are elements interacting within a cell. However, a broad level of interaction between different cells or different organism can be represented by using different types of biological networks, such as an ecological network to represent a food web in a microbial community [57].

2.2.2. Structural analysis of networks

For a given network, a number of structural characteristics can be obtained using certain algorithms. Here, a brief description of one of the most important characteristics, centrality, is provided. Two applications of the structural information of a network, detection of communities and development of control strategies, are discussed.

2.2.2.1 Centrality

Centrality is a measurement of importance of a vertex in a network, with a number of different definitions. The simplest centrality measure in a network is the degree of a vertex, the number of edges connected to it, which is referred to as degree centrality. A natural extension of the simple degree centrality is eigenvector centrality, which can be considered as a “central” point for the overall network [58]. However eigenvector

centrality has problems in directed networks, for example, vertices with no incoming edges will have a centrality of 0. In order to overcome this problem, the Katz centrality was proposed which gives each vertex a small amount of centrality 'for free' [59].

In 1996, one of the founders of Google, Larry Page, proposed an algorithm to rank web pages [60]. This algorithm is called 'PageRank'. Mathematically, it is a variant of the Katz centrality in which the centrality of a vertex is proportional to the sum of the centrality-to-out-degree ratios of its neighbours. There are other different measures of centrality, such as the closeness centrality, which measures the mean distance from a vertex to other vertices in a network, and the betweenness centrality, which measures the extent to which a vertex lies on paths between other vertices [61].

2.2.2.2 Detection of communities

The word "community" in network science represents a group of nodes that are relatively densely connected to each other but sparsely connected to the rest of the network. It plays very important roles in many subjects such as identifying people in different groups on Facebook. The algorithms for community detection are well studied.

In computer science, an algorithm was proposed by Brian Kernighan and Shen Lin in 1970 to partition electric circuits into boards so that nodes on different boards (corresponding to different communities) can be linked to each other using the fewest number of connections [62]. Communities can also be derived by betweenness centrality based algorithms. As introduced above, an edge has a high betweenness if it

lies on a large number of short paths between vertices. When implementing this algorithm, one can find the node with the largest betweenness centrality, remove it, and recalculate betweenness centrality for the remaining network and get rid of the node with largest centrality, until one obtains a set of isolated networks, which each can be regarded as a community [63].

One of the most popular algorithms for community detection is called modularity optimization. The modularity (denoted as Q) is the number of edges falling within a group of nodes minus the expected number of edges in an equivalent random network with nodes of the same degree. A high modularity means that there are more interactions than is expected within a community of nodes [64]. Another famous community detection algorithm is called k-clique percolation. k-clique represents a complete subnetwork of k nodes that are connected with all other nodes. A k-clique community is then defined as the union of all adjacent k-cliques. The k-clique percolation algorithm is an example of a local community finding method, which is a very effective way to detect communities by focusing on community overlap [65].

2.2.2.3 Employing structural information for control

With a carefully constructed network, a crucial goal is to control it. There are analytical tools developed to study the controllability of an arbitrary complex directed network, identifying the set of driver nodes to guide the system's entire dynamics. Liu et al. [66] studied the 'controllability' of networks, i.e. the ability to identify a suitable choice of

inputs to drive a network from any initial state to any desired final state within finite time. Based only on the topological structure of a directed network with linear dynamics, it was shown that full control can be gained on the network if and only if each so-called "unmatched" node (i.e. a node that is not matched by the maximum set of links that do not share start or end nodes) can be directly controlled and there are directed paths from the input signals to all matched nodes. In a more recent study [64], the theory was extended to microbial community networks with nonlinear dynamics. In this study, a similar conclusion was made: a microbial community network is structurally accessible (similar as controllable) if (1) each species is the end-node of a path that starts with a control input node, and (2) there is a disjoint union of cycles and paths that cover all species nodes [67].

Another type of network control strategy is called 'feedback vertex set (FVS)' control [68-69]. In this strategy, the purpose is to drive the system to long term dynamical behaviours of the whole system, such as steady states, periodic oscillations or quasi-periodic oscillations. It is worth to note that controlling all FVS nodes is a sufficient condition to control the dynamics of the overall network. In graph theory, a FVS is a subset of vertices in a directed graph, the removal of which leaves the graph without directed cycles. Atsushi et al. [68-69] demonstrated with empirical examples the ability to control network-based dynamical systems by using FVS nodes.

2.2.3. Microbial network construction from species abundance data

2.2.3.1 Undirected networks

An undirected network is a network with microbial species connecting with each other without a causal relation which would imply an abundance change of one species can lead to an abundance change of another, not in the opposite direction. Normally, such techniques often consider co-occurrence or sharing of a similar abundance pattern as a positive relationship between two species, and mutual exclusion as a negative relationship. While these principles appears to be simple, there are a number of difficulties in constructing undirected ecological networks from real data, which may suffer from compositionality or low availability.

Obtaining microbial communities data using high throughput genomic surveys, such as 16S rRNA gene profiling, can sometimes provide compositional data where only the relative abundance is available. Directly using relative abundance to construct ecological network can cause spurious correlation [70]. In Figure 2.1, the abundance of species 1 increases, and at the same time, the abundance of species 2 does not change. However, if measured by relative abundance, as the composition of species 1 increases, at the same time, the composition of species 2 decreases. If such a relation is measured by Pearson correlation, it will be determined as a negative correlation between the two species, which is apparently wrong.

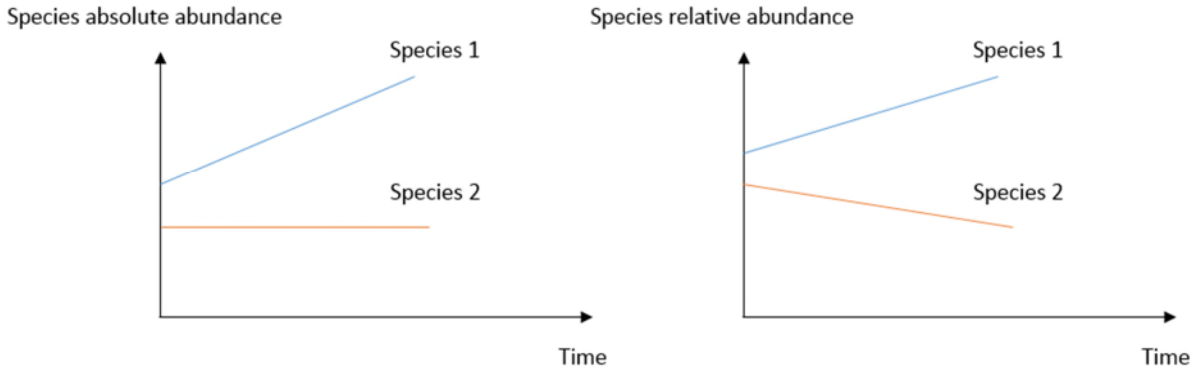


Figure 2.1. Spurious Correlation.

Friedman and Alm introduced a technique for inferring microbial species correlations from compositional data called SparCC (Sparse Correlations for Compositional data) [71]. It estimates the linear Pearson correlations between the log-transformed compositional abundance which cannot be computed exactly. SparCC is based on the assumptions that the number of components (ecological species) is large and the true correlation network is sparse. Similarly, Fang et al. [72] proposed a model called CCLasso (Correlation inference for Compositional data through Lasso) based on l_1-*penalty* on the off-diagonal variance-covariance matrix after log ratio transformation of compositional data. Compared to SparCC, CCLasso is able to guarantee a positive definite covariance matrix for inferred correlation.

Instead of inferring the correlation matrix, a l_1-*penalty* based graph model (Graphical Lasso) estimating the log-basis sparse precision matrix (inverse correlation matrix) was proposed by Kurtz et al. [73]. This model assumes variables (microbial species dynamics) follow a sparse Gaussian multi-variate distribution. The sparsity is

obtained by penalizing the l_1-*norm* of the elements in the precision matrix with a shrinkage parameter λ tuning the level of sparsity. The tuning parameter λ can be obtained by means of the Stability Approach to Regularization Selection (StARS) [74]. This approach repeatedly takes random subsamples of the data and estimates the edge stability under different candidate shrinkage parameter λ. Schwager et al. proposed a Bayesian approach to identify correlations, which is able to quantify uncertainty by using a graphical Lasso prior [75].

Techniques introduced above are able to solve problem of inferring species correlation from compositional data. Some of them are able to infer reliable species correlations from few data, such as by increasing the shrinkage parameter λ to get sparse correlation matrix. An undirected network is then able to be constructed from a sparse correlation matrix.

2.2.3.2 Directed networks

Different from correlation based methods for constructing undirected networks, one can infer (the trend in) the abundance of one species from the combined abundances of other taxa to obtain directed networks. In directed networks, there are edges each pointing one species to another, which indicates that a change of abundance of the former will lead to the abundance change of the latter. A number of studies have reported the construction of directed networks using time series data that record the temporal variation of microbial communities in a wide range of environments.

Normally the parameters in the model are fitted from temporal data through some kind of regularization (l_1 or l_2 norm). These parameters represent the edges linking the nodes in the network.

Ridenhour et al. [76] used an ARIMA (Autoregressive integrated moving average) model with elastic-net regularization to estimate a model of gut microbiome dynamics from white-throated woodrats. This regularization constrains estimates to allow for the possibility of many more interactions than data, and is scalable to handle data consisting of hundreds of taxa. Similarly, Gibbons et al. [77] applied the sparse vector autoregressive model (sVAR) to model the gut community dynamics. In the gut microbial community, the top 50 most abundant species were fit to a lag-3 sVAR model, with elastic net regularization to enforce very small (likely to be uncorrelated) coefficients to be zeros. Granger causality coefficients were then calculated for each non-zero sVAR coefficient, and statistical significance was assessed to infer causal relations between species. Chen et al. [78] proposed a linear state space model (SSM) to capture the dynamics of the vaginal microbiome. This model was able to measure system noise and measurement noise. Expectation maximization algorithm was applied to construct a dynamic microbial interaction network with a Lasso based variable selection method to achieve sparsity.

Models introduced above assume linear dynamics. However, there is another type of models called generalized Lotka Volterra model (gLV) [79] which assumes the

existence of mechanistic coupling between species or environmental factors in the community, shown in equation (2.1).

$$\frac{dx_i}{dt} = \mu_i x_i + x_i \sum_{j=1}^{L} \beta_{i,j} x_j + x_i \sum_{k=1}^{P} \gamma_{i,k} u_k \quad (2.1)$$

where x represents the abundance of a species, μ_i is the specific growth rate of species *i*, $\beta_{i,j}$ is the effect of the interaction of species *j* on species *i*, u_k represents the time dependent perturbation (for example, an antibiotic or feed), $\gamma_{i,k}$ represents the effect of interaction of perturbation to species *i*. L and P are numbers of microbial species and perturbations, respectively. Discretizing equation (2.1) gives:

$$\frac{\Delta \ln(x_{i,t})}{\Delta t} = \mu_i + \sum_{j=1}^{L} \beta_{i,j} x_{j,t} + \sum_{k=1}^{P} \gamma_{i,k} u_k \quad (2.2)$$

The model parameters can then be determined by a linear system of equations. When parameters are fitted, a directed network can be obtained. $\beta_{i,j}$ represents the link from species *j* to *i*, where a positive $\beta_{i,j}$ means a positive link where an increase of species *i* at time t is because of an increase of species *j* at time t+1, and a negative $\beta_{i,j}$ represents a negative link from species *j* to species *i*, and if $\beta_{i,j}$ is 0, it means species *i* is independent of species j, and hence no link exists from species *j* to species *i*. When *i*=*j*, $\beta_{i,i}$ denotes the decay of species *i*. The discretized gLV model was proposed by Stein et al. [79]. In a gLV model (equation 2.2) for a community with L species and P perturbations, there are L^2+L+LP parameters needed to be estimated, which may result in an under–determined system that a proper regularization is needed

(e.g. Ridge or Lasso). Several algorithms have been developed to regularize or select variables to obtain a sparse network.

Shaw et al. [80] proposed a tool called MetaMIS (Metagenomic Microbial Interaction Simulator) that applies partial least square regression for highly correlated time series data. Fisher and Mehta [81] proposed an approach with forward stepwise regression coupled with bootstrap aggregation to select covariates in order to infer the microbial network structure. The algorithm is named LIMIT (Learning Interactions from Microbial Time Series) and is able to reliably infer the network structure from the underlying gLV model. Bucci et al. [82] presented Microbial Dynamical Systems Inference Engine (MDSINE), which is an open source software system for modelling microbiome dynamics and inferring network structures from time series data, with the extended gLV equations as the underlying dynamical systems model. In the software, multiple algorithms were developed to achieve different modelling constraints and network inference requirements. A ridge constrained gLV model was constructed for microbiome dynamics prediction. In this algorithm, growth rates (μ_i) were constrained to be positive and the self-interaction $\beta_{i,i}$ was constrained to be negative to achieve their physical meanings. Two Bayesian dynamical systems inference algorithms were developed to model microbiome dynamics and infer microbial network structures. The first of the two Bayesian inference algorithms is called Bayesian Adaptive Lasso, a l_1-norm based approach to estimate gLV model parameters with sparsity. The second approach is called Bayesian Variable Selection, used to infer the presence/absence of

microbe-microbe interactions or microbe-perturbation effects. Lo and Marculescu proposed a framework named Microbial Time-series Prior Lasso (MTPLasso) which is able to integrate prior knowledge to a Lasso based form to achieve sparsity in order to obtain microbial interaction networks [83].

Algorithms and software developed above have enabled a number of research to be carried out. For example, anaerobic digestion microbiome interaction networks were inferred at both mesophilic and thermophilic temperatures using MDSINE [57]. After obtaining the topological structure of the microbial community, influential species were captured by means of centrality. As an early attempt, this work gained a systematic understanding of interactions between anaerobic digestion species [57]. In another study [84], a three-step classification scheme consisting of two tests and a model fitting process was developed to assess the prevalence of interactions from microbial time series data. The tests include one on the temporal data structure (noise types, whether the noise is white noise or coloured noise) and one on neutrality (potential for the existence of interactions in the microbial community). Once the two tests are passed, a gLV model is constructed with LIMIT to obtain a sparse network structure.

Dam et al. [85] applied slightly modified gLV models to capture and interpret the yearly abundance patterns of the species in a real-world system, Lake Mendota, WI, USA. A two-layered set of dynamic models was proposed with the first layer analysing the interactions between sub-communities and the second layer focusing on interactions between individual species and sub-communities. To capture the microbial network

structure accurately, Venturelli et al. proposed a systematic modelling and experimental pipeline to gain information of microbial interactions that mediate the community assembly [85].

2.3. Bioreactor design

A bioreactor is a vessel that carries out biological or biochemical processes that typically involve the anaerobic or aerobic culture of cells for delivering desired products. Bioreactors are used in various industries [86]. For example, in tissue engineering, bioreactors are tools used to mature and guide the development of tissue engineered constructs. They also guide tissue structure, organization and functions through the application of chemical and mechanical stimuli [87]. Bioreactors are also used in pharmaceutical industry, for example, for the production of enzymes in airlift and bubble column bioreactors [88]. In anaerobic digestion and biological wastewater treatment, bioreactors are applied for large-scale chemical and biological wastes conversion [89].

Bioreactor design considerations are crucial for achieving the desired featuresof a bioprocess, with respect to e.g. yield and selectivity of a desired product, economic cost, and environmental impacts. In order to achieve these goals, various configurations of bioreactors are applied. Traditionally, a commonly used configuration is called continuous stirred tank reactors (CSTRs). From modelling point of view, all chemicals in CSTR are assumed to be well-mixed, with no temperature or pressure gradient [86]. Beyond simple CSTRs, one of the research directions points to systems combining membranes, which are called membrane bioreactors. In this field, relevant literature has reviewed the improved properties of this types of bioreactors in terms of productivity and operational robustness [90]. For biological wastewater treatment, rotating

biological contactors were investigated to increase the contact area between the biofilm and wastewater by adding a rotating axle [91]. In order to overcome the problems related to oxygen transfer rate and excessive mycelia growth, applying a pulsing flow to fluidized-bed bioreactors was studied to control the pellet size effectively. In the current thesis, the focus is on the ecological interactions between microbial species, and CSTR has been assumed to be the bioreactor type, leaving the study on the impact of spatial distribution as a subject of future work.

2.4. Research Gaps

Although microbial community based engineering systems have been considered to play important roles in industry, systematic assessment of the potential is still lacking. In poly-culture of microalgae, although light complementary utilization is a widely-adopted hypothesis, there is no work to verify it. In dark-light fermentation bio-hydrogen production based system, although potentially there is a gain from the two species co-culture, again, there is no systematic assessment to analyse such a system by considering production and stability at the same time. Another question also arises as to what consequence the introduction of additional species into a community may have, particularly in terms of the interplay between the intended objective and unintended outcome.

The above questions could be explored with specific systems. On the other hand, generalised studies have been carried out on the dynamics of microbial communities, as reviewed earlier in this chapter, however, systematic studies particularly on generalised schemes that are particularly likely to be shared by small-scale, synthetic microbial communities are still lacking. For example, if one species feeds the other without any other interactions, is there a generalized dynamic pattern? What if the other species inhibits the first one? A generalized understanding of dynamics arising from typical interactions in such microbial communities would gain insights on the productivity and the stability of relevant bioreactors.

On the inference of network structure for large-scale microbial communities, a particular aspect which has received very little attention is the chemical-mediated nature of the microbial interactions. How well will an inference model considering only direct microbe-microbe interactions, such as the gLV model, perform for inferring the underlying microbial network, when the scale, the level of complexity, and information content of the available data from such chemical-mediated microbial communities? How would the existing inference models compare with one that does capture microbe-chemical interactions? These questions are yet to be answered.

References

[1]. Wade, M. J., Pattinson, R. W., Parker, N. G., & Dolfing, J. (2016). Emergent behaviour in a chlorophenol-mineralising three-tiered microbial 'food web'. Journal of theoretical biology, 389, 171-186.

[2]. Weedermann, M., Wolkowicz, G. S., & Sasara, J. (2015). Optimal biogas production in a model for anaerobic digestion. Nonlinear Dynamics, 81(3), 1097-1112.

[3]. Song, H., Ding, M. Z., Jia, X. Q., Ma, Q., & Yuan, Y. J. (2014). Synthetic microbial consortia: from systematic analysis to construction and applications. Chemical Society Reviews, 43(20), 6954-6981.

[4]. Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N., & Silver, P. A. (2015). Better together: engineering and application of microbial symbioses. Current Opinion in Biotechnology, 36, 40-49.

[5]. Shurin, J. B., Abbott, R. L., Deal, M. S., Kwan, G. T., Litchman, E., McBride, R. C., ... & Smith, V. H. (2013). Industrial‐strength ecology: trade‐offs and opportunities in algal biofuel production. Ecology letters, 16(11), 1393-1404.

[6]. Donlan, R. M. (2002). Biofilms: microbial life on surfaces. Emerging infectious diseases, 8(9), 881.

[7]. Lu, H., & Lee, P. K. (2015). Effects of cellulose concentrations on the syntrophic interactions between *Clostridium cellulovorans* 743B and *Rhodopseudomonas*

palustris CGA009 in coculture fermentation for biohydrogen production. International Journal of Hydrogen Energy, 40(35), 11800-11808.

[8]. Großkopf, T., & Soyer, O. S. (2014). Synthetic microbial communities. Current opinion in microbiology, 18, 72-77.

[9]. Cardinale, B. J. (2011). Biodiversity improves water quality through niche partitioning. Nature, 472(7341), 86.

[10]. Kazamia, E., Riseley, A. S., Howe, C. J., & Smith, A. G. (2014). An engineered community approach for industrial cultivation of microalgae. Industrial Biotechnology, 10(3), 184-190.

[11]. Tilman, D. (1980). Resources: a graphical-mechanistic approach to competition and predation. The American Naturalist, 116(3), 362-393.

[12]. Pavlou, S., Kevrekidis, I. G., & Lyberatos, G. (1990). On the coexistence of competing microbial species in a chemostat under cycling. Biotechnology and bioengineering, 35(3), 224-232.

[13]. Dikshitulu, S., Baltzis, B. C., Lewandowski, G. A., & Pavlou, S. (1993). Competition between two microbial populations in a sequencing fed‐batch reactor: Theory, experimental verification, and implications for waste treatment applications. Biotechnology and bioengineering, 42(5), 643-656.

[14]. Ajbar, A. (2002). Classification of static and dynamic behavior in chemostat for plasmid-bearing, plasmid-free mixed recombinant cultures. Chemical Engineering Communications, 189(8), 1130-1154.

[15]. Pulz, O., & Gross, W. (2004). Valuable products from biotechnology of microalgae. Applied microbiology and biotechnology, 65(6), 635-648.

[16]. Zhou, W., Chen, P., Min, M., Ma, X., Wang, J., Griffith, R., ... & Shi, J. (2014). Environment-enhancing algal biofuel production using wastewaters. Renewable and sustainable energy reviews, 36, 256-269.

[17]. Chiu, S. Y., Kao, C. Y., Chen, T. Y., Chang, Y. B., Kuo, C. M., & Lin, C. S. (2015). Cultivation of microalgal *Chlorella* for biomass and lipid production using wastewater as nutrient resource. Bioresource technology, 184, 179-189.

[18]. Brennan, L., & Owende, P. (2010). Biofuels from microalgae—a review of technologies for production, processing, and extractions of biofuels and co-products. Renewable and sustainable energy reviews, 14(2), 557-577.

[19]. Stockenreiter, M., Haupt, F., Graber, A. K., Seppälä, J., Spilling, K., Tamminen, T., & Stibor, H. (2013). Functional group richness: implications of biodiversity for light use and lipid yield in microalgae. Journal of phycology, 49(5), 838-847.

[20]. Striebel, M., Behl, S., Diehl, S., & Stibor, H. (2009). Spectral niche complementarity and carbon dynamics in pelagic ecosystems. The American Naturalist, 174(1), 141-147.

[21]. Behl, S., Donval, A., & Stibor, H. (2011). The relative importance of species diversity and functional group diversity on carbon uptake in phytoplankton communities. Limnology and Oceanography, 56(2), 683-694.

[22]. Béchet, Q., Shilton, A., & Guieysse, B. (2013). Modeling the effects of light and temperature on algae growth: state of the art and critical assessment for productivity prediction during outdoor cultivation. Biotechnology advances, 31(8), 1648-1663.

[23]. Rubio, F. C., Camacho, F. G., Sevilla, J. F., Chisti, Y., & Grima, E. M. (2003). A mechanistic model of photosynthesis in microalgae. Biotechnology and bioengineering, 81(4), 459-473.

[24]. Chojnacka, K., & Noworyta, A. (2004). Evaluation of *Spirulina sp.* growth in photoautotrophic, heterotrophic and mixotrophic cultures. Enzyme and Microbial Technology, 34(5), 461-465.

[25]. Megard, R. O., Tonkyn, D. W., & Senft, W. H. (1984). Kinetics of oxygenic photosynthesis in planktonic algae. Journal of Plankton Research, 6(2), 325-337.

[26]. Béchet, Q., Shilton, A., & Guieysse, B. (2013). Modeling the effects of light and temperature on algae growth: state of the art and critical assessment for productivity prediction during outdoor cultivation. Biotechnology advances, 31(8), 1648-1663.

[27]. Lee, E., Jalalizadeh, M., & Zhang, Q. (2015). Growth kinetic models for microalgae cultivation: a review. Algal Research, 12, 497-512.

[28]. Han, B. P. (2001). Photosynthesis–irradiance response at physiological level: a mechanistic model. Journal of theoretical biology, 213(2), 121-127.

[29]. Hartmann, P., Nikolaou, A., Chachuat, B., & Bernard, O. (2013, July). A dynamic model coupling photoacclimation and photoinhibition in microalgae. In 2013 European Control Conference (ECC) (pp. 4178-4183). IEEE.

[30]. Sari, T., & Harmand, J. (2016). A model of a syntrophic relationship between two microbial species in a chemostat including maintenance. Mathematical biosciences, 275, 1-9.

[31]. Basak, N., & Das, D. (2007). The prospect of purple non-sulfur (PNS) photosynthetic bacteria for hydrogen production: the present state of the art. World Journal of Microbiology and Biotechnology, 23(1), 31-42.

[32]. Argun, H., & Kargi, F. (2011). Bio-hydrogen production by different operational modes of dark and photo-fermentation: an overview. International Journal of Hydrogen Energy, 36(13), 7443-7459.

[33]. Wang, J., & Wan, W. (2009). Kinetic models for fermentative hydrogen production: a review. International Journal of Hydrogen Energy, 34(8), 3313-332.

[34]. Rai, P. K., & Singh, S. P. (2016). Integrated dark-and photo-fermentation: recent advances and provisions for improvement. International Journal of Hydrogen Energy, 41(44), 19957-19971.

[35]. Ding, J., Liu, B. F., Ren, N. Q., Xing, D. F., Guo, W. Q., Xu, J. F., & Xie, G. J. (2009). Hydrogen production from glucose by co-culture of *Clostridium butyricum* and immobilized Rhodopseudomonas faecalis RLD-53. International journal of hydrogen energy, 34(9), 3647-3652.

[36]. Uyar, B., Eroglu, I., Yücel, M., & Gündüz, U. (2009). Photofermentative hydrogen production from volatile fatty acids present in dark fermentation effluents. International Journal of hydrogen energy, 34(10), 4517-4523.

[37]. Jiao, Y., Navid, A., Stewart, B. J., McKinlay, J. B., Thelen, M. P., & Pett-Ridge, J. (2012). Syntrophic metabolism of a co-culture containing *Clostridium cellulolyticum* and *Rhodopseudomonas palustris* for hydrogen production. international journal of hydrogen energy, 37(16), 11719-11726.

[38]. Zagrodnik, R., & Łaniecki, M. (2017). Hydrogen production from starch by co-culture of *Clostridium acetobutylicum* and *Rhodobacter sphaeroides* in one step hybrid

dark-and photofermentation in repeated fed-batch reactor. Bioresource technology, 224, 298-306.

[39]. Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M., & Yukawa, H. (2005). Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing Escherichia coli strains. Appl. Environ. Microbiol., 71(11), 6762-6768.

[40]. Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C. H., Ahn, J. H., Foster, C. E., ... & Lin, X. N. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proceedings of the National Academy of Sciences, 110(36), 14592-14597.

[41]. Zuroff, T. R., Xiques, S. B., & Curtis, W. R. (2013). Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic *Clostridium phytofermentans*/yeast co-culture. Biotechnology for biofuels, 6(1), 59.

[42]. Zhou, K., Qiao, K., Edgar, S., & Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. Nature biotechnology, 33(4), 377.

[43]. Dolinšek, J., Goldschmidt, F., & Johnson, D. R. (2016). Synthetic microbial ecology and the dynamic interplay between microbial genotypes. FEMS microbiology reviews, 40(6), 961-979.

[44]. Liu, Y., & Sumpter, D. (2017). Insights into resource consumption, cross-feeding, system collapse, stability and biodiversity from an artificial ecosystem. Journal of The Royal Society Interface, 14(126), 20160816.

[45]. Friedman, J., Higgins, L. M., & Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. Nature ecology & evolution, 1(5), 0109.

[46]. Allesina, S., & Tang, S. (2012). Stability criteria for complex ecosystems. Nature, 483(7388), 205.

[47]. Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). The ecology of the microbiome: networks, competition, and stability. Science, 350(6261), 663-666.

[48]. Newman, M. (2010). Networks: an introduction. Oxford university press.

[49]. Bernard, H. R., Johnsen, E. C., Killworth, P. D., & Robinson, S. (1991). Estimating the size of an average personal network and of an event subpopulation: Some empirical results. Social science research, 20(2), 109-121.

[50]. Davis, G. F., & Greve, H. R. (1997). Corporate elite networks and governance changes in the 1980s. American journal of sociology, 103(1), 1-37.

[51]. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. Nature, 393(6684), 440.

[52]. Pitts, F. R. (1965). A graph theoretic approach to historical geography. The professional geographer, 17(5), 15-20.

[53]. Dobson, I., Carreras, B. A., Lynch, V. E., & Newman, D. E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. Chaos: An Interdisciplinary Journal of Nonlinear Science, 17(2), 026103.

[54]. Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. Physical Review E, 67(3), 036106.

[55]. Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? Nature biotechnology, 28(3), 245.

[56]. Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E., & Guthke, R. (2009). Gene regulatory network inference: data integration in dynamic models—a review. Biosystems, 96(1), 86-103.

[57]. Shaw, G. T. W., Liu, A. C., Weng, C. Y., Chou, C. Y., & Wang, D. (2017). Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. PloS one, 12(7), e0181395.

[58]. Newman, M. E. (2016). Mathematics of networks. The new Palgrave dictionary of economics, 1-8.

[59]. Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1), 39-43.

[60]. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

[61]. Freeman, L. C. (1977). A set of measures of centrality based on betweenness. Sociometry, 35-41.

[62]. Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. Bell system technical journal, 49(2), 291-307.

[63]. Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12), 7821-7826.

[64]. Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582.

[65]. Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043), 814.

[66]. Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2011). Controllability of complex networks. Nature, 473(7346), 167.

[67]. A theoretical framework for controlling complex microbial communities. Nature communications, 10.

[68]. Fiedler, B., Mochizuki, A., Kurosawa, G., & Saito, D. (2013). Dynamics and control at feedback vertex sets. I: Informative and determining nodes in regulatory networks. Journal of Dynamics and Differential Equations, 25(3), 563-604.

[69]. Mochizuki, A., Fiedler, B., Kurosawa, G., & Saito, D. (2013). Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. Journal of theoretical biology, 335, 130-146.

[70]. Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., ... & Birmingham, A. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME journal, 10(7), 1669.

[71]. Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. PLoS computational biology, 8(9), e1002687.

[72]. Fang, H., Huang, C., Zhao, H., & Deng, M. (2015). CCLasso: correlation inference for compositional data through Lasso. Bioinformatics, 31(19), 3172-3180.

[73]. Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. PLoS computational biology, 11(5), e1004226.

[74]. Liu, H., Roeder, K., & Wasserman, L. (2010). Stability approach to regularization selection (stars) for high dimensional graphical models. In Advances in neural information processing systems (pp. 1432-1440).

[75]. Schwager, E., Mallick, H., Ventz, S., & Huttenhower, C. (2017). A Bayesian method for detecting pairwise associations in compositional data. PLoS computational biology, 13(11), e1005852.

[76]. Ridenhour, B. J., Brooker, S. L., Williams, J. E., Van Leuven, J. T., Miller, A. W., Dearing, M. D., & Remien, C. H. (2017). Modeling time-series data from microbial communities. The ISME journal, 11(11), 2526.

[77]. Gibbons, S. M., Kearney, S. M., Smillie, C. S., & Alm, E. J. (2017). Two dynamic regimes in the human gut microbiome. PLoS computational biology, 13(2), e1005364.

[78]. Chen, I., Kelkar, Y. D., Gu, Y., Zhou, J., Qiu, X., & Wu, H. (2017). High-dimensional linear state space models for dynamic microbial interaction networks. PloS one, 12(11), e0187822.

[79]. Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., ... & Xavier, J. B. (2013). Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS computational biology, 9(12), e1003388.

[80]. Shaw, G. T. W., Pao, Y. Y., & Wang, D. (2016). MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. Bmc Bioinformatics, 17(1), 488.

[81]. Fisher, C. K., & Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PloS one, 9(7), e102451.

[82]. Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., ... & Olle, B. (2016). MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome biology, 17(1), 121.

[83]. Lo, C., & Marculescu, R. (2017, August). Inferring Microbial Interactions from Metagenomic Time-series Using Prior Biological Knowledge. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 168-177). ACM.

[84]. Shaw, G. T. W., Liu, A. C., Weng, C. Y., Chou, C. Y., & Wang, D. (2017). Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. PloS one, 12(7), e0181395.

[85]. Faust, K., Bauchinger, F., Laroche, B., De Buyl, S., Lahti, L., Washburne, A. D., ... & Widder, S. (2018). Signatures of ecological processes in microbial community time series. Microbiome, 6(1), 120.

[86]. Venturelli, O. S., Carr, A. V., Fisher, G., Hsu, R. H., Lau, R., Bowen, B. P., ... & Arkin, A. P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. Molecular systems biology, 14(6), e8157.

[86] Sinnott, R. A. Y. (2014). Chemical engineering design (Vol. 6). Elsevier.

[87] Blose, K. J., Krawiec, J. T., Weinbaum, J. S., & Vorp, D. A. (2014). Bioreactors for tissue engineering purposes. In Regenerative Medicine Applications in Organ Transplantation (pp. 177-185). Academic Press.

[88] Farinas, C. S. (2016). Advances in Cultivation Strategies of Aspergillus for Production of Enzymes Involved in the Saccharification of Lignocellulosic Feedstocks. In New and Future Developments in Microbial Biotechnology and Bioengineering (pp. 141-154). Elsevier.

[89] Henze, M., van Loosdrecht, M. C., Ekama, G. A., & Brdjanovic, D. (Eds.). (2008). Biological wastewater treatment. IWA publishing.

[90] Bakonyi, P., Nemestóthy, N., Simon, V., & Bélafi-Bakó, K. (2014). Fermentative hydrogen production in anaerobic membrane bioreactors: a review. Bioresource technology, 156, 357-363.

[91] Eberl, H., Morgenroth, E., Noguera, D., Picioreanu, C., Rittmann, B., van Loosdrecht, M., & Wanner, O. (2006). Mathematical modeling of biofilms (Vol. 18). London: IWA publishing.

[92] Moreira, M. T., Feijoo, G., Sanromán, A., & Lema, J. M. (1996). Effect of pulsation on morphology of Aspergillus niger and Phanerochaete chrysosporium in a fluidized-bed reactor. In Progress in Biotechnology (Vol. 11, pp. 518-523). Elsevier.

Chapter 3. Overyielding potential of microalgal polyculture with complementary light absorption spectra: A model-based analysis

Summary

This work was motivated by recent experiments which showed that the biodiversity of microalgal poly-culture was able to lead to overyielding, that is, the collective yield of microalgae species within the community exceeds that of cultivation of any single species under the same operation condition. A hypothetic reason to explain the observed overyielding has been the difference between algal species in their ability to absorb light segments of specific wavelength ranges. In terms of ecological interactions, this offers an interesting example of complementary resource utilisation between microbial species, and hence taken as a case study in this project with an aim understand the impact of a number of conditions on the potential of overyielding in such systems.

A mathematical model is developed for microalgae growth using different light segments. Light attenuation within a photobioreactor is modelled by means of Beer-Lambert law. Three growth regions differentiated by incident light intensity are considered with corresponding relations between photosynthetic rate and light absorption. For modelling complementary light absorption spectra, two species with a simplified (artificial) absorption profile are used to allow an easy manipulation of the light absorption complementarity. The results from this modelling exercise are subsequently verified with a co-culture system using real absorption spectra.

This study shows that the conditions affecting overyielding can be divided into operational settings and species traits. With a comprehensive analysis, light intensity, loss rate and light complementarity are shown to affect the possibility and the degree of overyielding, through a trade-off between the amount of light absorbed and the biological utilisation efficiency of the absorbed light. Specific growth rate and light use efficiency are shown to affect the competitiveness of the species, hence their co-existence which is a a prerequisite for overyielding. Overall, this work shows that complementary light utilization alone does not guarantee overyielding; other influencing factors also matter, such as light intensity and loss rate. More generally, this work suggests that complementary resource utilisation can be beneficial conditionally, and the conditions need to be carefully revealed when designing bioreactors.

This chapter comprises the following article:

Di, S., & Yang, A. (2018). Overyielding potential of microalgal polyculture with complementary light absorption spectra: A model-based analysis. Biomass and Bioenergy, 118, 141-148.

List of symbols

f_{I_i} : light factor for species i

$I(\lambda, z)$: light intensity of wavelength λ at depth z

$I_{in}(\lambda)$: incident light intensity of wavelength λ

$k_i(\lambda)$: specific light absorption spectrum of species *i*

k_{ij}: light absorption spectrum of species *j* on part *i* of the light

L_i : loss rate

N_i: cells density

$p_{i_{max}}$: maximum specific growth rate

q: degree of light absorption complementarity

z: vertical position in the bioreactor between the top (z=0) and the bottom ($z=z_m$)

$\gamma_i(z)$: light absorbed by species *i* at depth z, on a per cell basis

$\gamma_{i_{opt}}$: optimum level of light absorption corresponding to the maximum photosynthetic rate

ϕ_i: light use efficiency

3.1. Introduction

In the last few decades, microalgae have been considered to possess great potential for a variety of industrial applications. A lot of existing work has focused on developing practical algal biotechnology for biofuel production [1-3]. Microalgae-based systems have also frequently been considered for biological wastewater treatment [4] [5] and CO_2 sequestration [6, 7]. Microalgae can be cultured by either a mono-culture system, or a poly-culture system that grows multiple micro-algal species. Some of poly-culture systems aim at overyielding as introduced in section 2.1.1.2. A key mechanism that has been used to explain overyielding in a poly-culture system is the complementarity effect [8] [9]. Complementarity refers to resource partitioning or facilitation among different species causing a more complete consumption of the involved resources, which may lead to an increase in the overall yield. In particular, the complementary use of light at different wavelengths by different algal species has been suggested by several authors as a possible explanation of experimentally observed overyielding in poly-culture systems (e.g. [8], [9], [10], [11]).

Several model-based analyses have been made on the co-growth of multiple phytoplankton species in connection of light utilisation. The rather early work in [12] confirmed that a growth rate model that takes into account the effect of the spectral qualities of light could predict the co-existence of multiple photosynthetic microorganisms in a well-mixed environment. The more recent work in [13] and [14] showed that red and green picocyanobacteria, using the pigments phycoerythrin and

phycocyanin to absorb green and red light respectively, could co-exist due to the complementarity in light use, as confirmed by experimental observations. Theoretical analysis was also carried out on multiple species competing over light (as a homogeneous resource) and nutrients, which revealed conditions for co-existence in such circumstances [15] [16]. It should be noted that the above model-based or theoretical work has all focused on co-existence of multiple algal species, but not on overyielding.

Focusing on overyielding, this work aims to theoretically examine the hypothesis that complementary light utilisation leads to overyielding in microalgal poly-culture, as proposed in some previous work. A mathematical model is formulated to predict the potential of overyielding as a function of several influencing factors. The model is then used in the numerical simulation of a two-species system, with hypothetical (simplified) and actual light absorption spectra, to explore the conditions for overyielding to occur in a poly-culture system.

3.2. Methods

3.2.1 Modelling approach

This work considers light supply as the only limiting factor to the growth of microalgae; nutrients such as nitrogen and phosphorus are assumed sufficient for cultivation. Mass balance of microalgae species *i* in a well-mixed bioreactor follows equation (3.1):

$$\frac{dN_i}{dt} = \frac{N_i p_{i_{max}}}{z_m} \int_0^{z_m} f_{I_i}(z)dz - L_i N_i \quad (3.1)$$

where N_i represents the cell density (i.e. number of cells per unit reactor volume), $p_{i_{max}}$ is the maximum specific growth rate, z is the vertical position in the bioreactor between the top (z = 0) and the bottom ($z = z_m$) surface, L_i is the loss rate, assumed to be dominated by the dilution rate in a continuous bioreactor, a parameter that can easily be directly controlled. The loss rate will be a key parameter to study in this work, as it directly affects the steady-state concentration of cells which in turn influences the extent of light utilisation by the algal culture. f_{I_i} is referred to as the light factor, indicating the extent to which the algal growth is affected by light. During this work, we assume cells are homogeneously distributed throughout the bioreactor.

Light attenuation is modelled by applying the Beer-Lambert law:

$$I(\lambda, z) = I_{in}(\lambda)\exp(-\sum_{i=1}^{n} k_i(\lambda)N_i z) \quad (3.2)$$

$I(\lambda, z)$ represents the light intensity of wavelength λ at depth z, $I_{in}(\lambda)$ is the incident light intensity of wavelength λ, $k_i(\lambda)$ is the specific light absorption spectrum of

species *i*. The amount of light absorbed by species *i* at depth *z* over the photosynthetic active radiation (PAR) wavelength range of 400-700 nm, denoted as $\gamma_i(z)$, is

$$\gamma_i(z) = \int_{400}^{700} I(\lambda, z) k_i(\lambda) d\lambda \tag{3.3}$$

$\gamma_i(z)$ is light absorbed by species *i* at depth z, on a per cell basis.

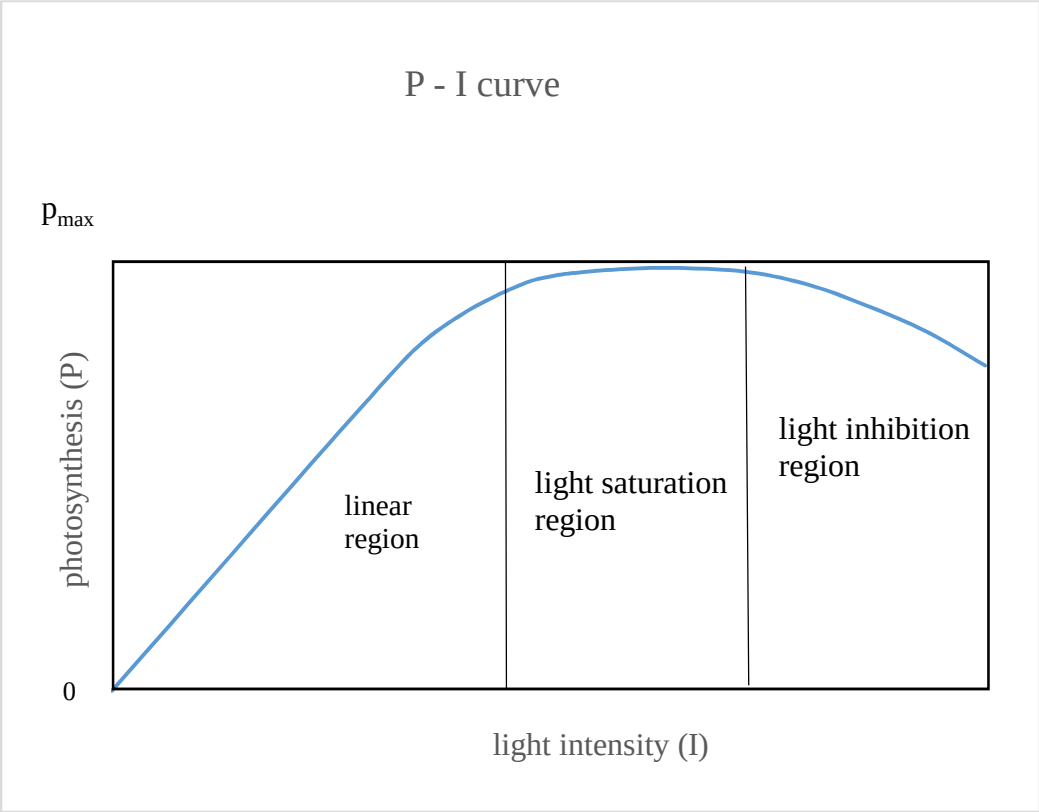


Figure 3. 1. Typical relationship between photosynthetic rate and light intensity.

To quantify the light factor f_{I_i} introduced in equation (3.1), the typical relationship between photosynthetic rate for microalgae (denoted by '*P*') and light intensity (denoted by '*I*') is considered, as shown in Figure 3.1. Three regimes are

distinguished according to the level of light intensity, namely linear, saturation and inhibition, as introduced in section 2.1.1.2.

The above *P-I* curve is widely adopted when no details of light absorption spectrum are considered. In this work, it is adapted by replacing light intensity (I) with light absorption (γ) to allow the photosynthetic rate to be directly related to the actual amount of absorbed light, which is similar to the modelling approach in a previous work [13]. Therefore, the above *P-I* curve essentially becomes a *P-*γ curve. Quantitatively, light factor in equation (3.1) is modelled as a nonlinear function of light absorption, reflecting the three regimes mentioned above and adapting a model previously formulated for the *P-I* curve [17]:

$$f_{I_i}(z) = \frac{\gamma_i(z)}{a_i\gamma_i(z)^2 + b_i\gamma_i(z) + c_i} \tag{3.4}$$

The three parameters are calculated by $a_i = \frac{p_{i_{max}}}{\gamma_{i_{opt}}^2 \phi_i}$, $b_i = 1 - \frac{2p_{i_{max}}}{\phi_i \gamma_{i_{opt}}}$, $c_i = \frac{p_{i_{max}}}{\phi_i}$, $\phi_i = p_{i_{max}} \left.\frac{df_{I_i}}{d\gamma}\right|_{\gamma=0}$ is the light use efficiency corresponding to the initial slope of the *P-*γ curve. $\gamma_{i_{opt}}$ is the optimum level of light absorption corresponding to the maximum photosynthetic rate.

3.2.2. Investigating a two-species community with simplified light absorption spectra

To theoretically assess the biomass overyielding potential due to complementary light absorption, it is desirable to obtain easy control over the degree of complementarity between different species. To this end, a hypothetic community of two species with a

pair of simplified light absorption spectra was considered. Each spectrum comprises two parts of equivalent size in terms of the PAR wavelength range, i.e. 400-550 nm and 550-700 nm, respectively. In the following, subscripts 1 and 2 refer to the two parts of light, and *a* and *b* are used to represent species *a* and *b*. Thus light intensity is divided into I_1 and I_2 corresponding to the wavelength 400 – 550 nm and 550 – 700 nm respectively. k_{ij} (i=1, 2; j=*a*, *b*) represents the light absorption spectrum of species j on part i of the light. The degree of complementarity between the two absorption spectra is controlled as follows:

$$k_{1a} = k_{2b} = 0.5 \times (1 + q)k_{max} \tag{3.5}$$

$$k_{2a} = k_{1b} = 0.5 \times (1 - q)k_{max} \tag{3.6}$$

In equations (3.5) and (3.6), k_{max} represents the upper limit of light absorption spectrum. q is an adjustable parameter and imposes an increasing degree of complementarity when its value increases from 0 to 1. An illustrative example corresponding to a medium level (q = -0.5) of complementarity is shown in Figure 3.2.

The default parameters that are used in the following work are shown in Table 3.1. The values for maximum specific growth rates, $p_{a_{max}}$ and $p_{b_{max}}$, are based on [13]. The upper limit of light absorption spectrum k_{max} is set so as to be consistent, in the order of magnitude, with the real light spectra reported in [13]. Given an average overall light absorption spectrum of 2×10^{-13} m^2 cell^{-1} as shown in Figure 3.2, the optimal light absorption $\gamma_{i_{opt}}$ is set to correspond to an optimal light intensity (i.e. the level

corresponding to the maximum growth rate) of 500 μmol photons m^{-2} s^{-1}. The literature reporting light saturation and inhibition in algal culture shows wide-ranging, species-dependent optimal light intensities ([18]). In this work, the absolute level of the optimal light intensity (and the corresponding optimal light absorption) does not matter per se; what *is* of importance is the adoption of incident light intensities in the numerical simulations with reference to the optimal level of light intensity such that all the three regimes (linear, saturation and inhibition) are tested. Besides, values smaller than those in [13] were adopted for ϕ_a and ϕ_b which, in combination with the highest incident light intensity adopted, allow the growth rate model to predict a rather clear inhibition effect, to ensure that the impact of this effect is well captured in the simulation analysis.

Symbol	value	Units
ϕ_a	4×10^{10}	cells day^{-1} (μmol photons s^{-1})$^{-1}$
ϕ_b	4×10^{10}	cells day^{-1} (μmol photons s^{-1})$^{-1}$
z_m	7	cm
$\gamma_{i_{opt}}$	1×10^{-10}	μmol photons s^{-1} $cell^{-1}$
$p_{a_{max}}$	1	day^{-1}
$p_{b_{max}}$	1	day^{-1}
k_{max}	4×10^{-13}	m^2 $cell^{-1}$

Table 3. 1. Default parameters used in simplified absorption spectra.

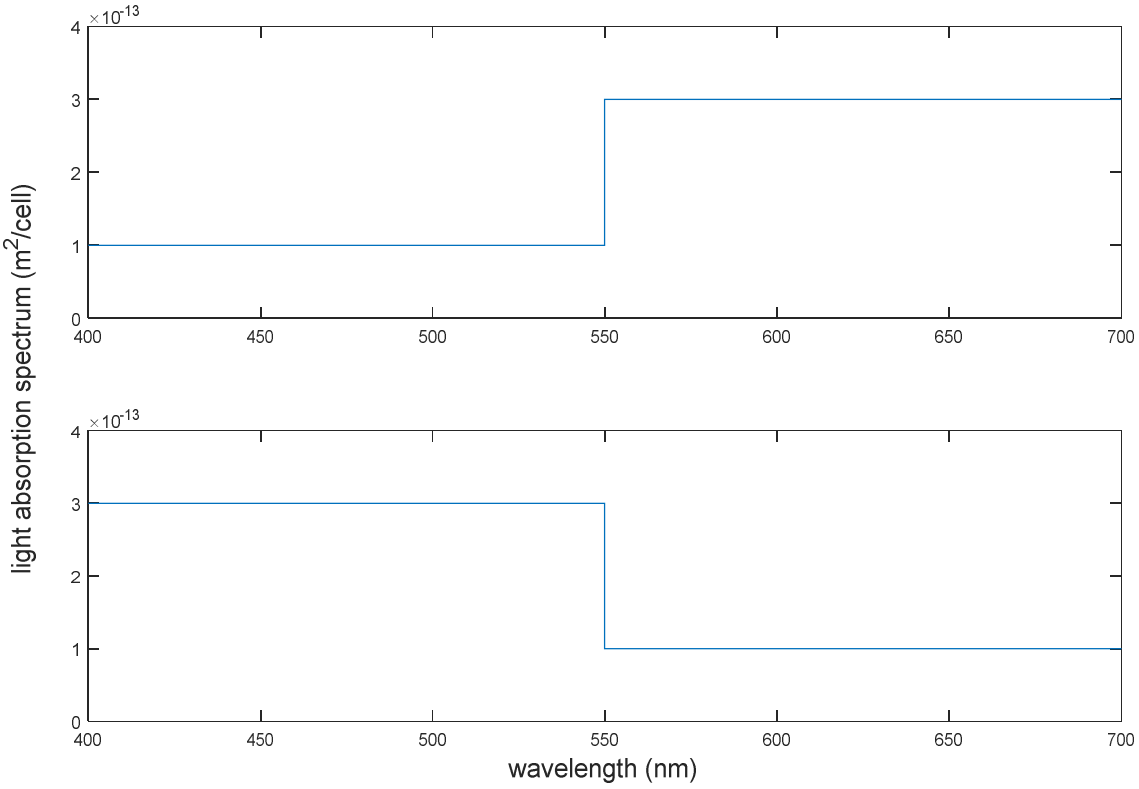


Figure 3. 2. Illustration of simplified light absorption spectra of species a (upper plot) and b (lower plot).

From equations (3.1-3.3), the factors that may affect the yield of co-culture, hence the degree of overyielding, include operation settings such as incident light intensity and cell density (which is largely controlled via loss rate), as well as species traits of the microalgae involved, such as light absorption spectrum, light use efficiency, and maximum growth rate. In this work, three different incident light intensities, namely 40, 400 and 800 μmol photons m^{-2} s^{-1} were considered. In reality, algal cultivation systems may be operated with a diverse and wide range of light intensities ([19] [20]). As explained earlier, the absolute light intensities adopted in this work are not of importance; these three levels are selected only for the purpose of locating the system

in the linear, saturation and inhibition regions of the $P-\gamma$ curve, respectively. Finally, light absorption spectra with various values of *q* were adopted in numerical experiments, together with the variation of the other parameters listed in Table 3.1.

3.2.3 Investigating a community of real species

Following the analysis using idealized absorption spectra, a community of two real species was considered to verify the findings from the former step. These species are chosen from the *synechococcus* group isolated from Baltic Sea named BS4 and BS5 as described in [13]. BS4 is a green picocyanobacteria containing pigment phycocyanin which absorbs red light whereas BS5 is a red picocyanobacteria with pigment phycoerythrin which can absorb blue and green light. Same as in [13], the reactor depth was set to z_m = 7 cm and an incident light intensity of 40 μmol photons m^{-2} s^{-1} (integrated over the range of 400 – 700 nm) was adopted. Monoculture of each species and co-culture of both species were simulated under two different loss rates to find out if and under which conditions overyielding can happen.

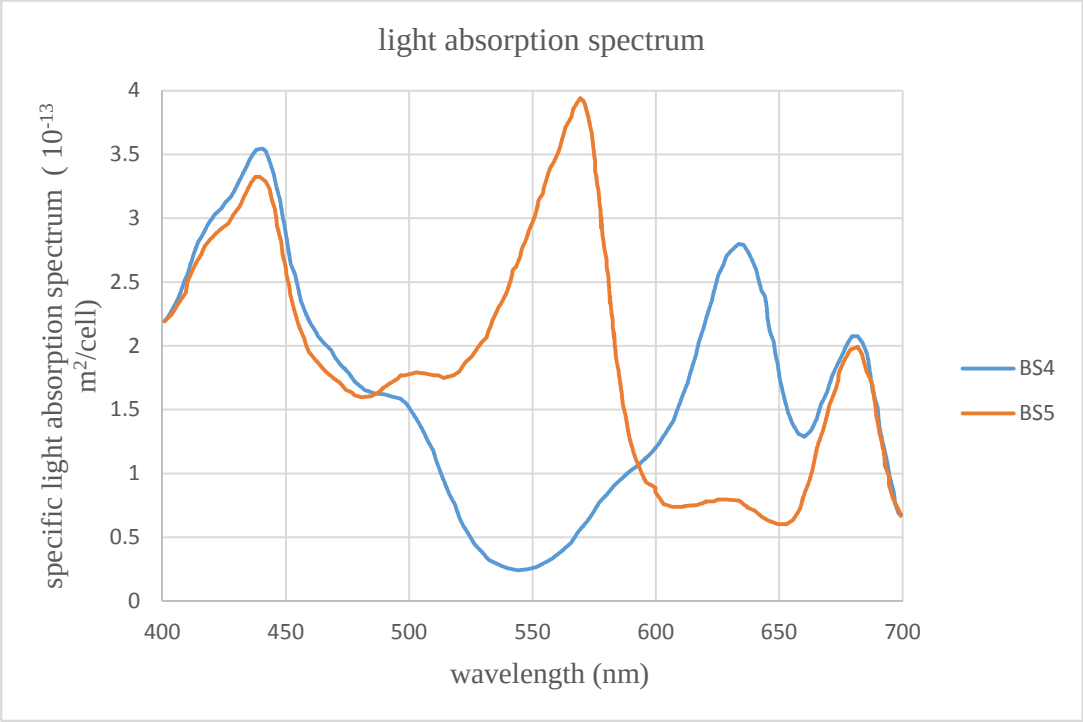


Figure 3. 3. Light absorption spectra of two species [13]. BS4 and BS5 represent (blue) picocyanobacteria and (red) picocyanobacteria, respectively.

3.3 Results and discussion

In this section, the results of numerical simulation of the system with the simplified spectra will be presented and analysed first, to show the impact of operation settings and species traits on overyielding potential when the system is subject to varying degrees of light absorption complementarity. The simulation results of the system with two real species will then be presented to offer a comparison.

3.3.1 Impact of operation settings

Assuming that the two algal species have the identical maximum specific growth rate and light use efficiency, simulation of the system was carried out to investigate the impact of different operation settings, in terms of incident light intensity and cell density. The overyielding potential affected by cell density (controlled via the loss rate) at different incident light intensities is indicated in Figure 3.4. The normalised overyielding level is calculated by the difference between the yield of co-culture and that of monoculture divided by the yield of mono-culture. Note that the latter is identical for the two species due to the fact that the only difference between the species *a* and *b* is their light absorption spectra which are symmetric to each other, therefore the two species will have an identical monoculture yield. A negative overyielding level implies the yield of the coculture is worse than the monocultures, which is termed "underyielding". From Figure 3.4, one can see that when incident light intensity is low (corresponding to the linear region of the light factor model, cf. equation 3.4), higher loss rates (lower cell density) and higher light absorption complementarity (i.e. q) tend

to result in overyielding. When incident light intensity is increased to the light saturation and light inhibition region, underyielding can occur particularly when the loss rate (cell density) is sufficiently low (high). In such cases, there exists a lowest yield (with underyielding) at a certain level of q.

The above result shows that, in a system operating at a light-saturation or light-inhibition regime with a high cell density, the yield of a co-culture may become increasingly worse than the monoculture while the complementarity in light absorption between the two species increases; this trend gets reversed and eventually overyielding starts to occur and improve but only after the complementary exceeds a certain level. To explain this rather counter-intuitive observation, Figure 3.5 shows plots of the specific growth rate curves of co-culture and mono-culture as a function of cell density, at different light absorption complementarity (i.e., q) and for a system operating in the light saturation regime. For each curve, its intersection with one of the two (low and high) loss rate lines represents the steady-state solution of the corresponding system. From Figure 3.5, one can see that at the steady state resulting from a higher loss rate, which gives rise to a lower cell density, the yield of the co-culture system is higher than that of the mono-culture system; this gap tends to increase with q. In terms of the specific growth rate, the monoculture would have a lower value than that of the co-culture at the steady-state cell density of the latter, indicating that the former would be less productive. Moving to the lower loss rate which corresponds to a higher cell density, which is the case leading to the counter-intuitive observation as presented earlier, one

can see that the steady-state cell density of the co-culture is lower than that of the mono-culture at low q values, with this gap initially increasing with q and then reversed afterwards. At a point where the co-culture yields less than the monoculture, the specific growth rate of the mono-culture would be higher than that of the co-culture at the steady-state cell density of the latter, indicating that the former would be more productive in this case.

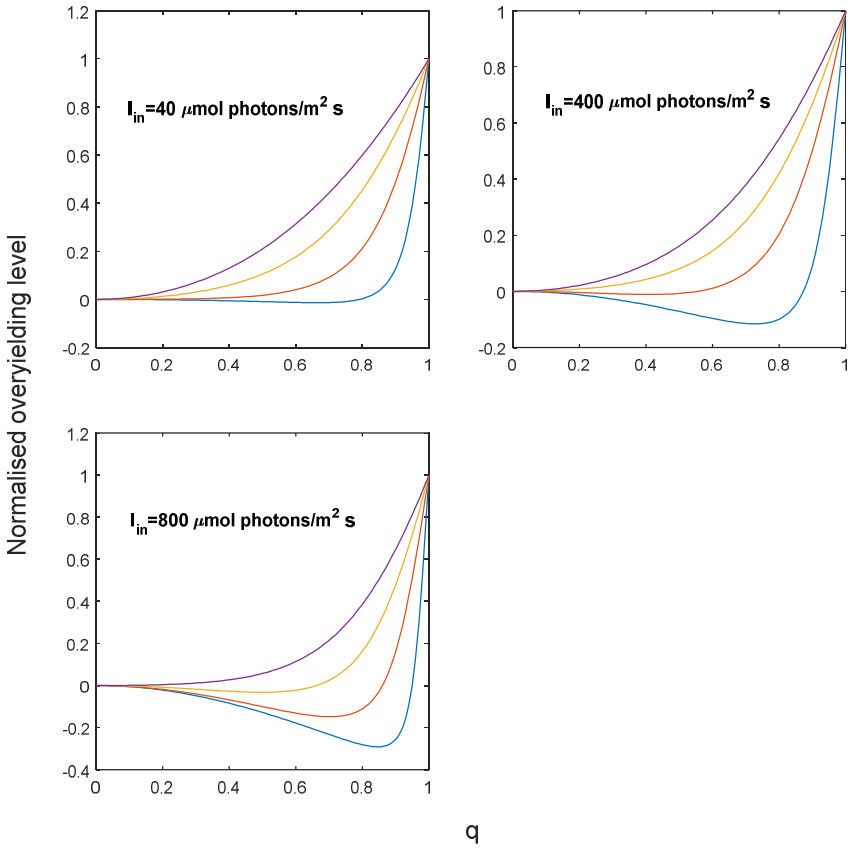


Figure 3. 4. Predicted overyielding level as a function of light absorption complementarity at different loss rates. On the top left graph, the loss rate is, from the bottom to the top, 0.02, 0.05, 0.1 and 0.2 day^{-1}. On the other two graphs, the loss rate is, from the bottom to the top, 0.2, 0.4, 0.6 and 0.8 day^{-1}.

How could the monoculture system become more productive than the co-culture system when a significant level of light absorption complementarity is present? The answer to this question appears to lie in the trade-off between the magnitude of light absorption and the degree of the saturation/inhibition effect. Using equation 3.4, this effect, denoted as η, is the denominator of the light factor expression:

$$\eta = \frac{1}{a_i\gamma_i(z)^2+b_i\gamma_i(z)+c_i} \tag{3.9}$$

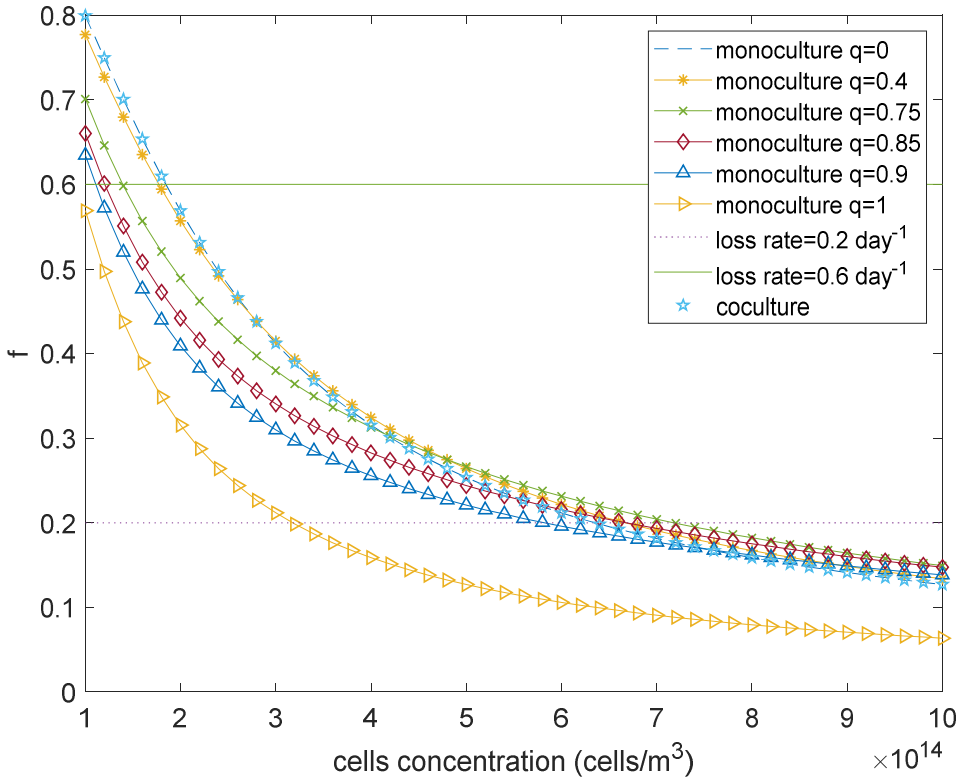


Figure 3. 5. Light factor predicted at different cells concentrations, affected by light absorption complementarity, loss rate and cultural type.

Clearly, η decreases with the increase of the light absorption γ; the decrease becomes increasingly pronounced when moving away from the linear region. Thus, the effect of

increased light absorption (which could be achieved by a co-culture of species with complementary light absorption) on light factor (hence the specific growth rate, see equation 3.1) is determined by the trade-off between the increase in γ and the decrease in η. To show these relations across different depth positions in a photobioreactor, Figure 3.6 presents multiple plots of a co-culture system and a mono-culture system at a medium cell density ($N = 5\times10^{14}$ cells/m^3) and a level of light absorption complementarity ($q = 0.75$) which falls in the region where, according to Figure 3.5, the co-culture system tends to be less productive compared to the mono-culture counterpart. One can see from Figure 3.6a that throughout the entire reactor, both parts of the light are equally and nearly completely absorbed by the co-culture, due to the complementarity of light absorption between the two species. In comparison, the monoculture can (nearly) completely absorb only one part of the light (denoted by the lower dashed curve in Figure 3.6a), but not the other part (denoted by the upper dashed curve in Figure 3.6a), thus leading to lower overall light absorption, as expected. In 3.6a, microalgae under monoculture has different ability to absorb different parts of the light, which leaves two parts of light intensity different in the pond (shown with two dashed lines).

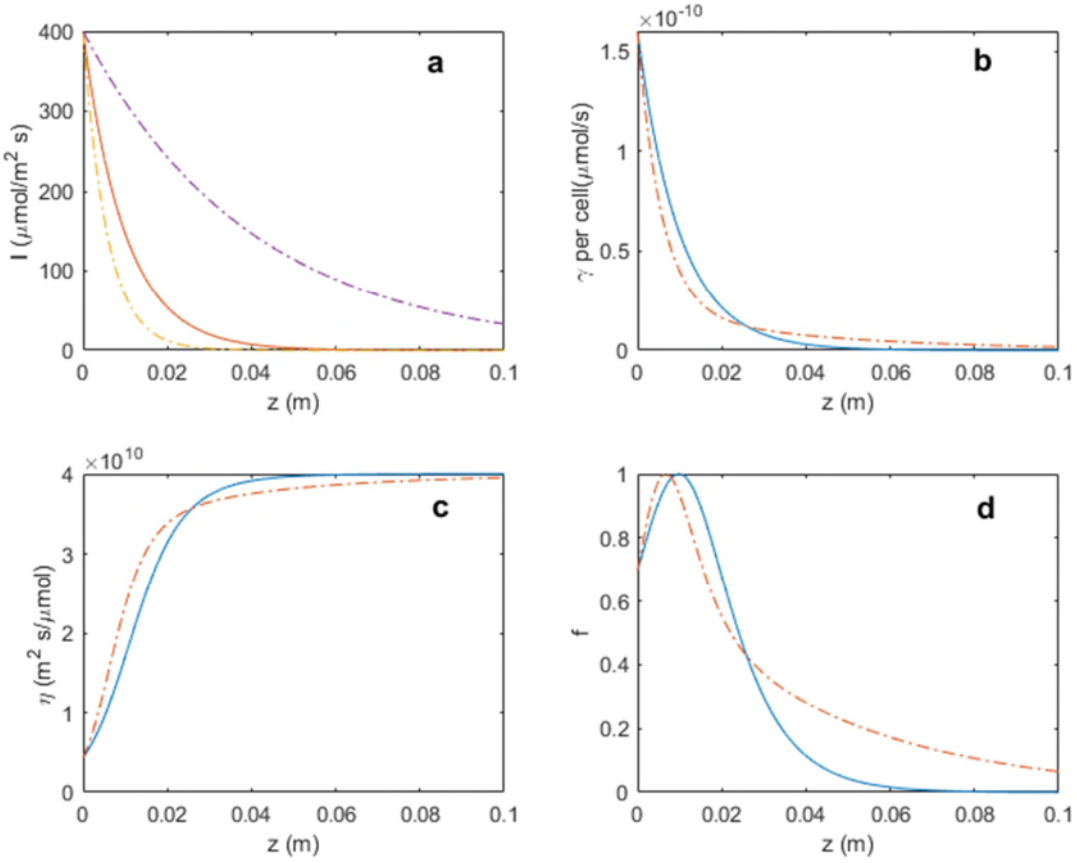


Figure 3. 6. Predicted changes of light intensity (a), light absorption (b), saturation/inhibition effect (c) and light factor (d) at different depths of the reactor. Dashed and solid lines represent monoculture and co-culture, respectively.

In terms of γ (Figure 3.6b), over the top layer of the reactor where most of the light is absorbed, light absorption achieved by the co-culture system is higher than that in the mono-culture system; it becomes lower afterwards only as the consequence of very limited amount of light available in the remaining layer of the reactor due to the stronger absorption in the upper layer. On the other hand, the saturation/inhibition effect, η, changes in an opposite fashion, as shown in Figure 3.6c. The trend of the combined influence of γ and η is manifested in the light factor as shown in Figure 3.6d. Numerical

integration across the depth of the bioreactor shows that the overall light factor (hence the value of the overall specific growth rate) of the monoculture is higher than that of the latter. This set of plots thus demonstrates the effect of the trade-off between γ and η and its impact on productivity, hence answering the question raised earlier.

3.3.2 Impact of species traits

Systems with two species with non-identical biological attributes, including light use efficiency (ϕ) and maximum specific growth rate (P_{max}), have been considered. From equation 1, increase in each of the two parameters will lead to the increased growth rate of the microalga for a given amount of absorbed light. Our simulation results show that the impact of the two parameters on the overyielding potential of the co-culture system under various operation settings and degrees of light absorption complementarity is similar. Therefore, only the impact of ϕ is presented and analysed here.

In Section 3.3.1, the photosynthetic efficiencies of both species, ϕ_a and ϕ_b, were set to 4×10^{10}. In this section, ϕ_b remains unchanged, while three different values for ϕ_a, 4.2, 4.5 and 5×10^{10}, are adopted in simulation studies. As the two species adopt different photosynthetic efficiencies, their monocultures will have different yields. In such circumstances, the normalised overyielding level is defined as the difference of the yield of the co-culture and the yield of the most productive monoculture divided by the yield of most productive monoculture. With this definition, a negative value of overyielding level may or may not mean underyielding, if this term refers to the cases

where the yield of co-culture is lower than the worst yield of monoculture of any species; this will depends on whether the yield of co-culture lies between the yields of the two monocultures or below the lower one.

Simulation results with different values for ϕ_a, incident light intensity and light absorption complementarity are summarised in Figure 3.7. The most noticeable change compared to the results presented in Section 3.3.1, caused by the difference in the species traits (more specifically, light use efficiency as simulated here) of the two species, is that the different competitiveness of the two species with respect to light conversion leads to the failure of co-existence when light absorption complementarity (q) is low. As further indicated in Figure 3.8a, the minimum level of q for supporting co-existence (denoted as q_{min}) increases with the difference between the light use efficiency of the two species, but decreases when the incident light intensity is higher. The loss rate also affects q_{min}, as shown in Figure 8b; a higher loss rate demands for a higher level of light absorption complementarity to achieve co-existence.

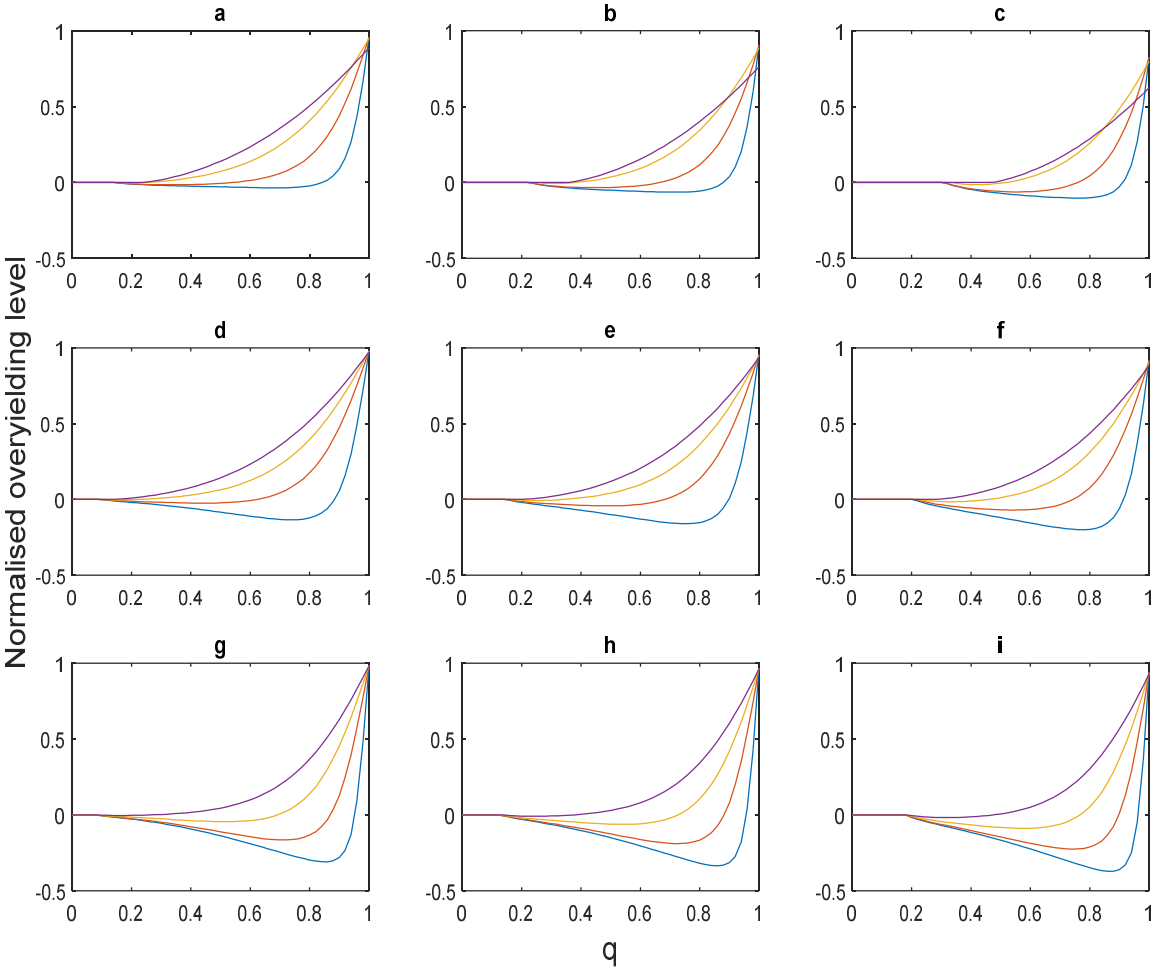


Figure 3. 7. Predicted overyielding level changes in co-culture of species with different light use efficiencies. The incident light intensities are 40, 400, 800 μmol photons s^{-1}/m^2 s for the 1st, 2nd and 3rd rows, respectively. ϕ_a for the 1st, 2nd and 3rd columns are 4.2×10^{10}, 4.5×10^{10} and 5×10^{10} respectively. The loss rates are 0.02, 0.05, 0.1 and 0.2 day^{-1} for curves in each plot of the first row, from bottom to top, respectively, and 0.2, 0.4, 0.6 and 0.8 day^{-1} for the other two rows.

Under conditions where the two species can co-exist, the influence of various settings on overyielding under different values of ϕ_a is similar to that observed in

Section 3.3.1. A lower cell density generally tends to promote overyielding. At a high cell density, negative overyielding occurs and worsens with the initial increase of light absorption complementarity (q) until the "worst" value of q is exceeded; from that point on the degree of overyielding will improve.

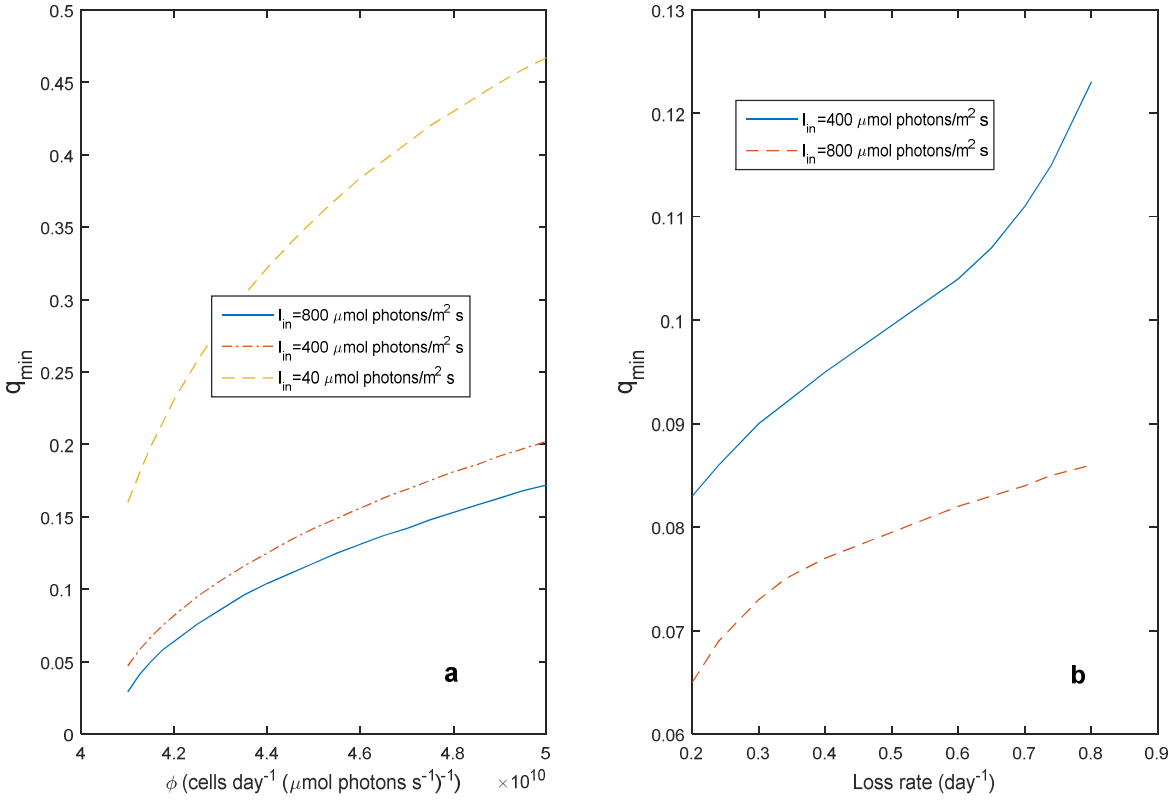


Figure 3. 8. Minimum level of light absorption complementarity required for co-existence. (a) assumes loss rate = 0.2 day^{-1}. (b) assumes $\phi_a = 4.2 \times 10^{10}$.

3.3.3 Simulation of a community of two species with real spectra

Following the simulation studies on the hypothetic system with simplified light absorption spectra, two real species (BS4 and BS5) with the light absorption spectra

shown in Figure 3.3 are used in the final set of simulation. Other parameters were adopted from [14] and shown in Table 3.2. The simulation results for two different loss rates are shown in Figure 3.9.

One can see from the simulation results in 3.3.1 and 3.3.2 that the co-culture of the two species with real light absorption spectra could result in overyielding and underyielding at a higher and lower loss rate (i.e. a lower and higher cell density), respectively. It is consistent with the findings from the simulation of the two-species community with simplified light absorption spectra, as presented earlier.

Symbol	Value	Unit
ϕ_a	2×10^{12}	cells day^{-1} (μmol photons s^{-1})$^{-1}$
ϕ_b	2×10^{12}	cells day^{-1} (μmol photons s^{-1})$^{-1}$
z_m	7	cm
I_{in}	40	μmol photons m^{-2} s^{-1}
$\gamma_{i_{opt}}$	1×10^{-10}	μmol photons s^{-1} $cell^{-1}$
$p_{a_{max}}$	1	day^{-1}
$p_{b_{max}}$	1	day^{-1}

Table 3. 2. Parameter values adopted in simulation of real species.

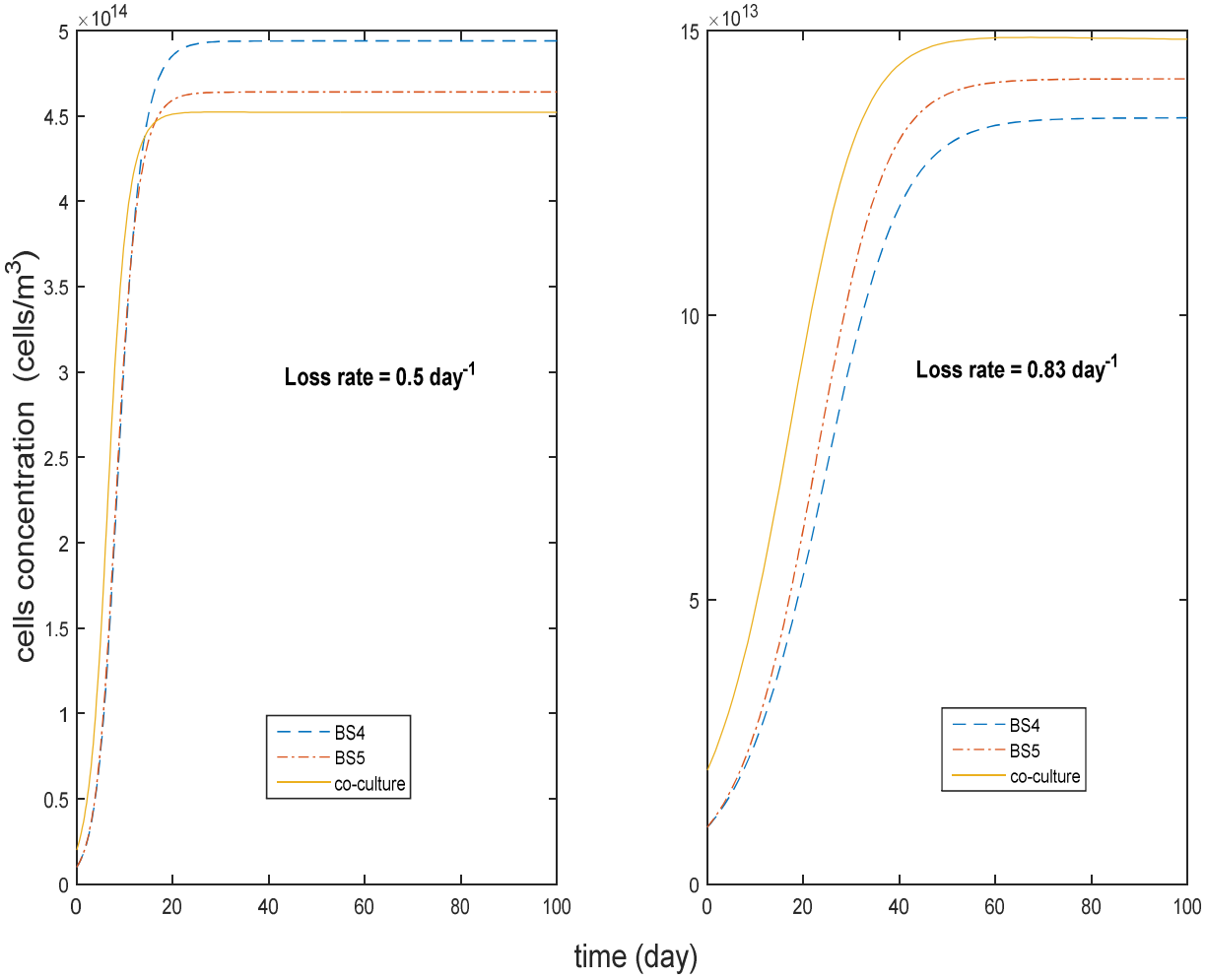


Figure 3. 9. Simulation results of the cultivation of BS4 and BS5.

3.3 Conclusion

In this work, conditions for biomass overyielding have been investigated for co-cultures of microalgal species exhibiting complementarity in light absorption. Numerical simulation indicates that complementary light use alone does not guarantee overyielding; both operation settings and species traits also play an important role. When a difference exists in the light use efficiency or the maximum growth rate of different species, co-culture might not be feasible at all due to the co-existence issue.

In such circumstances, co-existence needs to be supported by a minimum level of complementarity in light absorption; the required level varies with the agree of inter-species difference in light use efficiency (or maximum growth rate) as well as operating conditions such as loss rate and incident light intensity. When co-existence is possible, overyielding can be affected by both species traits and operation settings. Generally, a lower cell density (achieved by a higher loss rate) and an increasing degree of light use complementarity tend to better promote overyielding. However, in a system operating in the light saturation or inhibition regime and with a higher cell density, underyielding may occur and worsen when the complementarity initially increases; this trend reverses only after the level of complementarity becomes sufficiently high which eventually will bring about overyielding. These observed trends appear to stem from the trade-off between light absorption and effect of light saturation/inhibition.

In summary, this work has shown that overyielding in a polycultute of multiple microalgal species from light absorption complementarity alone is conditional and could occur only when these species are with favourable traits and are cultured under suitable operating settings. When such conditions are absent, experimental observations of overyielding by a polyculture system will have to be attributed to other factors than complementary light absorption, which should be focused on to guide the successful application of such systems for bioenergy and other purposes.

References

[1] M. Hannon, J. Gimpel, M. Tran, et al., Biofuels from algae: challenges and potential, Biofuels 1 (5) (2010) 763–784.

[2] L. Brennan, P. Owende, Biofuels from microalgae—a review of technologies for production, processing, and extractions of biofuels and co-products, Renew. Sustain. Energy Rev. 14 (2) (2010) 557–577.

[3] R. Slade, A. Bauen, Micro-algae cultivation for biofuels: cost, energy balance, environmental impacts and future prospects, Biomass Bioenergy 53 (2013) 29–38.

[4] L. Wang, M. Min, Y. Li, et al., Cultivation of green algae Chlorella sp. in different wastewaters from municipal wastewater treatment plant, Appl. Biochem. Biotechnol. 162 (4) (2010) 1174–1186.

[5] L.E. de-Bashan, Y. Bashan, Immobilized microalgae for removing pollutants: review of practical aspects, Bioresour. Technol. 101 (6) (2010) 1611–1627.

[6] L. Meier, P. Barros, A. Torres, C. Vilchez, D. Jeison, Photosynthetic biogas upgrading using microalgae: effect of light/dark photoperiod, Renew. Energy 106 (2017) 17–23.

[7] A. Yang, Modeling and evaluation of CO2 supply and utilization in algal ponds, Ind. Eng. Chem. Res. 50 (19) (2011) 11181–11192.

[8] M. Stockenreiter, F. Haupt, A.K. Graber, et al., Functional group richness: implications of biodiversity for light use and lipid yield in microalgae, J. Phycol. 49 (5) (2013) 838–847.

[9] E. Kazamia, A.S. Riseley, C.J. Howe, et al., An engineered community approach for industrial cultivation of microalgae, Ind. Biotechnol. 10 (3) (2014) 184–190.

[10] S. Behl, A. Donval, H. Stibor, The relative importance of species diversity and functional group diversity on carbon uptake in phytoplankton communities, Limnol. Oceanogr. 56 (2) (2011) 683–694.

[11] J.O. Nalley, M. Stockenreiter, E. Litchman, Community ecology of algal biofuels: complementarity and trait-based approaches, Ind. Biotechnol. 10 (3) (2014) 191–201.

[12] G. Stephanopoulos, A.G. Fredrickson, Coexistence of photosynthetic microorganisms with growth rates depending on the spectral quality of light, Bull. Math. Biol. 41 (4) (1979) 525–542.

[13] M. Stomp, J. Huisman, L. Vörös, et al., Colourful coexistence of red and green picocyanobacteria in lakes and seas, Ecol. Lett. 10 (4) (2007) 290–298.

[14] M. Stomp, J. Huisman, F. De Jongh, et al., Adaptive divergence in pigment composition promotes phytoplankton biodiversity, Nature 432 (7013) (2004) 104–107.

[15] J. Huisman, F.J. Weissing, Competition for nutrients and light in a mixed water column: a theoretical analysis, Am. Nat. 146 (1995) 536–564.

[16] J. Huisman, F.J. Weissing, Light-limited growth and competition for light in well-mixed aquatic environments: an elementary model, Ecology 75 (2) (1994) 507–520.

[17] P. Duarte, J.G. Ferreira, Dynamic modelling of photosynthesis in marine and estuarine ecosystems, Environ. Model. Assess. 2 (1997) 83–93.

[18] H.Y. Lee, L.E. Erickson, S.S. Yang, Kinetics and bioenergetics of light-limited photoautotrophic growth of Spirulina platensis, Biotechnol. Bioeng. 29 (7) (1987) 832–843.

[19] N. Kurano, S. Miyachi, Selection of microalgal growth model for describing specific growth rate-light response using extended information criterion, J. Biosci. Bioeng. 100 (4) (2005) 403–408.

[20] H.L. MacIntyre, T.M. Kana, T. Anning, R.J. Geider, Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria1, J. Phycol. 38 (1) (2002) 17–38.

Chapter 4. Modelling and analysis of microbial community based bioreactors for hydrogen production

Summary

This chapter is the second case study for model-based analysis of microbial communities, which considers multiple types of ecological interactions that co-exist. Two specific systems are investigated which produce hydrogen from the decomposition of cellulose: a two-species system with *C.cellulovorans* and *R.palustris* and a three-species system where a hypothetical *E.coli* is introduced additionally with the hope to improve hydrogen production. The aim of this study is to understand the dynamics of the bioreactors and the factors affecting their hydrogen production and stability, particularly the impact and trade-offs associated with the introduction of additional species.

Monod-type equations are applied as the basis to describe the uptake/growth rate of all microorganisms. The pH inhibition is modelled by a sigmoid-type equation. All parameters are fitted according to literature data. Simulation is conducted for both batch and continuous reactors, and bifurcation analysis is carried out for the cases of continuous operation with inlet cellulose concentration as the free parameter to be varied.

For the two species system, the analysis has revealed a bi-stable region and the associated risk for the system to shift, when exposed to disturbance, from a high yield

state to a low yield state. The three species system shows improvement in both stability and productivity; a bi-stable region still exists but the productivity gap between the two stable states is much smaller than the two-species case which means switching from one state to the other would not lead to a significant change of hydrogen production. Among other factors, relative pH tolerance of different species is recognised as having a significant impact on both stability and productivity.

This case study shows that a relatively simple microbial community can present sophisticated behaviours which are potentially important for the design of the community and the operation of the bioreactor.

This chapter comprises the full report corresponding to the following conference abstract published for an oral presentation:

Model-based analysis and optimisation of hydrogen production with a synthetic microbial community. Sihao Di, Aidong Yang, 2017. World Congress of Chemical Engineering, Barcelona, Spain.

Abstract is shown in Appendix.

List of abbreviations and symbols:

Abbreviations and subscripts:

Ac: acetate

Bu: butyrate

c: cellulose

Cc: *C. cellulovorans*

Fc: formate

HEC: hypothetical *E. coli*

La: lactate

Rp: *R.palustris*

Symbols:

S: concentration of substrates

X: concentration of biomass

$I_{pH_{Cc}}$: pH inhibition term for C.cellulovorance

$I_{pH_{HEC,Ac}}$: pH tolerance of HEC on uptaking acetate

$I_{pH_{HEC,Fc}}$: pH tolerance of HEC on uptaking formate

$I_{pH_{Rp}}$: pH inhibition term of *R. palustris*

$K_{1_{Cc}}$, $K_{pH_{Cc}}$: constants

$K_{1_{HEC}}$, $K_{2_{HEC}}$, $K_{HEC,Fc,1}$, pH_{UL} and pH_{LL}: constants.

$K_{1_{Rp}}$ and $K_{pH_{Rp}}$: constants

$K_{Cc,c}$: half saturation constant for cellulose uptake for *C. cellulovorans*

$K_{j,HEC}$: half saturation constant for HEC

$K_{Rp,j}$: half saturation constant of *R. palustris* on VFA *j*

p_{VFA_k}: production rate of VFA *k* by *C. cellulovorans*

$r_{Cc,c}$: update rate of cellulose for *C. cellulovorans*

$r_{Cc,c,max}$: maximum uptake rate of cellulose for *C. cellulovorans*

$r_{H_2{}_{Cc}}$: hydrogen production rate

$r_{H_2{}_{HEC}}$: hydrogen production by HEC

$r_{j,HEC}$: HEC growth rate

$r_{j,HEC,max}$: maximum uptake rate of HEC on VFA *j*

$r_{Rp,j}$: consumption rate of *R. palustris* on each of the VFAs

$r_{Rp,j,max}$: maximum uptake rate of *R. palustris* on VFA *j*

$Y_{Cc/_c}$: cellulose-to-*C. cellulovorans* yield coefficient

$Y_{H_2/_c}$: cellulose-to-hydrogen yield constant for *C. cellulovorans*

$Y_{H_2,HEC/_{Ac}}$, $Y_{H_2,HEC/_{Fc}}$: VFAs-to-H2 yield constants of HEC

$Y_{H_2,Rp/_{Ac}}$, $Y_{H_2,Rp/_{Bu}}$ and $Y_{H_2,Rp/_{La}}$: VFAs-to-hydrogen yield constants of *R. palustris*

$Y_{HEC/_{Ac}}$: acetate-to-HEC yield coefficient

$Y_{Rp/_{Ac}}$, $Y_{Rp/_{Bu}}$ and $Y_{Rp/_{La}}$: VFAs-to-*R. palustris* yields constant

$Y_{VFA,k}$: yield coefficient for producing VFA *k* for *C. cellulovorans*

μ_{Cc}: growth rate of *C. cellulovorans*

μ_{HEC}: specific growth rate of HEC

4.1. Introduction

Naturally occurring communities of microorganisms have long been used for engineering purposes, for example in anaerobic digesters for treating waste organic streams (Gunaseelan, 1997; Mata-Alvarez et al., 2000; Appels et al., 2008). In recent years, a number of attempts have been made in developing bioprocesses based on artificially formed consortia of multiple microorganisms, referred to as synthetic microbial communities, to gain advantages over the conventional, mono-culture based processes (Rollié, et al., 2012; Jia et al., 2016). On the other hand, such systems inevitably involve a greater level of complexity, originated from the interactions between the multiple species, which requires careful analysis and design in order to achieve their intended engineering objectives.

In order to develop a better understanding towards the inter-species mechanisms, theoretical analysis based on either analytical models (Sari and Harmand, 2016) or numerical simulation (Yoshida et al., 2003) plays a very important role. Existing literature has reported a number of investigations on the interactions of species within natural or artificial communities, particularly focusing on inter-species competition. It is well-known that two microbial species competing for a single rate-limiting substrate are not able to co-exist under steady-state operation (Tilman, 1980), however, there are specific operational settings which are able to make the co-existence possible (Dikshitulu et al., 1993; Pavlou et al., 1990; Lenas and Pavlou, 1994). Besides, a microbial consortium involving plasmid-bearing and plasmid-free cell populations

competing on a single resource can lead to co-existence by extra mechanisms with which plasmid-bearing organisms lose or modify the recombinant DNA during reproduction (Ajbar, 2002). Apart from competition, a two-species synthetic symbiotic ecosystem was studied, where quorum sensing imposed a significant impact on the stability of the consortium (Kambam et al., 2008). Existing literature also investigated ecosystems with more than two species, such as a three-species system competing on two resources (Thomopoulos et al., 1998). Besides, several studies have analysed syntrophic consortia where producers and consumers critically interdepend on each other, particularly in the anaerobic digestion process (Wade et al., 2016; Weedermann et al., 2015).

The above studies have mostly focused on the stability of microbial communities within the setting of a bioreactor; little attention has been given at the same time to the efficiency or productivity of the bioreactor which is naturally of equal importance for engineering applications. Besides, the analysed systems commonly involve a single type of relationship between any two species in a community, e.g. competition or syntrophy, whereas in an artificially designed engineering system this is not necessarily the case, therefore an understanding of the impact of more complex interactions is needed.

In this work, an in silico analysis is carried out on the productivity and stability of microbial community based conversion of cellulose to hydrogen. Based on the experimental data for the co-culture of C. cellulovorans and R. palustris (Lu and Lee,

2015), a hypothetical third species is proposed on the basis of literature data of E. coli, with the intention to further improve the performance of hydrogen production. Kinetics of these species is modelled and calibrated against available data. The productivity of the two-species and the three-species system is predicted for both batch and continuous operations, as dictated by the sophisticated inter-species interactions within specific bioreactor settings. For the continuous operation, the dynamic behaviour and stability of each system are studied through bifurcation analysis and dynamic simulation, to reveal the impact of operational parameters and species traits.

4.2. Modelling kinetics of individual species

Monod type equations are adopted to represent the kinetics of microorganisms. *C. cellulovorans* uptakes cellulose and produces hydrogen and VFAs (Volatile Fatty Acids). Throughout the work, S is used to denote substrates and X is used to denote biomass. The update rate of cellulose, $r_{Cc,c}$ (hr^{-1}) is:

$$r_{Cc,c} = r_{Cc,c,max} \frac{S_c}{S_c + K_{Cc,c}} I_{pH_{Cc}} \tag{4.1}$$

where $r_{Cc,c,max}$ (hr^{-1}) is the maximum uptake rate of cellulose, S_c (g/L) is the cellulose concentration, $K_{Cc,c}$ (g/L) is the half saturation constant for cellulose uptake, and $I_{pH_{Cc}}$ is the pH inhibition term. *Cc* and *c* represent *C. cellulovorans* and cellulose respectively. The production rate of VFA *k* by *C. cellulovorans*, p_{VFA_k} (mol/ (10^8 cells) hr) is:

$$p_{VFA_k} = r_{Cc,c} Y_{VFA,k} \tag{4.2}$$

where $Y_{VFA,k}$ (mol/(10^8 cells)) represents the yield coefficient for VFA *k*, *k* refers to individual VFAs, including acetate (*Ac*), butyrate (*Bu*), formate (*Fc*) and lactate (*La*). The growth rate of *C. cellulovorans*, μ_{Cc} ((10^8 cells)/g hr) is modelled by equation (4.3):

$$\mu_{Cc} = r_{Cc,c} Y_{Cc/c} \tag{4.3}$$

where $Y_{Cc/c}$ ((10^8 cells)/g) is the cellulose-to-*C. cellulovorans* yield coefficient. The pH inhibition term is:

$$I_{pH_{Cc}} = 1 - \frac{1}{1+\exp(K_{1_{Cc}}(pH - K_{pH_{Cc}}))} \quad (4.4)$$

where $K_{1_{Cc}}$ and $K_{pH_{Cc}}$ are constants. The hydrogen production rate $r_{H_{2_{Cc}}}$ (hr^{-1}) is proportional to the cellulose uptake rate by a cellulose-to-hydrogen yield constant $Y_{H_2/_c}$(ml/(g/L)):

$$r_{H_{2_{Cc}}} = r_{Cc,c} Y_{H_2/_c} \quad (4.5)$$

R. palustris grows on acetate, butyrate and lactate (Lu and Lee, 2015); the consumption rate of each of the VFAs, $r_{Rp,j}$ (hr^{-1}) is:

$$r_{Rp,j} = r_{\mathrm{Rp},j,max} \frac{S_j}{S_j + K_{Rp,j}} I_{pH_{Rp}} \quad (4.6)$$

$r_{Rp,j,max}$ (hr^{-1}) is the maximum uptake rate of *R. palustris* on VFA *j*, $K_{Rp,j}$ (mol/l) is the half saturation constant of VFA *j*, S_j (mol/l) is the concentration of VFA *j* and $I_{pH_{Rp}}$ is the pH inhibition term of *R. palustris*. The growth rate of *R. palustris* is modelled by equation (4.7); $Y_{Rp/_{Ac}}$, $Y_{Rp/_{Bu}}$ and $Y_{Rp/_{La}}$ ((10^8 cells)/mol) are VFAs-to-*R. palustris* yield constants. The hydrogen production rate is expressed in equation (4.8) where $Y_{H_2,Rp/_{Ac}}$, $Y_{H_2,Rp/_{Bu}}$ and $Y_{H_2,Rp/_{La}}$ (ml/(mol/L)) are VFAs-to-hydrogen yield constants of *R. palustris*. The pH inhibition of *R. palustris* is expressed in equation (4.9), with $K_{1_{Rp}}$ and $K_{pH_{Rp}}$ as two constants.

$$\mu_{Rp} = r_{Rp,Ac} Y_{Rp/_{Ac}} + r_{Rp,Bu} Y_{Rp/_{Bu}} + r_{Rp,La} Y_{Rp/_{La}} \quad (4.7)$$

$$r_{H_{2_{Rp}}} = r_{Rp,Ac} Y_{H_2,Rp/_{Ac}} + r_{Rp,Bu} Y_{H_2,Rp/_{Bu}} + r_{Rp,La} Y_{H_2,Rp/_{La}} \quad (4.8)$$

$$I_{pH_{Rp}} = 1 - \frac{1}{1+\exp(K_{1_{Rp}}(pH - K_{pH_{Rp}}))} \quad (4.9)$$

The VFAs produced by *C. cellulovorans* include a considerable amount of formic acid, which is not consumed by *R. palustris* and contributes significantly to the decrease of pH in the system (Lu and Lee, 2015). In order to further improve the hydrogen production process, introduction of a third organism was considered, primarily aiming to uptake formate. In particular, literature shows that *E. coli* is able to convert formate into hydrogen (Yoshida et al., 2005), and also to uptake acetate as carbon source (Trchounian et al., 2015). We thus introduce a third species modelled on *E. coli* according to these data sources, which is referred to as "hypothetic *E. coli"* or HEC. Although there is no evidence showing such a strain actually exists, HEC is introduced in this *in silico* study with two considerations:

1. It converts formate, hence fulfilling the original purpose of introducing a third species into the *C. cellulovorans – R. palustris* consortium;
2. Like *R. palustris*, it uptakes acetate as the carbon source, hence introducing a competition relationship into the system. This allows us to study the potentially complex behaviours arising from multiple types of inter-species relationship that may exist in a microbial consortium synthesised for engineering purposes.

HEC's uptake rates on acetate and formate are expressed in equation (4.10), its growth rate is by equation (4.11)

$$r_{j,HEC} = r_{j,HEC,max} \frac{S_j}{S_j + K_{j,HEC}} I_{pH_{HEC,j}} \quad (4.10)$$

$$\mu_{HEC} = r_{Ac,HEC} Y_{HEC/Ac} \quad (4.11)$$

where $r_{j,HEC,max}$ (hr^{-1}) represents the maximum uptake rate on VFA j, μ_{HEC} represents the specific growth rate of HEC; HEC is assumed to only grow on acetate. $K_{j,HEC}$ is the half saturation constant and $Y_{HEC/Ac}$ (10^8 cells/mol) is the acetate-to-HEC yield coefficient.

The hydrogen production by HEC is expressed in equation (4.12), with $Y_{H_2,HEC/Ac}$(ml/(mol/L)) and $Y_{H_2,HEC/Fc}$ (ml/(mol/L)) as yield constants:

$$r_{H_2 HEC} = r_{HEC,Ac} Y_{H_2,HEC/Ac} + r_{HEC,Fc} Y_{H_2,HEC/Fc} \quad (4.12)$$

Different pH tolerance correlations are adopted for the uptake of acetate and formate by HEC, for a better fitting to the available data:

$$I_{pH_{HEC,Ac}} = 1 - \frac{1}{1+\exp(K_{1_{HEC}}(pH - K_{2_{HEC}}))} \quad (4.13)$$

$$I_{pH_{HEC,Fc}} = \exp\left(\left(-K_{HEC,Fc,1} \frac{(pH - pH_{UL})}{(pH_{UL} - pH_{LL})}\right)^2\right) \quad (4.14)$$

where $K_{1_{HEC}}$, $K_{2_{HEC}}$, $K_{HEC,Fc,1}$, pH_{UL} and pH_{LL} are all constants.

4.3. Model parameterisation and batch bioreactor simulation

4.3.1. The two-species system

Figure 4.1 shows a schematic representation of the interactions between *C. cellulovorans* and *R. palustris* in the two-species system. Cellulose is taken as the only feed substrate and is fermented to H_2 and four different VFAs (volatile fatty acids) by the dark fermentative bacterium *C. cellulovorans*. At the same time, *R. palustris* uptakes acetate, butyrate and lactate as carbon source for growth and additive H_2 production. In a batch reactor, cellulose consumption is modelled by equation (4.15); the growth of *C. cellulovorans* is modelled by equation (4.16):

$$\frac{dS_c}{dt} = -r_{Cc,c}X_{Cc} \tag{4.15}$$

$$\frac{dX_{Cc}}{dt} = \mu_{Cc}X_{Cc} \tag{4.16}$$

X denotes the amount of cells throughout the work. The growth of *R. palustris* and the change in VFAs (considering both production and consumption) are modelled by equation (4.17) and equations (4.18-4.21), respectively:

$$\frac{dX_{Rp}}{dt} = \mu_{Rp}X_{Rp} \tag{4.17}$$

$$\frac{dS_{Ac}}{dt} = p_{VFA_{Ac}}X_{Cc} - r_{Rp,Ac}X_{Rp} \tag{4.18}$$

$$\frac{dS_{Bu}}{dt} = p_{VFA_{Bu}}X_{Cc} - r_{Rp,Bu}X_{Rp} \tag{4.19}$$

$$\frac{dS_{La}}{dt} = p_{VFA_{La}}X_{Cc} - r_{Rp,La}X_{Rp} \tag{4.20}$$

$$\frac{dS_{Fc}}{dt} = p_{VFA_{Fc}} X_{Cc} \tag{4.21}$$

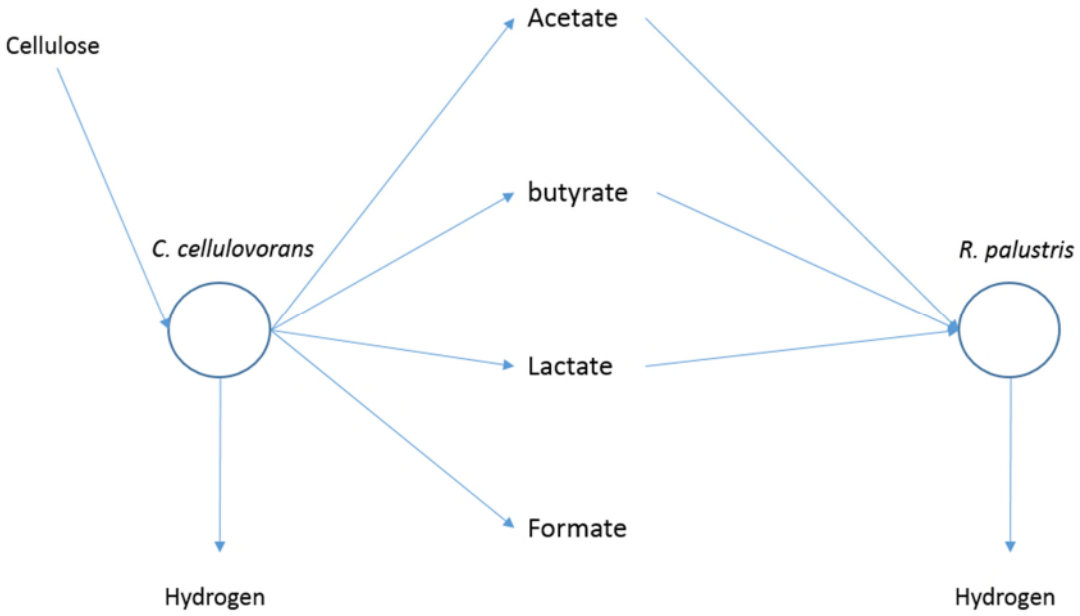


Figure 4. 1. Interactions between *C. cellulovorans* and *R. palustris*.

Hydrogen production is expressed in equation (4.22).

$$\frac{dH_2}{dt} = r_{H_2 Cc} X_{Cc} + r_{H_2 Rp} X_{Rp} \tag{4.22}$$

The above model was fitted to the experimental data reported in (Lu and Lee, 2015); the estimated parameter values are shown in Table 4.1.

As shown in Figure 4.2, the model prediction agrees with the experimental data reasonably well. Compared to *C. cellulovorans* monoculture, the model predicts an increase of the cellulose consumption and the hydrogen production in the co-culture. *R. palustris* uptakes VFAs and hence moderates the drop of pH, which is able to ease pH

inhibition to *C. cellulovorans* resulting in more cellulose converted to hydrogen. In doing so, *R. palustris* also produces hydrogen, further increasing its yield.

Parameter	Value	Unit	Parameter	Value	Unit
$r_{Cc,c,max}$	3.05×10^{-3}	hr^{-1}	$K_{Ac,Rp}$	1.85×10^{-3}	mol/L
$Y_{c/Cc}$	8.74×10^{-2}	g/(10^8 cells)	$K_{Bu,Rp}$	1.90×10^{-1}	mol/L
$Y_{VFA,Ac}$	2.7×10^{-3}	mol/(10^8 cells)	$K_{La,Rp}$	1.11×10^{-4}	mol/L
$Y_{VFA,Bu}$	3.3×10^{-3}	mol/(10^8 cells)	$K_{1_{Rp}}$	8.38	
$Y_{VFA,Fc}$	3.8×10^{-3}	mol/(10^8 cells)	$K_{2_{Rp}}$	6.99	
$Y_{VFA,La}$	1.2×10^{-3}	mol/(10^8 cells)	$Y_{Rp/La}$	1.47×10^{4}	(10^8 cells)/mol
K_c	0.872	g/L	$Y_{Rp/Ac}$	4.74×10^{3}	(10^8 cells)/mol
$K_{1_{Cc}}$	3.78		$Y_{Rp/Bu}$	0	(10^8 cells)/mol
$K_{2_{Cc}}$	5.92		$Y_{H_2,Rp/La}$	0	ml/(mol/L)
$r_{Ac,Rp,max}$	1.022×10^{-5}	hr^{-1}	$Y_{H_2,Rp/Ac}$	8.4×10^{3}	ml/(mol/L)
$r_{Bu,Rp,max}$	3.681×10^{-6}	hr^{-1}	$Y_{H_2,Rp/Bu}$	1.44×10^{3}	ml/(mol/L)
$r_{La,Rp,max}$	2.760×10^{-5}	hr^{-1}			

Table 4. 1. Model parameters for *C. cellulovorans* and *R. palustris*.

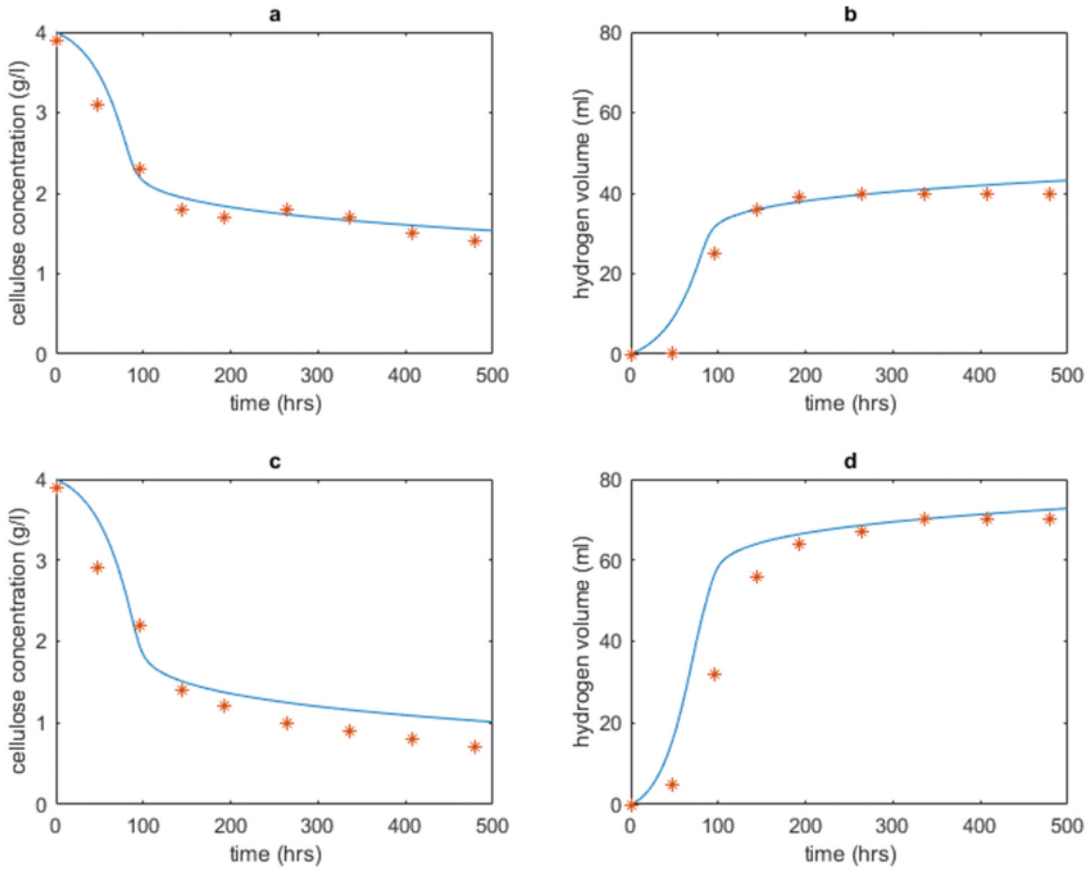


Figure 4. 2. Comparison of experimental data (Lu and Lee, 2015) and numerical simulation results. Solid line and "*" represent simulation results and experimental data, respectively. a, b: mono-culture; c, d: co-culture.

4.3.2. The three-species system

As shown in Figure 4.3, the three-species system further includes HEC's conversion of formate and uptake of acetate as carbon source for growth; the latter implies that HEC emerges as a competitor with *R. palustris*. The growth of microorganisms and the change of all products and substrates in the three-species system are modelled by:

$$\frac{dS_c}{dt} = -r_{Cc,c}X_{Cc} \quad (4.15)$$

$$\frac{dX_{Cc}}{dt} = \mu_{Cc}X_{Cc} \quad (4.16)$$

$$\frac{dX_{Rp}}{dt} = \mu_{Rp}X_{Rp} \quad (4.17)$$

$$\frac{dX_{HEC}}{dt} = \mu_{HEC}X_{HEC} \quad (4.23)$$

$$\frac{dS_{Ac}}{dt} = p_{VFA_{Ac}}X_{Cc} - r_{Rp,Ac}X_{Rp} - r_{HEC,Ac}X_{HEC} \quad (4.24)$$

$$\frac{dS_{Bu}}{dt} = p_{VFA_{Bu}}X_{Cc} - r_{Rp,Bu}X_{Rp} \quad (4.19)$$

$$\frac{dS_{La}}{dt} = p_{VFA_{La}}X_{Cc} - r_{Rp,La}X_{Rp} \quad (4.20)$$

$$\frac{dS_{Fc}}{dt} = p_{VFA_{Fc}}X_{Cc} - r_{HEC,Fc}X_{HEC} \quad (4.25)$$

$$\frac{dH_2}{dt} = r_{H_2Cc}X_{Cc} + r_{H_2Rp}X_{Rp} + r_{H_2HEC}X_{HEC} \quad (4.26)$$

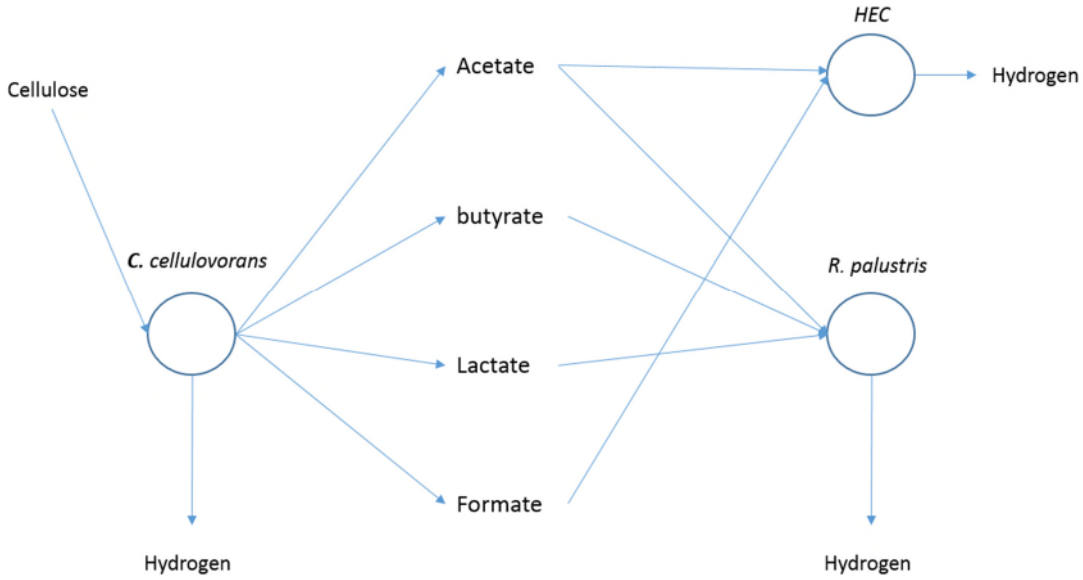


Figure 4. 3. Interactions between *C. cellulovorans*, *R. palustris* and HEC.

Parameters fitted to literature data (Trchounian et al., 2015; Yoshida et al., 2005) for HEC are shown in Table 4.2 In the lack of relevant knowledge, it is assumed that the introduction of HEC does not change the growth parameters of *C. cellulovorans* and *R. palustris* (as shown in Table 4.1).

Parameter	Value	Unit
$r_{Ac,HEC,max}$	2.26×10^{-4}	ml/10^8 cells hr^{-1}
$r_{Fc,HEC,max}$	7.08×10^{-6}	ml/10^8 cells hr^{-1}
$K_{Fc,HEC}$	0.07	mol/l
$K_{Ac,HEC}$	0.022	mol/l
$Y_{HEC/Ac}$	8000	10^8 cells/mol
$Y_{H_2,HEC/Ac}$	4000	ml/(mol/l)
$Y_{H_2,HEC/Fc}$	1272	ml/(mol/l)
$K_{1_{HEC}}$	2.58	-
$K_{2_{HEC}}$	5.62	-
$K_{HEC,Fc,1}$	7.4	-
pH_{UL}	6.38	-
pH_{LL}	3.01	-

Table 4. 2. Model parameters for the third species (HEC).

With the parameterised dynamic model, the batch reactor with three species has been simulated; the results are shown in Figure 4.4 Compared with the two-species system, the simulation shows that the introduction of HEC would increase hydrogen production

significantly and achieve complete conversion of cellulose (when the initial cellulose concentration is 3g/l). At the start of the process, *R. palustris* grows faster than HEC because it uptakes multiple VFAs and has a higher acetate uptake rate than HEC. After about 70 hrs, *R. palustris* is inhibited due to its pH tolerance lower than that of HEC, whereas HEC starts to grow faster, uptaking acetate and converting formate, and leading to the increase of pH. After pH has increased beyond a certain level, *R. palustris* is "rescued" and starts to grow again. After all available carbon sources are consumed, it stops growing, whereas HEC stops its growth after acetate is depleted.

The above simulation results predict clear advantages of the three species system within a batch reactor. In the next section, the two- and three-species systems operating continuously as a continuous stirred tank bioreactor (CSTR) will be studied, with respect to both productivity and stability.

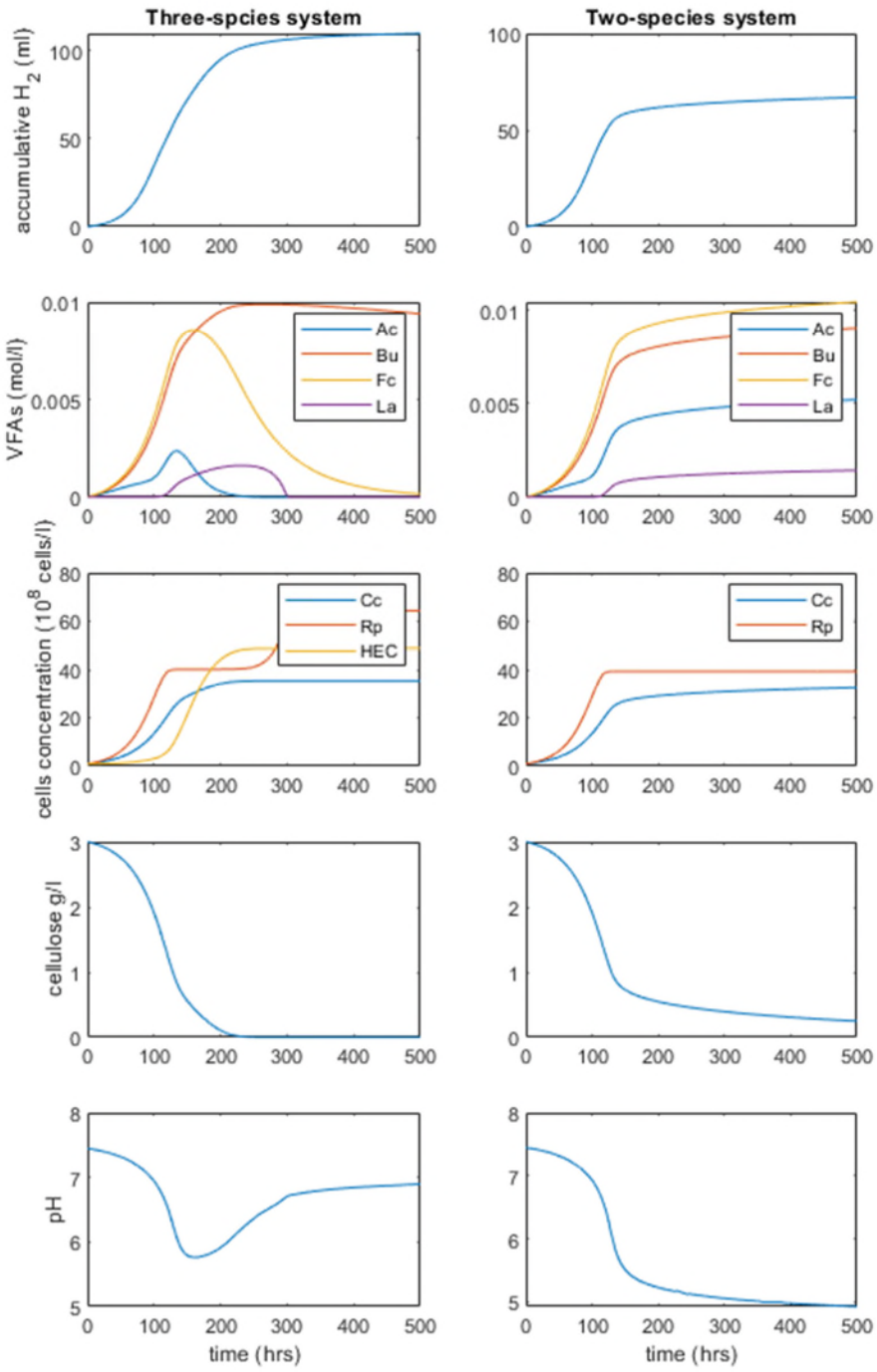


Figure 4. 4. Comparison of hydrogen production, VFAs concentration and cells concentration for two-species and three-species system at initial cellulose concentration 3g/l. Ac, Bu, Fc and La represent acetate, butyrate, formate and lactate, respectively. Cc and Rp represent *C. cellulovorans* and *R. palustris*, respectively.

4.4. Simulation and analysis of CSTR

A CSTR features continuous supply of the substrate and continuous discharge of the content in the reactor, and can reach one or multiple steady states. Compared to a batch reactor, the behaviour arising from the inter-species interactions can be rather different and needs further study in order to understand how the reactor can be better designed and operated. In this part of study, bifurcation analysis and dynamic simulation of both two and three species systems were conducted, to reveal their productivity and stability as one of the key operating parameters, namely the inlet cellulose concentration (ICC), varies while another parameter, namely the hydraulic retention time (HRT) is fixed to a certain level. Several variations to the original models were also simulated to explore a wider range of behaviour. Note that the ranges of ICC shown in the bifurcation diagrams in this section are different from case to case and are always chosen as needed for revealing all the important features of the respective cases, such as the existence of multiple steady states, Hopf points and maximum hydrogen production rates. That is, if a relatively narrow range of ICC is displayed, it means that beyond the displayed range, no additional features exist.

4.4.1. Simulation and analysis of the two-species system

The bifurcation diagrams for two-species system with respect to a range of ICC (HRT = 60 hours) are shown in Figure 4.5 The diagrams show two stable branches connected by an unstable branch in the middle. In the region where all three steady states exist, the unstable steady state separates the system to a high conversion state with two species coexisting and a low conversion state where only *C. cellulovorans* remains. One can see that the highest productivity occurs at an ICC of 2g/l, which is outside the bi-stable region, therefore the complication of bi-stability might not affect the (optimal) operation of the reactor. However, if for any reason the system does operate in the bi-stable region, certain disturbance could cause the system to jump from the high conversion state to the low conversion state.

Figure 4.6 is provided here to show an illustrative example where the system can dynamically evolve to the two different stable steady states, when different initial community compositions are adopted.

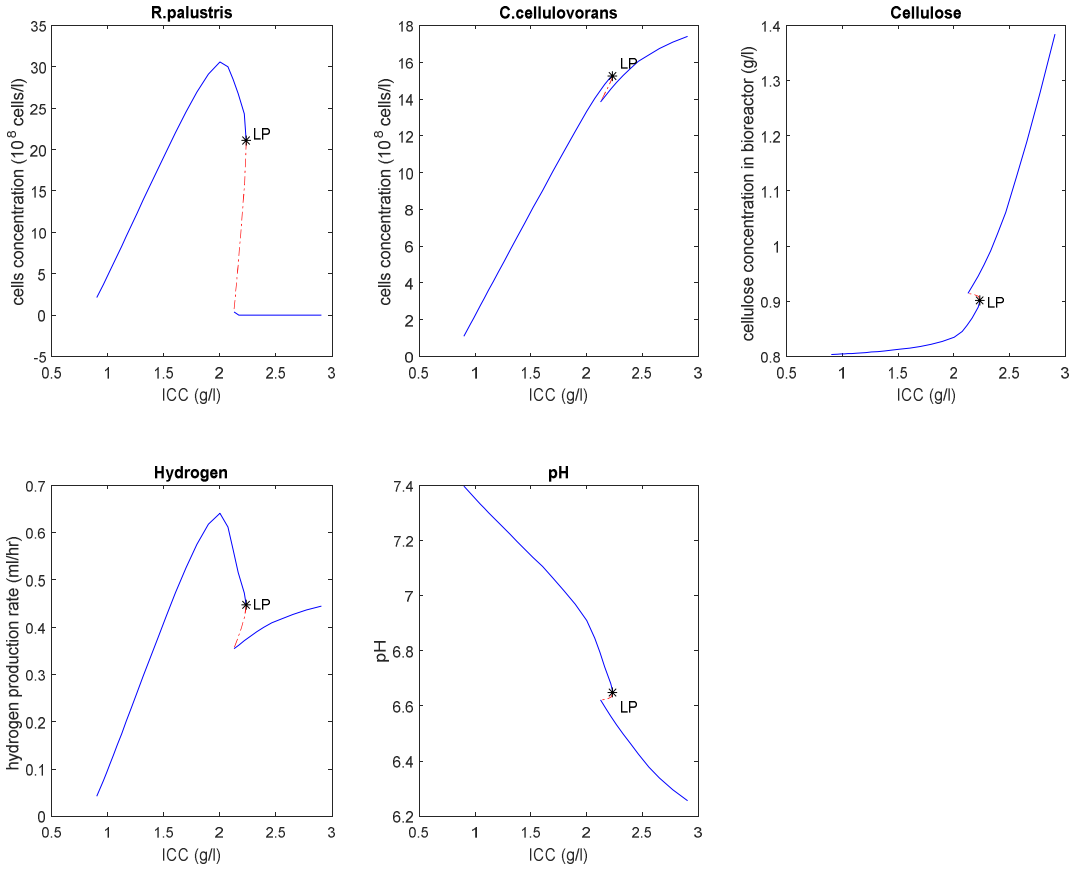


Figure 4. 5. Bifurcation diagrams for the two-species CSTR system. Solid line: stable branch; dash line: unstable branch; '*': static limit point (marked as "LP").

Impact of relative pH tolerance in the two-species system. Our simulation analysis further showed that the behaviour of the system was rather sensitive to the relative pH tolerance of the two species. Here, we present the analysis of two variations to the original model of *R. palustris* (Figure 4.7). In variation 1, pH tolerance of *R. palustris* is increased, compared to the original model. With this change, *R. palustris* and *C. cellulovorans* are able to co-exist within a wider range of ICC (Figure 4.8). There is no bifurcation in the operating region simulated, hence complications due to bi-stability do not exist in such a system at all.

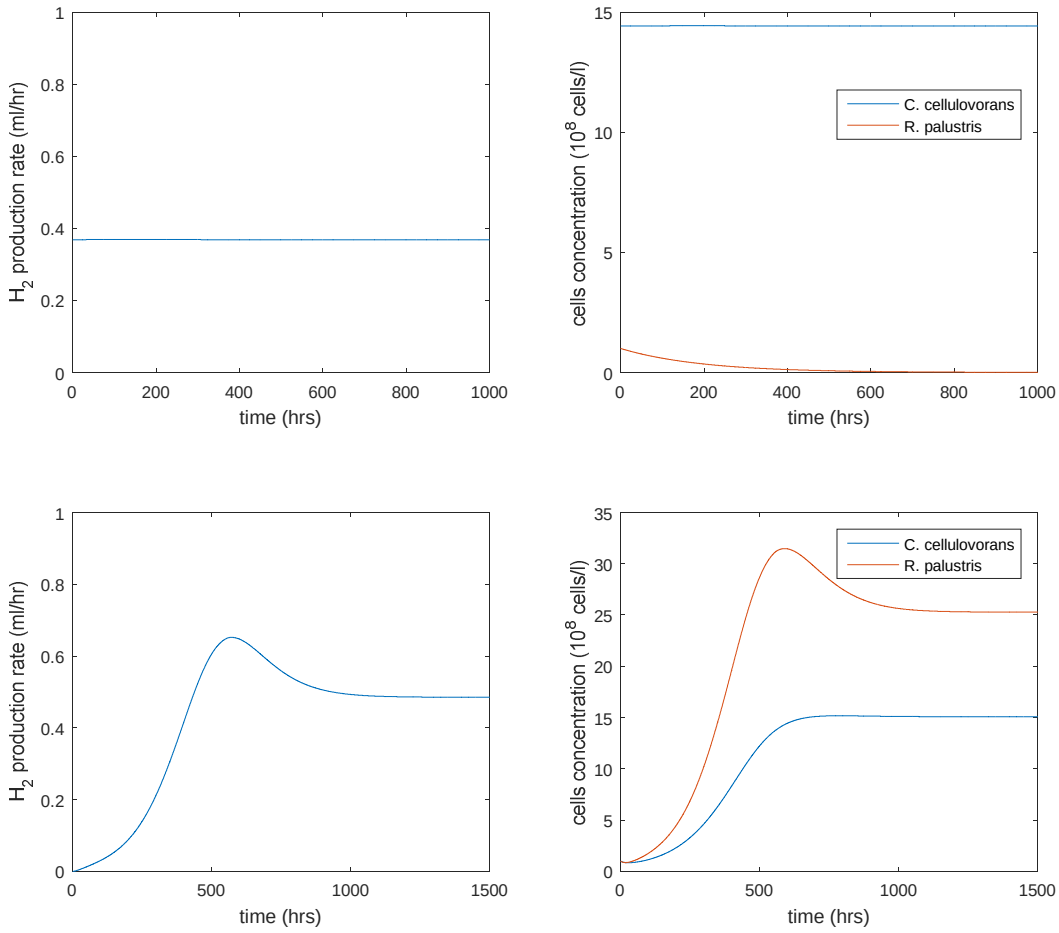


Figure 4. 6. Evolvement of the two-species CSTR system to two different stable steady states. Cc and Rp represent *C. cellulovorans* and *R. palustris*. The upper two plots show the transition to a *C. cellulovorans* only state; the lower two plots show the transition to a co-existence state. HRT=60 hours; ICC = 2.2 g/l.

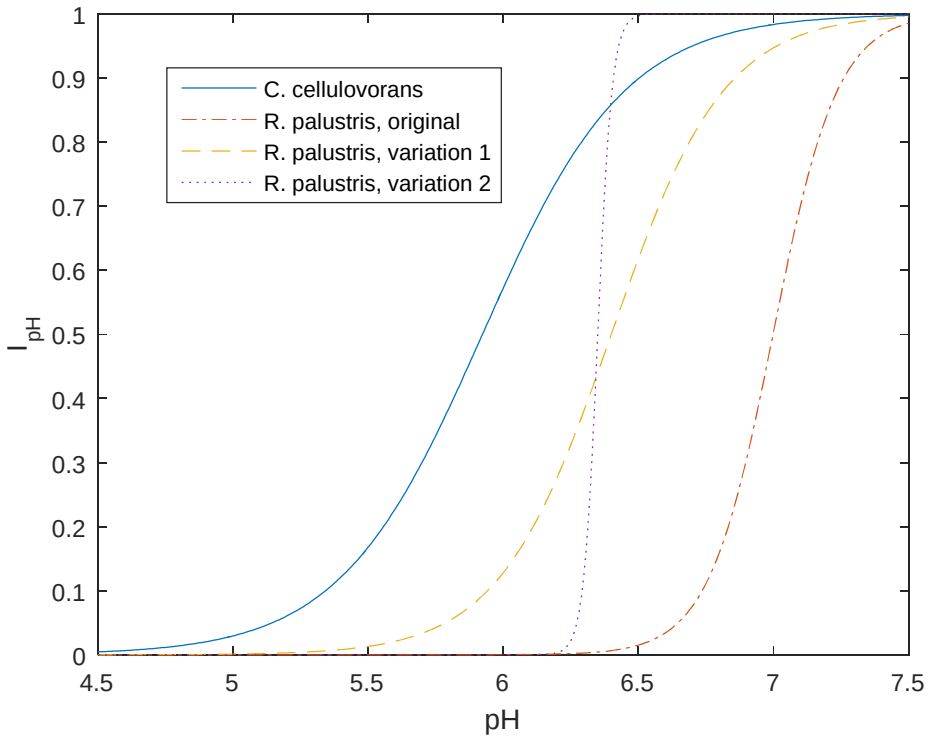


Figure 4. 7. pH inhibition curves of *R. palustris* and *C. cellulovorans* in the original model, variation 1 and variation 2.

In variation 2, a sharper pH effect of *R. palustris* is adopted. While the system behaves similarly to that depicted by the original model when HRT remains as 60 hrs (not presented), it shows a Hopf bifurcation point when HRT is changed, for example, to 39 hrs (Figure 4.9); oscillations are expected to occur beyond the Hopf point, as illustrated in Figure 4.10. The oscillatory behaviour arises from the rapid switch of *R. palustris* from growth to being fully inhibited as pH drops to a threshold level. This subsequently leads to the further dropping of pH which eventually inhibits *C. cellulovorans*. With the decrease of the growth rate of *C. cellulovorans*, VFAs production reduces to be lower than the dilution rate, and pH increases. This thus allows *R. palustris* and *C. cellulovorans* to recover, until the next cycle begins.

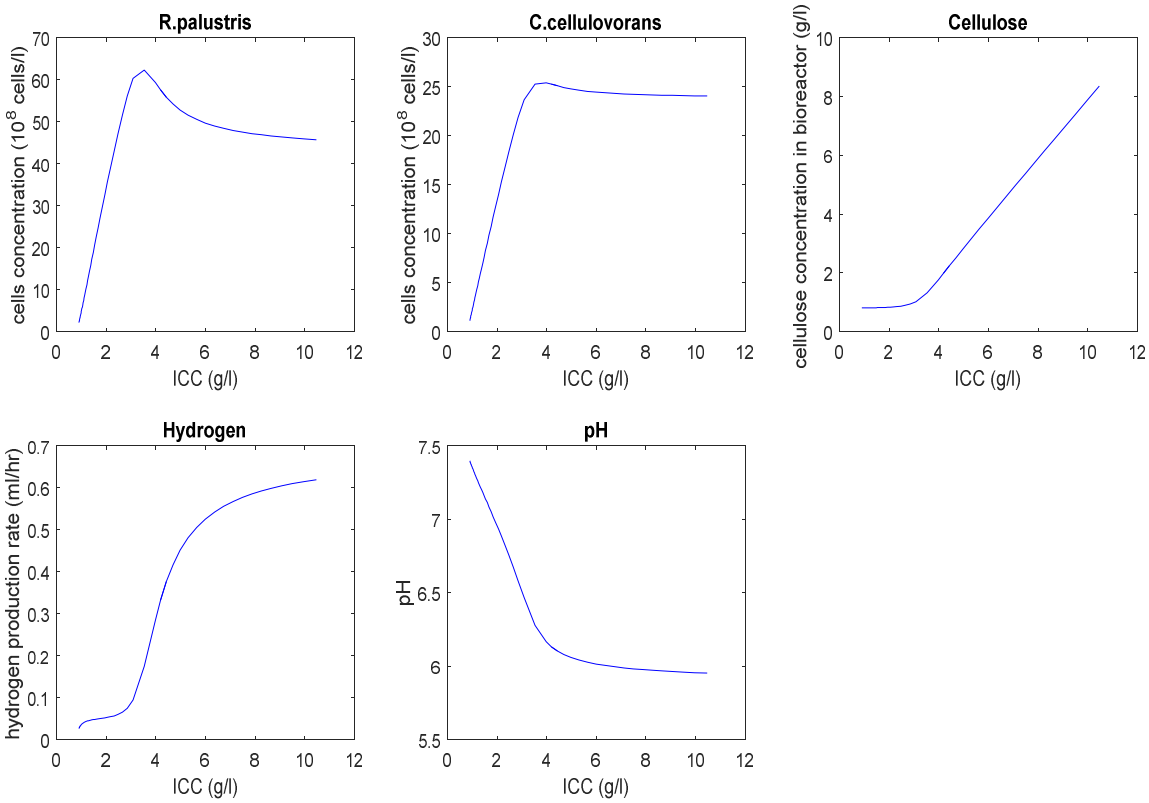


Figure 4. 8. Continuity diagram for the two-species CSTR system in variation 1; HRT = 60hrs.

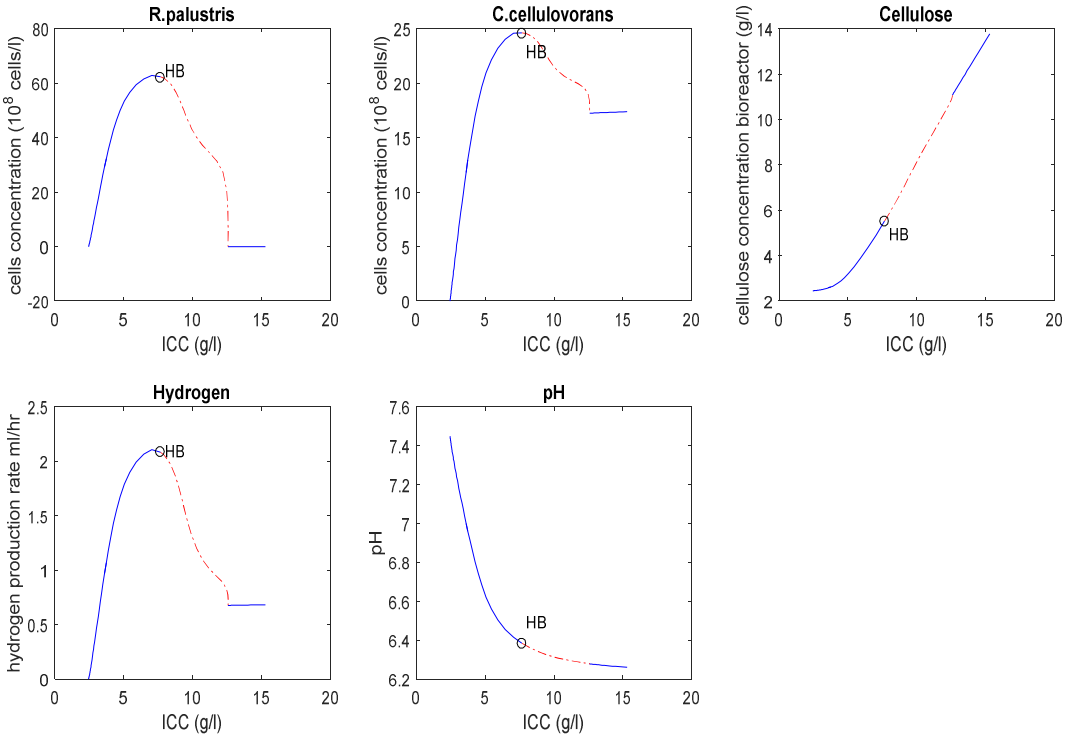


Figure 4. 9. Continuity diagram for the two-species CSTR system in variation 2. Solid line: stable branch; dash line: unstable branch; '*': static limit point; 'o': Hopf bifurcation (marked as "HB").

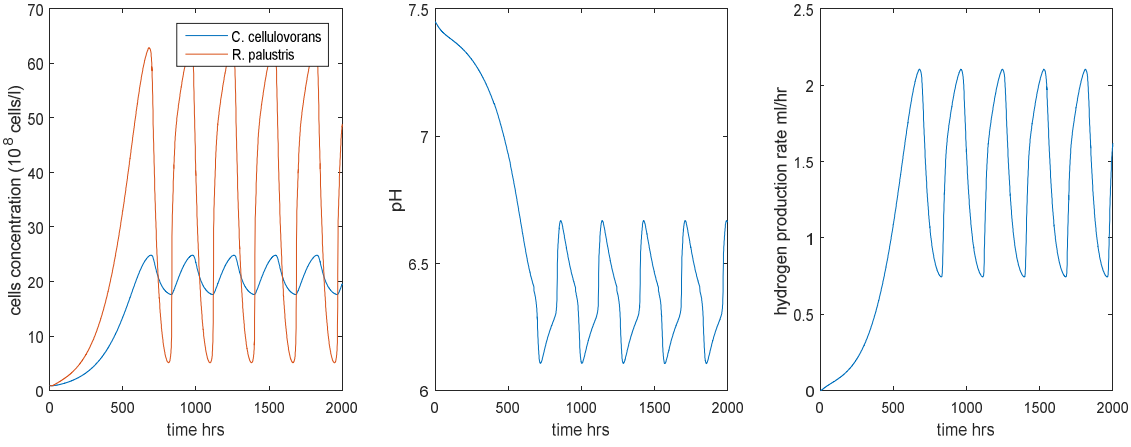


Figure 4. 10. Oscillatory behaviour for the two-species CSTR system, variation 2. HRT = 39 hrs; ICC = 9 g/l.

4.4.2. Simulation and analysis of the three-species system

Figure 4.11 shows the performance of the three-species system constructed by introducing hypothetic HEC to the two-species system. Compared to the system with *C. cellulovorans* and *R. palustris* (Figure 4.5), the stability and hydrogen productivity are generally both improved. On stability, although bi-stability still exists, the switch to a stable steady state with lower hydrogen production yield will not cause too much loss. Furthermore, unlike the two species system which exhibits a peak hydrogen production rate, the hydrogen output largely continuously grows in the three-species system (apart from the narrow bi-stable region) with the increase of ICC. In terms of the structure of the community, when the ICC is low (less than around 1.9 g/l), HEC is not able to co-exist with *R. palustris* because HEC has a lower rate of uptaking acetate, losing in the competition with *R. palustris*. However, HEC has higher pH tolerance compared to *R. palustris*. This results in the ability of co-existence with *R. palustris* through uptaking VFAs, which benefits the latter by allowing it to survive at higher concentrations of the ICC. When the ICC further increases, *R. palustris* is not able to survive because of the more severe dropping of pH, thus *C. cellulovorans* co-exists only with HEC. These inter-species mechanisms, more sophisticated than those in the two-species system, underpin the predicted overall improvement in hydrogen production.

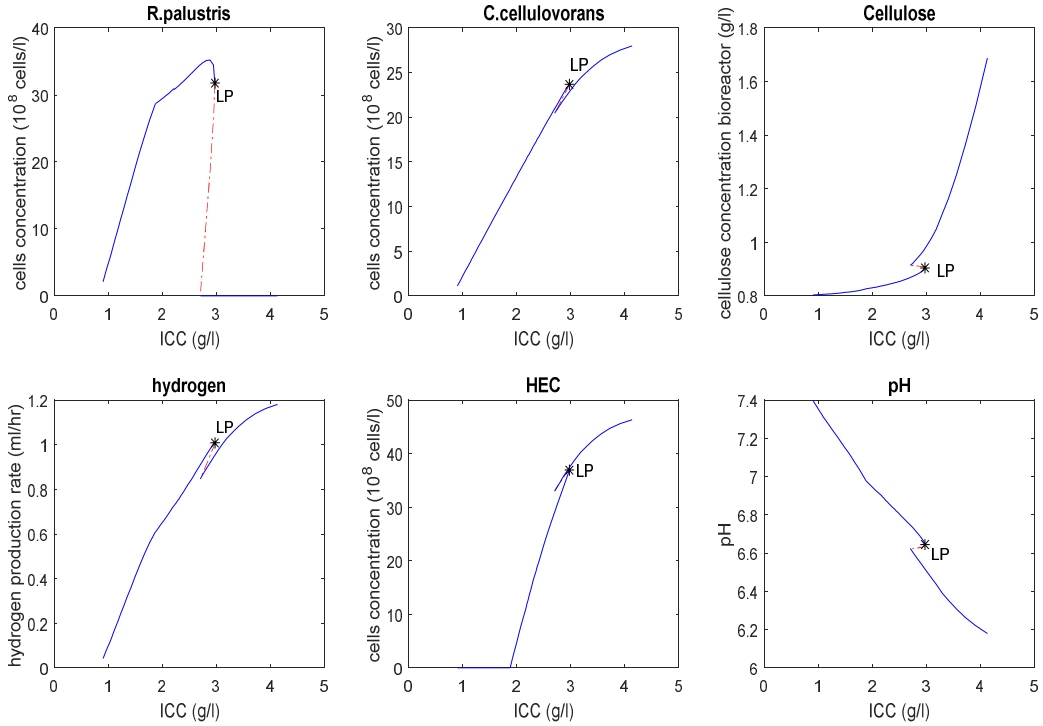


Figure 4. 11. Continuity diagram for the three-species CSTR system in the original model. Solid line: stable branch; dash line: unstable branch; '*': static limit point (marked as "LP"). HRT = 60 hrs.

Impact of improved complementarity in VFAs utilisation. In the original model of the three-species system, the competition between HEC and *R. palustris* on acetate is an important relationship. Here we introduce a change such that HEC uptakes butyrate as carbon source instead of acetate (but with same uptake rate). In the original model, *R. palustris* has a much lower uptake rate in butyrate than that in acetate, thus such a change will not only eliminate the competition on acetate between HEC and *R. palustris*, but could also lead to improved complementary utilisation of VFAs. The possible effects of this change include ease of the dropping in pH, which benefits all species, and increase in the overall conversion of VFAs to hydrogen.

As shown by the results presented in Figure 4.12, the hydrogen production rate can indeed increase significantly, by more than 3 folds as compared to Figure 4.11. On the other hand, bi-stability occurs in a region where the highest productivity is attained, therefore the operation of such a system needs to avoid migration towards the undesirable state under disturbances.

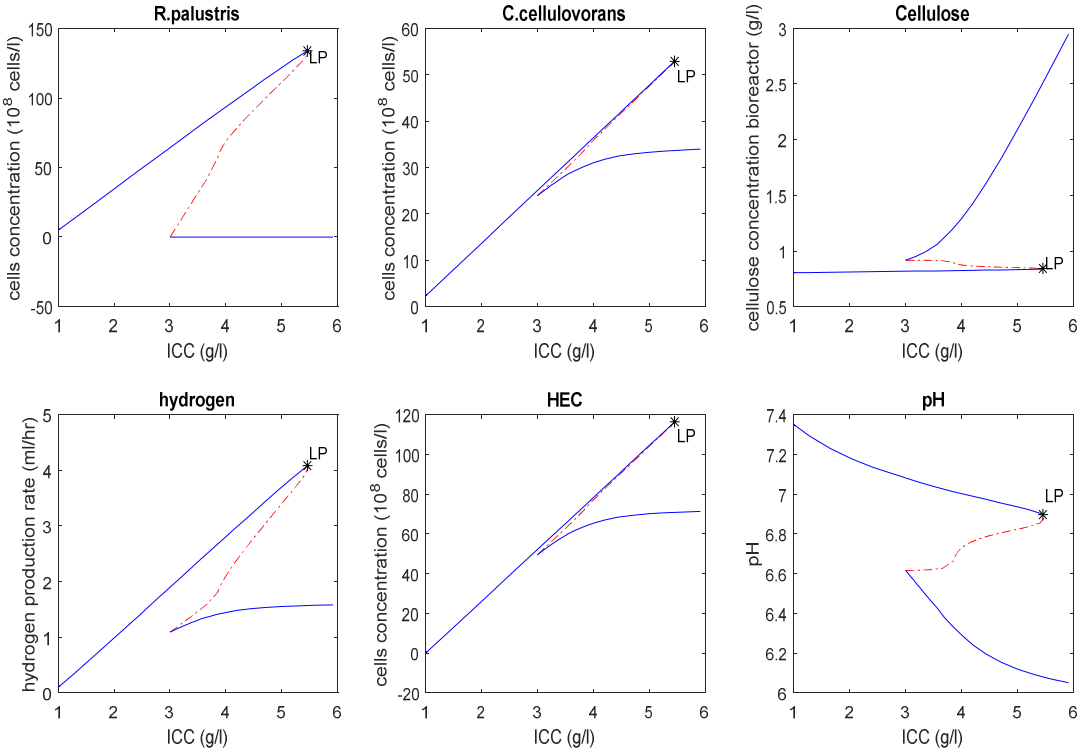


Figure 4. 12. Continuity diagram for the three-species CSTR system with improved complementarity in VFAs utilization. Solid line: stable branch; dash line: unstable branch; '*': static limit point (marked as "LP"), HRT = 60 hrs.

Impact of alteration in pH inhibition effect. In the two-species system, a sharper pH tolerance curve of *R. palustris* was shown to lead to an oscillatory behaviour. We now examine the impact of the same variation on the three-species system. In Figure 4.13,

HEC and *R. palustris* are shown to be able to co-exist only in an oscillatory process. Like with variation 2 of the two-species system, pH drop triggers oscillation. However, the inter-species mechanism has changed in the three species system; HEC cannot survive and co-exist with *R. palustris* when the ICC is low. When the feed concentration increases, *C. cellulovorans* produces more VFAs and once reaching a threshold, *R. palustris* is inhibited and oscillation starts. In the oscillation, HEC competes with *R. palustris* on acetate, as shown in Figure 4.14. Accompanying the drop of pH, the populations of *R. palustris* and *C. cellulovorans* decrease, while that of HEC increases with the improved availability of acetate. The pH "conditioning" function of HEC gradually allows *R. palustris* (and *C. cellulovorans*) to recover, eventually leading to the decline of HEC due to the intensified competition with *R. palustris* on acetate, which continues until the next cycle starts.

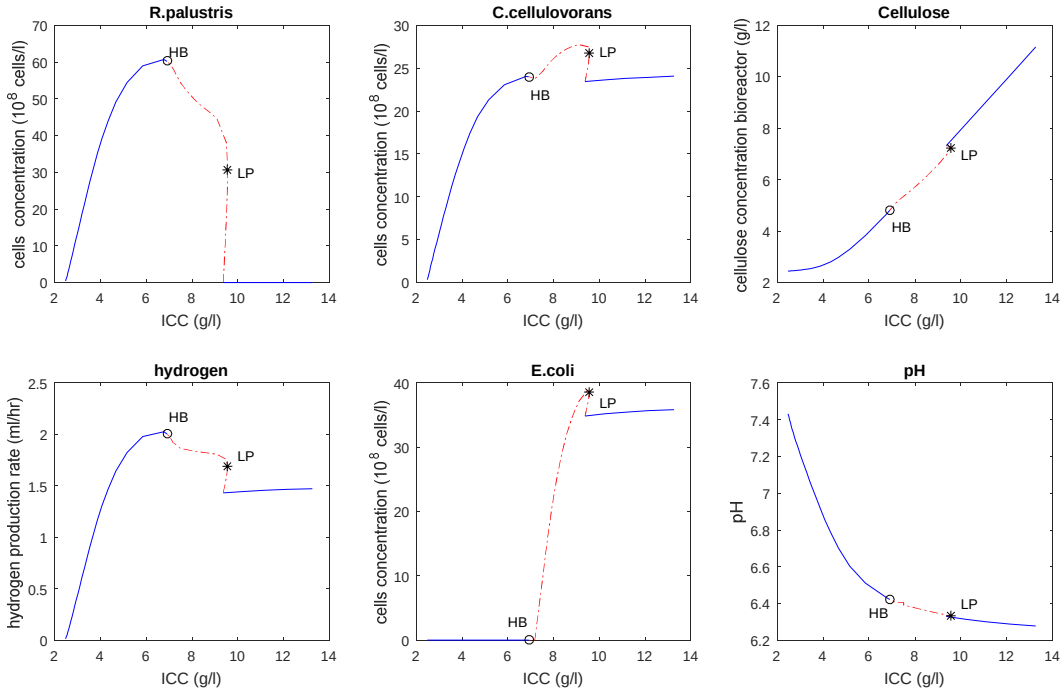


Figure 4. 13. Continuity diagram for the three-species CSTR system with altered pH inhibition effect. Solid line: stable branch; dash line: unstable branch; '*': static limit point (marked as "LP"); 'o': Hopf bifurcation (marked as "HB"). HRT = 39 hrs.

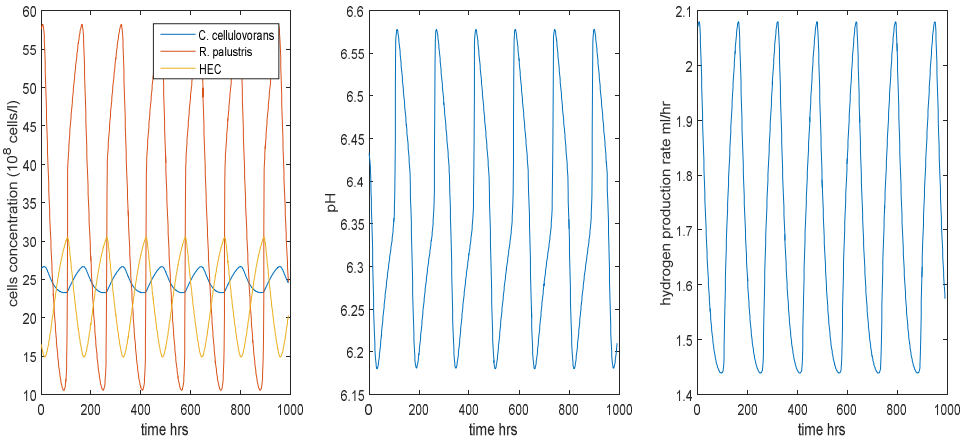


Figure 4. 14. Oscillatory behaviour of the three-species CSTR system, with altered pH inhibition effect. HRT = 39 hrs; ICC = 9 g/l.

4.5. Discussion

Synthetic microbial assemblages consisting of two species have previously been investigated in several studies, typically for engineering purposes involving sequential conversion which benefits from the mutualistic relationship between the pair. In particular, the second species in a few cases is introduced to convert acidic intermediate metabolite(s) from the first species, to increase the final product and at the same time to help maintain pH to a level beneficial to the first micro-organism (e.g. Rai and Singh, 2016; Zhou et al., 2015). In the present work, the two-species system shares this characteristic, and, together with the other literature examples, suggests that such a design has wide engineering applicability.

Compared to two-species systems, designing a bioreactor involving a synthetic community of three or more species is still rare. In (Liu et al., 2017), a two-species microbial consortia for producing electricity in a bioelectrochemical system was enhanced by introducing a third species according to the principle of 'division-of-labour'. In the present work, a two-species system is augmented with a third species primarily to address specific limitations of the former, namely pH regulation and feedstock utilisation due to the lack of conversion of formic acid. While this additional species proves *in silico* the potential to overcome these limitations (hence being commensal to the existing two species), it also introduces competition with one of the existing species (*R. palustris*), leading to rather complex behaviours. Furthermore, the dynamics and productivity of the system, particularly in a continuous reactor,

significantly depend on the biological traits such as pH tolerance, as well as the ranges of operating parameters such as feed concentration and HRT. These factors should thus be carefully considered in the future applications of synthetic microbial communities.

In terms of modelling the growth kinetics of microorganisms, different outcomes may be predicted depending on the choice of kinetic models. In literature, Andrew's inhibitory kinetics model, which assumes exceeding supply of substrate inhibits the growth of microorganisms, is used in some cases to replace the Monod type equation adopted in this work. Based on Andrew's kinetics, an increase of complexity in dynamics was observed (Ajbar and Alhumanzi, 2000). Other relevant work also indicated, in terms of uptaking multiple substrates (as is the case of *R. palustris* in this work), a cybernetic model can be used to describe the sequential uptake orders of each of substrates (Vasudeva et al, 2009). However, based on the information available from the data source (Lu and Lee, 2015), we have considered only low pH as the inhibitory factor; and according to the experimental data (Lu and Lee, 2015), we have assumed that *R. palustris* uptakes multiple acids simultaneously. Moreover, the process of *C. cellulovorans* uptaking cellulose can be divided into two steps: decomposition of cellulose to glucose and uptake of glucose. Research on this mechanism suggested that the rate determining step could be the decomposition process (Schwarz, 2001; Shoham et al., 1999). In our work, a single rate equation has been adopted without explicitly incorporating such detailed mechanisms to avoid the need for estimating additional parameters. Overall, the simulation results of the relatively simple model agreed with

the experimental data of the two-species system reasonably well. However, for a wider range of the operating conditions beyond what is covered by the experimental data, and more obviously for the case involving the third species, the model is still required to be experimentally validated in the future work.

4.6. Conclusion

In this work, mathematical models of two- and three-species systems for hydrogen production from cellulose were developed according to literature data. Productivity of the bioreactor operating in the batch and the continuous mode reactor was predicted, and the stability of the continuous reactor (CSTR) was analysed through dynamic simulation and bifurcation analysis. In all cases, the three-species system was predicted to offer an improved productivity compared to the two-species counterpart, as expected. In the two-species CSTR, co-existence was found to become infeasible when the inlet cellulose concentration was high, which could be eased by the introduction of the third species. On the other hand, the third species brought both commensalism and competition, hence remarkably further complicated the inter-species interactions. Overall, this case study on bio-hydrogen production shows that when artificially selecting species to design or modify a microbial community system for specific engineering purposes, attention needs to be given to biological traits such as pH tolerance as well as the location of the operating window of key parameters such as feed concentration and hydraulic retention time, to attain desirable combination of production and stability and to avoid unwanted consequences.

References

Ajbar A, Alhumazi K., 2000. Microbial competition: study of global branching ph enomena. AlChE J. 46, 321-334.

Ajbar A., 2002. Classification of static and dynamic behavior in chemostat for pla smid-bearing, plasmid-free mixed recombinant cultures. Chem. Eng. Commun. 18 9, 1130-1154.

Appels L, Baeyens J, Degrève J, Dewil R., 2008. Principles and potential of the anaerobic digestion of waste-activated sludge. Prog. Energy Combust. Sci. 34, 75 5-781.

Dikshitulu S, Baltzis BC, Lewandowski GA, Pavlou S., 1993. Competition betwe en two microbial populations in a sequencing fed-batch reactor: Theory, experime ntal verification, and implications for waste treatment applications. Biotechnol. Bi oeng. 42, 643-656.

Gunaseelan VN., 1997. Anaerobic digestion of biomass for methane production: a review. Biomass Bioenergy. 13, 83-114.Mata-Alvarez J, Mace S, Llabres P., 200 0. Anaerobic digestion of organic solid wastes. An overview of research achieve ments and perspectives. Bioresour. Technol. 74, 3-16.

Jia X, Liu C, Song H, Ding M, Du J, Ma Q, Yuan Y., 2016. Design, analysis an d application of synthetic microbial consortia. Synth. Syst. Biotechnol. 1, 109-11 7.

Kambam PKR, Henson MA, Sun L., 2008. Design and mathematical modelling o f a synthetic symbiotic ecosystem. IET syst. Biol. 2, 33-38.

Lenas P, Pavlou S., 1994. Periodic, quasi-periodic, and chaotic coexistence of two competing microbial populations in a periodically operated chemostat. Math. Bio sci. 121, 61-110.

Yue Liu, Mingzhu Ding, Wei Ling, Yun Yang, Xiao Zhou, Bing-Zhi Li, Tao Che n, Yong Nie , Miaoxiao Wang, Boxuan Zeng, Xia Li, Hong Liu, Baode Sun, He ming Xu, Jiamei Zhang, Yi Jiao, Yanan Hou, Hui Yang, Sijia Xiao, Qucheng Li n, Xinzi He, Wenjie Liao, Zeqi Jin, Yufei Xie, Bofeng Zhang, Tianyu Li, Xi Lu, Jiabei Li, Fan Zhang, Xiao-Lei Wu, Hao Song and Ying-Jin Yuan, 2017. A three -species microbial consortium for power generation. Energy & Environmental Sci ence, 10, 1600-1609.

Lu H, Lee PK., 2015. Effects of cellulose concentrations on the syntrophic intera ctions between Clostridium cellulovorans 743B and Rhodopseudomonas palustris CGA009 in coculture fermentation for biohydrogen production. Int. J. Hydrogen Energy. 40, 11800-11808.

Pavlou S, Kevrekidis IG, Lyberatos G., 1990. On the coexistence of competing m icrobial species in a chemostat under cycling. Biotechnol. Bioeng. 35, 224-232.

Rai PK, Singh SP., 2016. Integrated dark-and photo-fermentation: Recent advance s and provisions for improvement. Int. J. Hydrogen Energy. 41, 19957-19971.

Rollié S, Mangold M, Sundmacher K., 2012. Designing biological systems: syste ms engineering meets synthetic biology. Chem. Eng. Sci. 69, 1-29.

Sari T, Harmand J., 2016. A model of a syntrophic relationship between two mic robial species in a chemostat including maintenance. Math. Biosci. 275, 1-9.

Schwarz W., 2001. The cellulosome and cellulose degradation by anaerobic bacter ia. Appl. Microbiol. Biotechnol. 56, 634-649.

Shoham Y, Lamed R, Bayer EA., 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microb iol. 7, 275-281.

Thomopoulos NA, Vayenas DV, Pavlou S., 1998. On the coexistence of three mi crobial populations competing for two complementary substrates in configurations of interconnected chemostats. Math. Biosci. 154, 87-102.

Tilman, D., 1980. Resources: a graphical-mechanistic approach to competition and predation. The American Naturalist. 116, 362-393.

Trchounian K, Sargsyan H, Trchounian A., 2015. H2 production by Escherichia c oli batch cultures during utilization of acetate and mixture of glycerol and acetate. Int. J. Hydrogen Energy. 40, 12187-12192.

Vasudeva KM, Zeyer KP, Kienle A, Pushpavanam S., 2009. Conceptual Analysis of the Effect of Kinetics on the Stability and Multiplicity of a Coupled Bioreactor

−Separator System Using a Cybernetic Modeling Approach. Ind. Eng. Chem. Res. 48, 10962-10975.

Wade MJ, Pattinson RW, Parker NG, Dolfing J., 2016. Emergent behaviour in a chlorophenol-mineralising three-tiered microbial 'food web'. J. theor. Biol. 389, 1 71-186.

Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H., 2005. Enhanced hy drogen production from formic acid by formate hydrogen lyase-overexpressing Es cherichia coli strains. Appl. Environ. Microbiol. 71, 6762-6768.

Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F., & Hairston Jr, N. G., 2 003. Rapid evolution drives ecological dynamics in a predator-prey system. Natur e. 424, 303.

Zhou K, Qiao K, Edgar S, Stephanopoulos G., 2015. Distributing a metabolic pat hway among a microbial consortium enhances production of natural products. Na t. Biotechnol. 33, 377-383.

Appendix:

Model-based analysis and optimisation of hydrogen production with a synthetic microbial community

Sihao Di, Aidong Yang

Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK

Naturally occurring communities of microorganisms have long been used for engineering purposes, for example in anaerobic digesters for treating waste organic streams. In recent years, a number of attempts have been made in developing bioprocesses based on artificially formed consortia of multiple microorganisms, referred to as synthetic microbial communities, to gain advantages over the conventional, mono-culture based processes. One of such processes is a two-bacteria system for the purpose of hydrogen production [1]. In this system, *C. cellulovorans* carries out the dark fermentation to convert cellulose into VFAs while producing hydrogen. The second bacterium, *R. palustris* converts some of the VFAs into more hydrogen through a photo fermentation process. At the same time, it also helps to moderate the increase of acidity and hence ease the pH inhibition to *C. cellulovorans* and increase cellulose consumption. However, Limitations exists with both pH regulation and VFA utilisation achievable by *R. palustris,* as revealed by the experimental study [1].

The aim of this work is to explore the possible solutions to these limitations by means of mathematical modelling. Monod type equations incorporating inhibition effects of pH and other possible sources were used to describe the interactions between the two bacteria in a batch reactor. The resulting model was able to capture the main features of the process, such as the contrast in performance between the case with dark fermentation only and that of co-culture, and the impact of different substrate concentrations. It was also able to reveal the difference in the rate of acid production and consumption, as well as the difference between the two bacteria in pH tolerance; these differences appear to significantly affect the performance of the system.

As the next step, possible manipulations were studied through numerical simulation in order to cope with the limitations revealed by the model. Firstly, different cellulose feeding regimes were tested in order to optimize the production of hydrogen during a given period of time, through balancing substrate supply and pH control. Secondly, simulation experiments were conducted to assess the impact of changing the composition of the consortium, to overcome the inherent limitations posed by the original consortium and further enhance hydrogen production.

Through the above application, this work demonstrates how mathematical modelling can be used to guide the construction, analysis and optimisation of synthetic microbial community based systems with respect to both operational and biological settings, as a systematic approach to the development of successful industrial applications of such systems.

Reference

[1] Lu, Hongyuan, and Patrick KH Lee. "Effects of cellulose concentrations on the syntrophic interactions between Clostridium cellulovorans 743B and Rhodopseudomonas palustris CGA009 in coculture fermentation for biohydrogen production." *International Journal of Hydrogen Energy* 40.35 (2015): 11800-11808.

Chapter 5. Analysis of productivity and stability of synthetic microbial communities

Summary

Motivated by the observation of rather sophisticated behaviours from the case study in Chapter 4 on microbial community based bioreactors for bio-hydrogen production, this chapter is tended to offer a more generalised study on a set of microbial community structures involving two or three species, with sequential cross-feeding as the backbone of a sophisticated set of relations including particularly inhibitions which were shown to be important in chapter 4. The aim is to gain more generalised insights on how system structure and parameterization would affect stability and productivity of synthetic microbial communities.

To achieve this goal, we first define a base model with a sequential cross-feeding structure involving two substrates and two species. This base model is then extended to encompass three layers of complexity: the first layer contains cases each involving a single inhibition between two species (via an exchanged metabolite), the second layer with cases each involving a combination of two different inhibitions, and the third layer focuses on two cases where a third species is introduced with an expected function and associated ecological interactions with the two existing species.

The study of base system and the layer-one systems reveals key factors that affect the productivity, the location and the size of the coexistence region, the system's ability

re-stabilise from disturbances, and the onset of oscillatory behaviour. The study of the systems in layer 2 shows that the combination of individual inhibition connections can lead to (1) additive behaviour where the overall effect is the further reduction of productivity and stability, (2) an easing effect by one inhibition on another due to opposing impacts, or (3) complex behaviours such as oscillation and bi-stability particularly due to the formation of feedback loops. When a third species is introduced to the bioreactor for the purpose to improve productivity, the aim could be achieved but accompanied by a reduction of stability.

Overall, this work renders a systematic view on the general design and analysis of synthetic microbial communities with commonly occurring structures and parameterization, in the context of a continuous bioreactor.

This chapter comprises the following article:

Di, S., & Yang, A. (2019). Analysis of productivity and stability of synthetic microbial communities. Journal of the Royal Society Interface, 16(150), 20180859.

List of symbols and abbreviations

D : dilution rate

HRT: hydraulic retention time

$I_{i,j}$: inhibition factor on species *i* by chemical or species *j*

K_i: half saturation constant of species *i*

Re_i: real part of the i^{th} of n eigenvalues

r_i: growth rate of species *i*

S_k: concentration of chemical *k*

$S_{k,\,in}$ inlet concentration of chemical *k*

X_i: concentration of species *i*

$Y_{k,i}$: yield factor, the amount of chemical *k* that is produced or consumed by the growth of a unit amount of the biomass of microbial species *i*

$\mu_{max,i}$: maximum specific growth rate of species *i*

$\phi_{i,j}$: inhibition constant on species *i* by chemical or species j

5.1. Introduction

In recent years, a number of attempts have been made in developing bioprocesses based on artificially formed consortia of multiple microorganisms, referred to as synthetic microbial communities, to gain advantages over the conventional, mono-culture based processes [1-4]. A synthetic microbial community has the potential to complete complex tasks, which are either impossible or inefficient to be carried out by a single wild-type or genetically-engineered species, through making the best use of (i) the properties and capacities of each member of the community and (ii) the interactions between different members [4-5]. On the other hand, such systems inevitably involve a greater level of complexity, originating from the interactions between the multiple species, which requires careful analysis and design in order to achieve the intended engineering objectives [6].

In order to develop a better understanding of the inter-species mechanisms, theoretical analysis based on either analytical models [7] or numerical simulation [8] could play an important role. Existing literature has reported a number of investigations on the interactions of microbial species within natural or artificial communities, particularly focusing on inter-species competition. It is well-known that two microbial species competing for a single rate-limiting substrate are not able to co-exist under steady-state operation [9], however, there are specific operational settings which are able to make the co-existence possible [10-13]. In addition to competitive communities, a two-species symbiotic ecosystem was studied, where quorum sensing, a mechanism

that affects the growth rate of a species according to the cells concentration, was shown to impose a significant impact on the stability of the consortium [14]. Existing literature also investigated ecosystems with more than two species, such as a three-species system competing for two resources [15]. Further, several studies have analysed syntrophic consortia where producers and consumers critically interdepend on each other, particularly in the anaerobic digestion process [16-17].

The above studies have mostly focused on only the stability of microbial communities, but not their productivity. One exception is a modelling study on anaerobic digestion which recognised that a bistable state of the system could also be where an increased yield occurs [17]. In general, the productivity of a microbial community represents an important aspect of its function (e.g. [34]), which however is yet to be more generally understood for engineering applications together with stability. Besides, the systems analysed in the past commonly involve a single type of relationship between any two species in a community, e.g. competition or syntrophy, whereas in an artificially designed engineering system multiple relationships may co-exist [18], which could be further complicated by the presence of interactions between chemical and biological species, for example, through feed or product inhibition. Therefore, an understanding of the impact of such more complex inter-connectedness is needed.

In this work, an *in silico*, generalised analysis is carried out on bioreactors based on synthetic microbial communities. The aim is to gain insights on how the structures

and parameterisation of such systems could affect their productivity and stability, to provide guidance to the engineering of such bioreactors. In particular, a base system with two cross-feeding species is used, on top of which a variety of two-species systems with inhibition(s) are constructed and studied. These systems, demonstrating the principle of division of labour by realising a complex conversion task through a sequential process, represent the most common schemes of synthetic communities based bioreactors that have been reported in the literature (see Table 5.1). Furthermore, two representative three-species systems are investigated. These two- and three-species systems are described by generic mathematical models, and their behaviour and performance studied by bifurcation analysis and numerical simulation.

Ref.	Feed	Product	1st species	2nd species	Inhibition
[19]	cellulose	hydrogen	*C. cellulovorans*, converting the feed to acids	*R. palustris*, converting acids to hydrogen	low pH, on both species, reduced by acids consumption of the 2nd species
[20]	cellulose	isobutanol	*T. reesei*, hydrolyzing the feed to soluble carbohydrates	*E. coli*, converting soluble carbohydrates to product	isobutanol, on both species
[21]	cellulose	ethanol	*C. phytofermentans*, hydrolyzing the feed to soluble carbohydrates	Yeast, converting soluble carbohydrates to product	oxygen and cellodextrin (produced by 1st species), on the 1st species, reduced by consumption of both by the 2nd species
[5]	xylose	oxygenated taxanes	*E. coli*, converting xylose and producing taxadience and acetate	*S. cerevisiae*, consuming acetate and converting taxadience to product	acetate, on *E. coli*, reduced through consumption by the 2nd species

Table 5. 1. Literature examples of engineered cross-feeding microbial consortia

5.2. Methods

5.2.1. Construction of representative systems

To construct representative microbial communities, we consider a set of building blocks involving two parts: microbial species (denoted as X) and chemicals processed or produced by the species (denoted as *S*). With this notation, we first consider a *base system* that forms a simple cross (forward) feeding chain (Figure 5.1), where species X_1 is produced by consuming chemical S_1 which itself is produced by species X_2 through consuming chemical S_2 that acts as the substrate of the overall system. In reality, X_1 would be the species to produce the final, desired chemical product. For the sake of simplicity, we assume such a product would have no feedback effect on the behaviour of the microbial community, therefore it is not explicitly represented in the model. Instead, the abundance of X_1 is assumed to be proportional to the quantity of the product and is taken as a proxy of the latter.

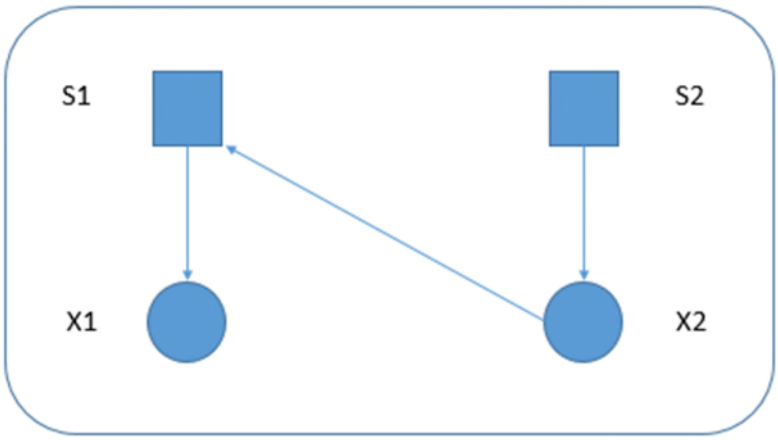


Figure 5. 1. The base system.

On top of this base system, we construct its variations with three layers of complexity. In layer 1 (Figure 5.2), four two-species cases (numbered as cases 1-1 to 1-4) are considered, each of which contains an inhibition between a chemical and a species or between the two species. In reality, an inter-species inhibition may still be mediated by a certain chemical, which is however not explicitly represented in the system structure. Note that, an inhibition between S_1 and X_1 and that between S_2 and X_2 are not considered, as they will not affect the interactions between the two species and are therefore not relevant to this study.

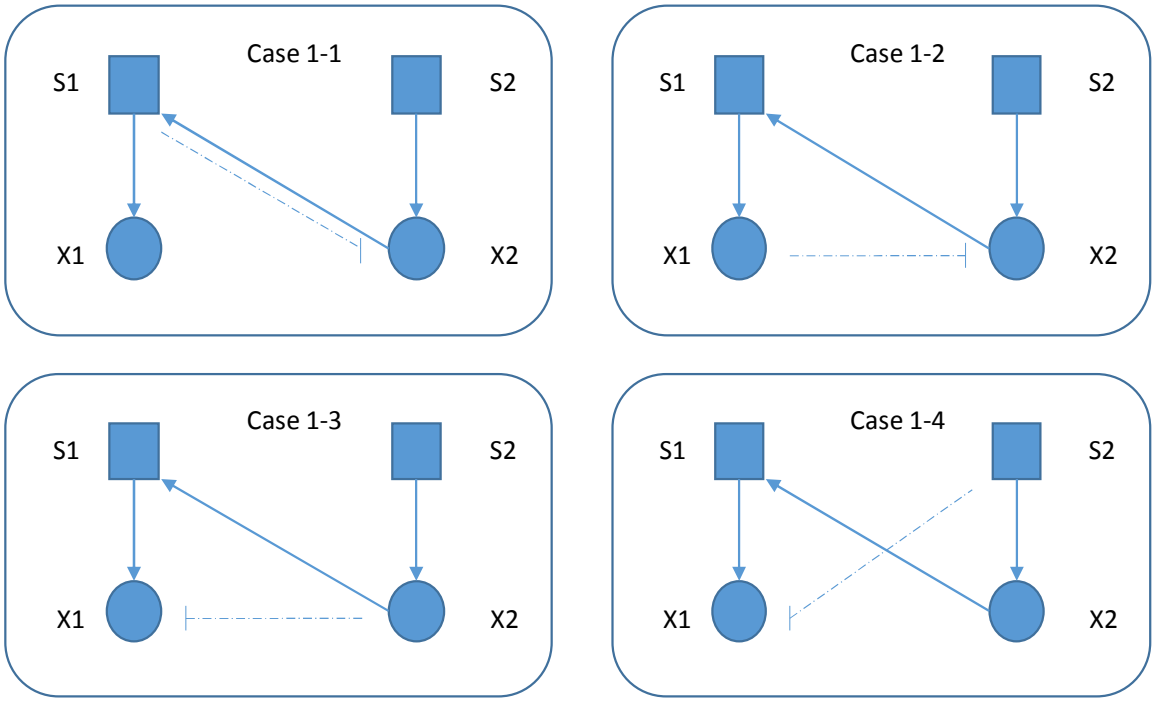


Figure 5. 2. Two-species systems with a single inhibition (Layer-1 cases).

In the next step, single inhibitions from layer 1 are combined, leading to six cases (numbered as cases 2-1 to 2-6), each of which includes two different inhibitions; these

cases are referred as those of layer 2 (Figure 5.3 and Table 2). The layer-2 cases are considered in order to investigate the compound effects of multiple inhibitions that may exist in a two-species system.

Finally, we introduce layer 3 where systems involving three species are considered. Compared to two-species systems, synthetic communities containing 3 (or more) species are still rare, albeit intuitively they have the potential to bring more opportunities for engineering design. On the other hand, the number of their possible structures is much larger due to the combinatorial effect, which would be difficult to investigate exhaustively. In this work, we take a view of incremental design and focus on the circumstances where a third species is introduced to a two-species system for a specific improvement. In particular, two important ecological functions of the third species are considered, namely (i) complementary resource utilisation and (ii) facilitation. These considerations lead to the construction of two representative three-species cases for layer 3 (Figure 5.4).

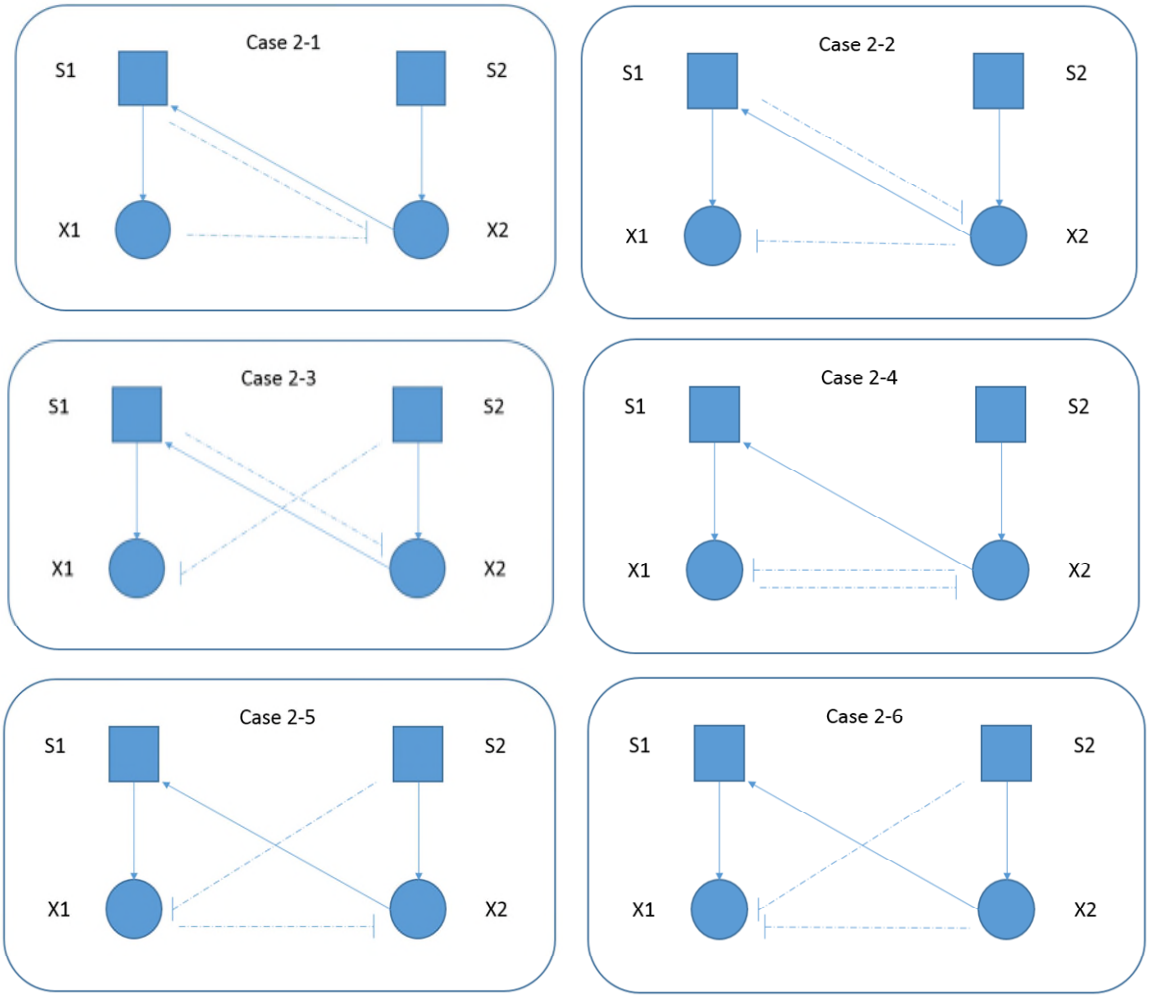


Figure 5. 3. Two-species systems with double inhibitions (Layer-2 cases).

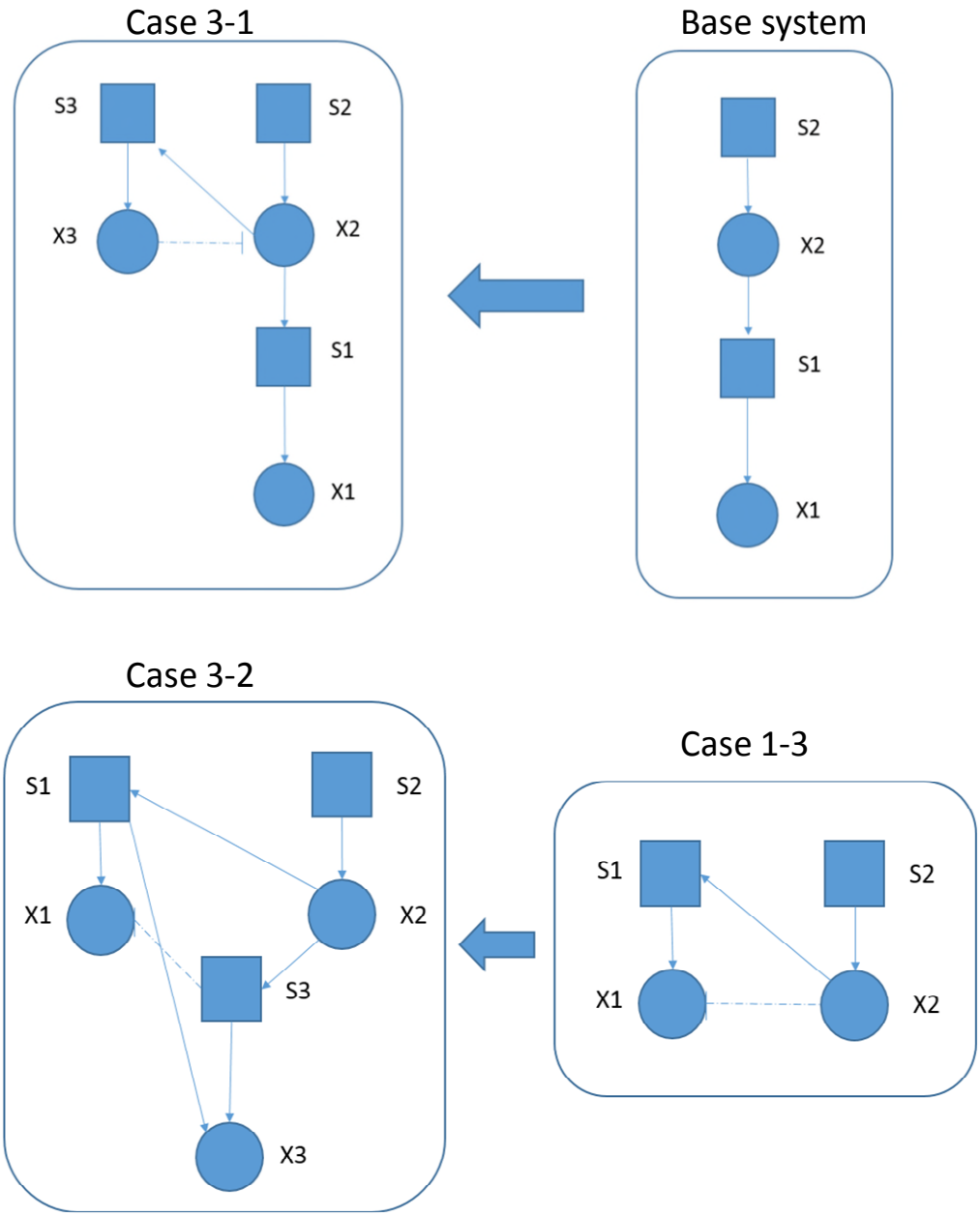


Figure 5 4. Construction of the two cases of three-species system (layer-3 cases).

The first one (case 3-1) is derived from the base system, now with X_2 producing not only S_1 but also S_3, as frequently encountered in industrial applications such as mixed-acid fermentation. To allow the additional intermediate product S_3 to be utilised, complementary to the utilisation of S_1 by X_1, we further introduce the third species, X_3,

to convert S_3 to a useful product (represented by X_3 itself). One of the complexities of engineering microbial communities that needs to be recognised when introducing a new species is the multiple effects of the species, which means it may cause unwanted complications (such as excreting a metabolite which is toxic to an existing species) while providing the intended function to the system. Reflecting this reality, we consider in case 3-1 that X_3 has a negative effect on X_2, modelled by inhibition.

The second three-species case (case 3-2) is constructed from case 1-3, with a newly introduced species X_3 to offer a facilitating function: easing the inhibition of X_2 on X_1. To incorporate this effect, we assume the inhibition is via a chemical S_3 produced by X_2 which negatively affects X_1, and the role of the new species, X_3, is to consume S_3, hence reducing the inhibition on X_2. Again, to include an unwanted side-effect for this case, we assume that the new species X_3 competes with X_1 for S_2.

Layer-2 case	Layer-1 cases being combined
2-1	1-1, 1-2
2-2	1-1, 1-3
2-3	1-1, 1-4
2-4	1-2, 1-3
2-5	1-2, 1-4
2-6	1-3, 1-4

Table 5. 2. Relations between layer 1 and layer 2 models.

5.2.2. Mathematical models

Well-mixed continuous bioreactors (i.e. chemostats) for the culture of multiple species are modelled with the commonly used Monod-type kinetics. For a two-species system, its mass balance is modelled by:

$$\frac{dX_i}{dt} = -DX_i + r_i \tag{5.1}$$

$$\frac{dS_1}{dt} = D\left(S_{1,\,in} - S_1\right) + Y_{1,2}r_2 - Y_{1,1}r_1 \tag{5.2}$$

$$\frac{dS_2}{dt} = D\left(S_{2,\,in} - S_2\right) - Y_{2,2}r_2 \tag{5.3}$$

$$D = {}^{1}/_{HRT} \tag{5.4}$$

where X_i represents the concentration of species *i*, *i* = 1,2; S_k is the concentration of chemical *k*, *k* = 1,2; $S_{k,\,in}$ is the inlet concentration of chemical *k*; HRT and *D* are the hydraulic retention time and the dilution rate, respectively; $Y_{k,i}$ is a yield factor, referring to the amount of chemical *k* that is produced or consumed by the growth of a unit amount of the biomass of microbial species *i*. The growth rates of the two species, r_i, are further modelled by

$$r_i = \mu_{max,i}\frac{S_i}{S_i + K_i}X_i \prod I_{\mathrm{i,j}} \tag{5.5}$$

where, for species *i*, $\mu_{max,i}$ is the maximum specific growth rate, K_i is the half saturation constant. $I_{\mathrm{i,j}}$ represents the inhibition factor on species *i* by chemical or species *j*, and is modelled as follows:

$$I_{i,j} = \begin{cases} \frac{1}{1+\phi_{i,j}\times(X_j or\ S_j)}, if\ inhibition\ is\ present \\ 1, otherwise \end{cases} \quad (5.6)$$

where *i* represents biological species being inhibited, $\phi_{i,j}$ is the inhibition constant.

To reduce the number of parameters, the above model is further converted to a dimensionless form. Below are equations for the dimensionless quantities to be discussed in the main text; other equations of the dimensionless model are given in Appendix A.

$$\theta_1 = \mu_{max,1} \times HRT \quad (5.7)$$

$$\theta_2 = \mu_{max,2} \times HRT \quad (5.8)$$

$$\overline{X_1} = Y_{1,1}X_1/K_1 \quad (5.9)$$

$$\overline{X_2} = Y_{1,1}X_2/K_1 \quad (5.10)$$

$$\gamma = K_2/K_1 \quad (5.11)$$

$$\bar{t} = t/HRT \quad (5.12)$$

$$\alpha_1 = Y_{1,2}/Y_{1,1} \quad (5.13)$$

$$\alpha_2 = Y_{2,2}/Y_{1,1} \quad (5.14)$$

$$\overline{s_{2,in}} = s_{2,in}/K_1 \quad (5.15)$$

The two three-species systems are modelled by extending the two-species model; details are given in Appendix B.

5.2.3. Quantification of stability and productivity

Three aspects of a multi-species bioreactor, collectively referred to as its "stability behaviour", are examined: (i) how fast the system which has deviated from a stable steady-state (with the co-existence of all species) following a disturbance can recover itself, that is, the rate of (re-)convergence to the previous steady-state; (ii) the size and position of the (productive) operating window, depicted by the range(s) of operating parameter(s), in which the system can reach a stable steady-state (i.e. with the co-existence of all species); and (iii) the occurrence of multiple steady states (i.e. bi-stability) or stable oscillation (i.e. stable periodic orbits).

To quantify the first aspect, the largest eigenvalue of the Jacobian matrix of a linearized dynamical system at a stable steady-state is calculated, and its extent of negativity (i.e. how far it is below zero) is used as an indicator, referred to as "stability" in the rest of the paper:

$$Stability = \max(|Re_i|, i = 1,2, \dots n) \tag{5.16}$$

where Re_i is the real part of the i^{th} of n eigenvalues; stability is calculated only for cases where $\max(\mathrm{Re}_i, i = 1,2, \dots n) < 0$.

The operating space of the bioreactor is represented by the range of the inlet concentration of substrate S_2, assuming S_1 is only produced internally by X_1; thus S_2 becomes the only feed to the bioreactor. For each of the cases in the three layers presented in Section 5.2.1, steady-state solutions are obtained analytically for each of

the selected points in the operating space, the model is linearised at each steady-state solution, and eigenvalues of the Jacobian matrix of the linearised model are calculated. This thus forms the basis for calculating the stability indicator (equation 5.16), as well as for quantifying the other two aspects of the stability behaviour. The operating window with stable co-existence corresponds to the range of inlet substrate concentration in which (i) each species has a positive steady-state concentration and (ii) the largest eigenvalue is negative. The multiplicity of stable steady states (i.e. bi-stability) can readily be detected by examining the steady-state solutions and the corresponding eigenvalues. The possible occurrence of a (stable) oscillatory behaviour is firstly identified by the detection of the eigenvalue with the largest real part becoming purely imaginary. Subsequently, dynamic simulation is carried out with parameter settings leading to such cases to confirm the occurrence of stable oscillation.

The productivity of a bioreactor at a steady state is defined as the amount of the produced output per unit amount of substrate supplied, and is quantified by:

$$Productivity = \frac{\overline{X_1}}{\overline{S_{2,in}}} \; or \; \frac{\overline{X_1}+\overline{X_3}}{\overline{S_{2,in}}} \tag{5.17}$$

depending on whether the bioreactor has X_1 as the only product or has both X_1 and X_3 as its products.

5.3. Results

5.3.1. The base system: two species with no inhibition

We start by analysing the behaviour and performance of the base system, where two species form a simple cross-feeding chain, without any inhibition effects. The model presented in Section 5.2.2 contains several kinetic- and yield-related parameters. Here, the variation in three parameters important for predicting the collective behaviour of the two species was considered: θ_1 and θ_2, representing the maximum amount of growth that can be achieved by the two species over a single hydraulic retention period; and α_1, as the ratio of yield factors of the two species that link the growth of the two species with the consumption or production of the intermediate substrate, S_1. As shown in Table 5.3, eight sets of parameter values were tested, and the change in productivity and stability (as defined in Section 5.2.3) along with the inlet substrate concentration is predicted for each parameter set. For clarity, only part of the results is shown in Figure 5.5 to demonstrate the main trends.

From Figures 5.5a and 5.5b, one can see (simply by visual inspection) that for both productivity and stability, high values for θ_1, θ_2 and α_1 generally lead to improvement, with parameter sets "e" (with highest values for all the three parameters) and "d" (with lowest values for all the three parameters) corresponding to the best and the worst performance, respectively. Another common trend is that, for each parameter set, both productivity and stability of this base system appear to improve as the inlet substrate

concentration increases. On the other hand, the difference in the parameter values is shown to cause variations in the lower bound of the inlet substrate concentration required for the co-existence of the two species (corresponding to the leftmost point on each curve in Figures 5.5a and 5.5b; at an inlet substrate concentration lower than this point, washout occurs to one or both species). Again, higher parameter values appear to allow the co-existence region to start at a lower inlet substrate concentration, and because of this, parameter sets “e” and “d” represent the two extremes also in this respect. No multiple steady states or stable oscillations were detected for the base system.

Parameter set	a	b	c	d	e	f	g	h
θ_1	3	3	2	2	3	3	2	2
θ_2	3	2	3	2	3	2	3	2
α_1	1	1	1	1	1.5	1.5	1.5	1.5
γ_1	1							
α_2	1							

Table 5. 3. Sets of parameters tested for the base system.

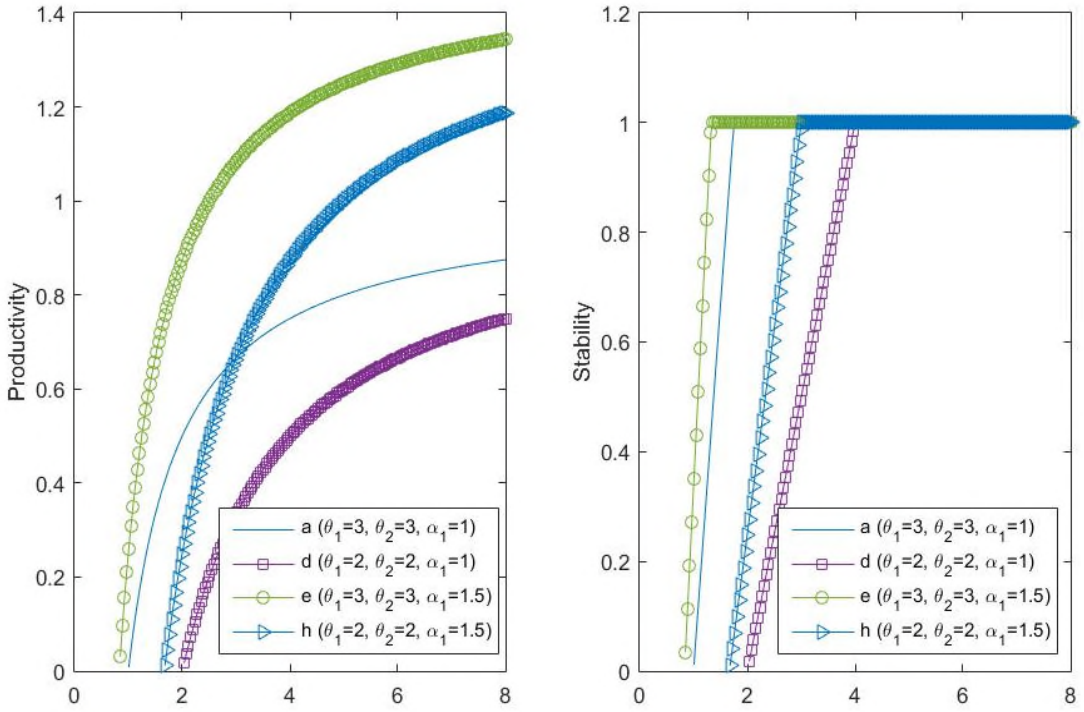


Figure 5. 5. Productivity (a) and stability (b) of the base system under different parameter settings.

5.3.2 Layer-1 cases: two species with one inhibition

The four cases in layer 1 are studied here. The results with parameter set "a" are shown in Figure 5.6; other parameter sets produced similar findings and are given in Appendix C. Each case was evaluated with three levels of inhibition strength. Note, case 1-3 adopted lower values of the inhibition constant ϕ than those for the other cases, because higher inhibitions would in this case lead to the extinction of X_1 under any inlet substrate concentrations. Besides, as the inlet concentration varies, different eigenvalues of the system may become the one that has the largest real part. When this happens, the stability curve becomes unsmooth.

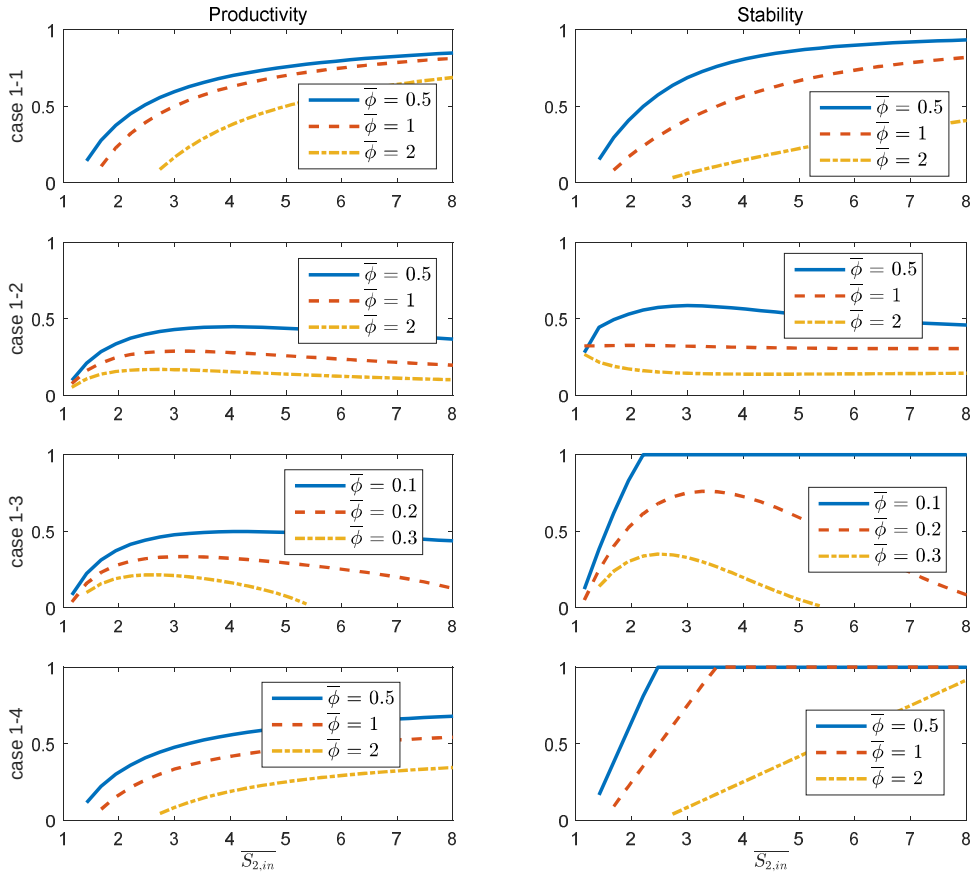


Figure 5. 6. Results for the layer-1 cases (θ_1=3, θ_2=3, α_1=1).

From Figure 5.6, it can be seen that, in all cases, an increase in the strength of a particular inhibition always leads to reduction in both productivity and stability. Case 1-1 (with the intermediate substrate inhibiting its producer) and case 1-4 (with the intermediate substrate inhibiting its consumer) demonstrate behaviours similar to the base system, in that the productivity and the stability generally both increase with the inlet substrate concentration.

In case 1-2 (with the downstream species X_1 inhibiting the upstream species X_2), when the inlet substrate concentration increases, X_2 tends to produce more S_1, and hence more X_1; however, the accumulation of X_1 will inhibit the activity of X_2, hence forming an indirect feedback loop. Beyond a certain inlet substrate concentration, the production of X_2 will begin to be reduced and may eventually stablise at a moderate (but not diminishing) level. Furthermore, with this specific community structure and at high inhibition strengths, the increase in productivity is not always accompanied by the increase in stability, see for example the curves for $\bar{\phi}$ = 2 in Figure 5.6 when the inlet substrate concentration is larger than 2. When the inhibition effect is significant enough, the microbial community loses the ability to stabilise at a steady-state and starts to oscillate at high inlet substrate concentrations, as shown in Figure 5.7, which is also a consequence of the feedback loop mentioned above.

In case 1-3 (with the upstream species X_1 inhibiting the downstream species X_2), the most significant difference compared with the other cases is that, the inhibition effect of the upstream species on the downstream species in this case tends to eliminate

the latter from the system, which occurs when the inlet substrate concentration is high. This means that an operating window with a limited size (as opposed to being open-ended) exists for case 1-3. Beyond the best inlet substrate concentration, both the productivity and the stability decline, until the downstream species ceases to survive. Besides, although the higher values of the three model parameters ($\theta_1, \theta_2, \alpha_1$) lead to the improvement in productivity and stability across all the layer-1 cases (see Appendix C), case 1-3 appears to be specifically sensitive to the change in parameter θ_1: when θ_1 becomes small, the uptake rate of S_1 decreases, and the inhibition affects the system to a greater extent. This makes the co-existence of the two species impossible at the highest inhibition strength for parameter sets “c”, “d”, “g” and “h” and also impossible at the medium inhibition strength for parameter sets “c” and “d”, regardless of the level of the inlet substrate concentration.

As a final note on layer 1, the inhibition in cases 1-1 and 1-4 shows an additional effect (see Figure 5.6): the minimum inlet substrate concentration needed for co-existence increases in comparison with the base system, and the shift becomes greater with the increase in the inhibition strength.

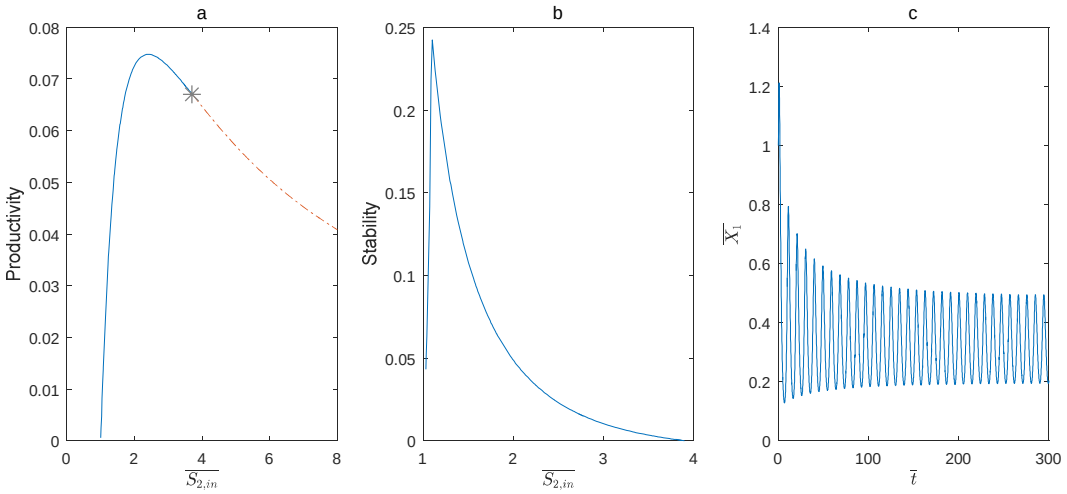
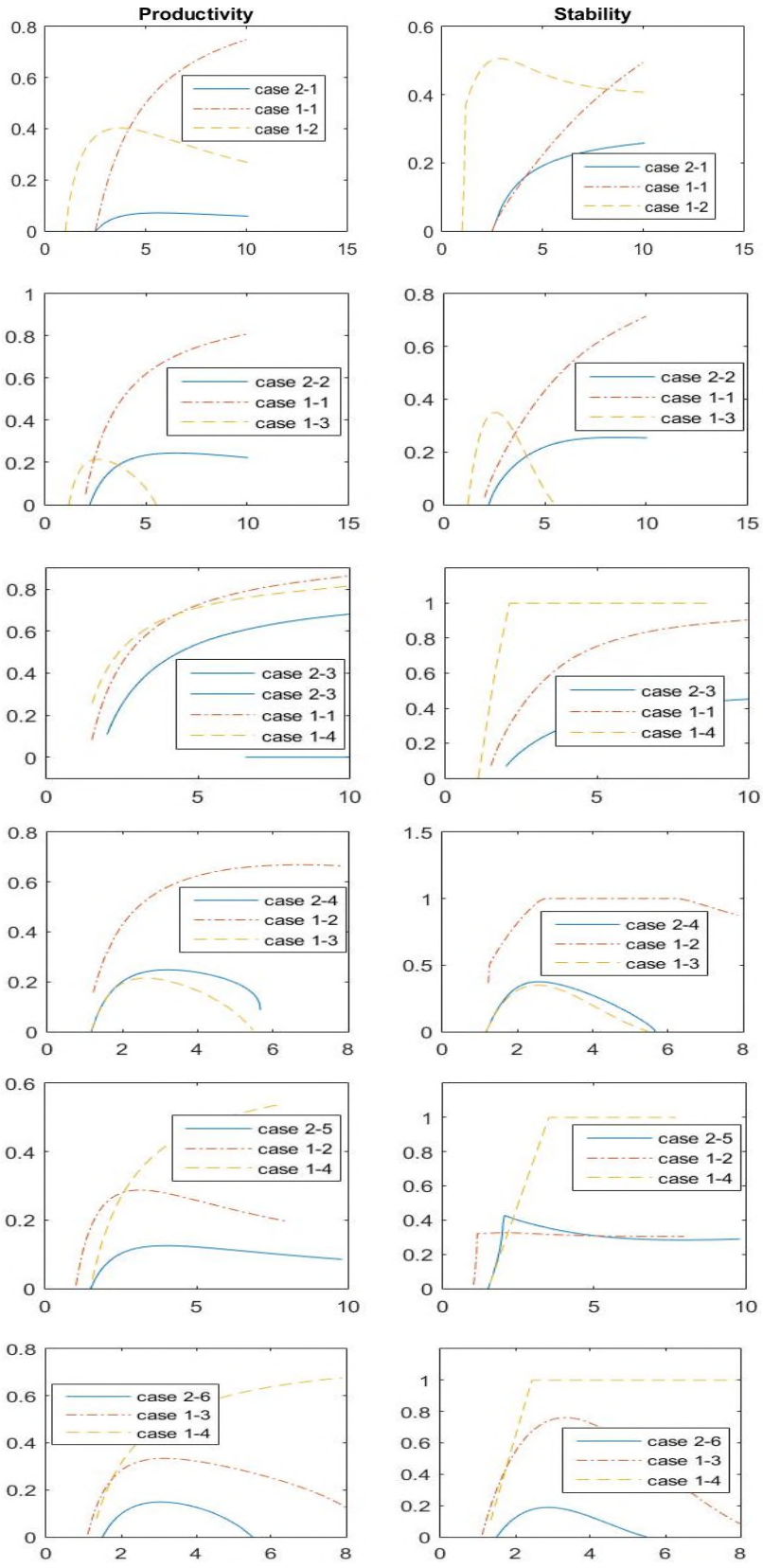


Figure 5. 7. Oscillation in case 1-2. $\bar{\phi}$ = 5, θ_1 = 3, θ_2 = 3 and α_1 = 1. a. Productivity, solid line represents stable steady state; dashed line represents oscillation region; '*' represents onset of stable oscillation. b. Stability at stable steady state. c. Numerical simulation of the stable oscillation, $\overline{s_{2,in}} = 6$.

5.3.3. Layer-2 cases: two-species with double inhibitions

As described earlier, the six layer-2 cases each contain two of the four individual inhibitions. Therefore, a primary concern of this part of the study is on the effect of the combination of two inhibitions. Since the study of the layer-1 cases has shown that applying different values for parameters θ_1, θ_2 and α_1 does not affect the qualitative characteristics of the single-inhibition systems (Appendix C), here only parameter set "a" was applied together with inhibition parameters (Table 5.4) to examine the layer-2 cases. The results are shown in Figure 5.8. From preliminary test, different cases require different range of parameters value. In order to maintain the feasibility of the simulated

systems, particularly in terms of co-existence, parameters in table 5.4 have been selected specifically for each case.



$\overline{S_{2,in}}$ $\overline{S_{2,in}}$

Figure 5. 8 Results of layer-2 cases in comparison with corresponding layer-1 cases.

	case 2-1	case 2-2	case2-3	case 2-4	case 2-5	case 2-6
Inhibition from case 1-1 ($\bar{\phi}_1$)	2	1.5	0.8			
inhibition from case 1-2 ($\bar{\phi}_2$)	0.6			0.2	1	
inhibition from case 1-3 ($\bar{\phi}_3$)		0.3		0.3		0.2
inhibition from case 1-4 ($\bar{\phi}_4$)			0.2		1	0.5
θ_1	3					
θ_2	3					
α_1	1					
α_2	1					
γ_1	1					

Table 5. 4 Parameter values adopted in the layer-2 cases.

The behaviour of the six layer-2 cases falls largely into two categories. In the first category, the two inhibitions impose their inhibitive effect on the system rather independently. It includes cases 2-1 (both S_1 and X_1 inhibiting X_2), 2-3 (S_1 inhibiting X_2 and S_2 inhibiting X_1), 2-5 (S_2 inhibiting X_1 and X_1 inhibiting X_2) and 2-6 (both S_2 and X_2 inhibiting X_1). In these cases, the combined effect of two inhibitions results in greater reduction in productivity and stability compared to the cases with any of the individual inhibitions, and may reduce the size of the operating window for co-existence as well (as in cases 2-3 and 2-6).

The two remaining layer-2 cases belong to the second category, where the two inhibitions are inter-connected, with the presence of one inhibition bringing an easing effect on the other. In case 2-4 (X_1 and X_2 inhibiting each other), this is achieved through a direct counter-action: the inhibition of X_1 to X_2 can give X_1 a better chance to survive the inhibition by X_2 at high inlet substrate concentrations. In case 2-2 (X_2 inhibiting X_1 and S_1 inhibiting X_2), the easing effect is realised through a negative feedback loop: reduced consumption of S_1 by X_1 following the inhibited growth of X_1 due to X_2 (first inhibition) can lead to the increased inhibition to X_2 by S_1 (second inhibition), due to the increased accumulation of S_1 resulting from the first inhibition. The presence of the second inhibition will reduce the growth of X_2, hence easing the first inhibition. This negative feedback loop thus allows X_1 to survive in this system at a high inlet substrate concentration which is impossible when only the first inhibition exists (i.e. as in case 1-3).

In addition to the aforementioned changes in productivity and stability, some of the layer-2 cases also demonstrate complex behaviours such as oscillation and bi-stability. The first ones to mention are cases 2-1, 2-4 and 2-5, all of which inherit the indirect feedback loop from case 1-2 which has been shown earlier to cause oscillation. All these three layer-2 cases were found to be able to give rise to oscillation at specific parameter settings, as shown in Figure 9 for case 2-5 as an example. In case 2-3, on the other hand, two stable steady states emerge when the inlet substrate concentration is high (see Figure 8, 3rd row, left), which means a disturbance could cause the system to

switch from the desirable steady state of co-existence and high productivity to the other one where X_1 is not able to survive. This bi-stability behaviour can be attributed to the positive feedback loop of X_2->S_2->X_1->S_1->X_2, a new structure formed due to the combination of two inhibitions, which can make the system drift away from a stable co-existence regime after it experiences a disturbance that causes a dip in X_2. The presence of bi-stability represents another situation where balanced consideration is needed on productivity and stability, as the bi-stability region is also where the productivity is higher.

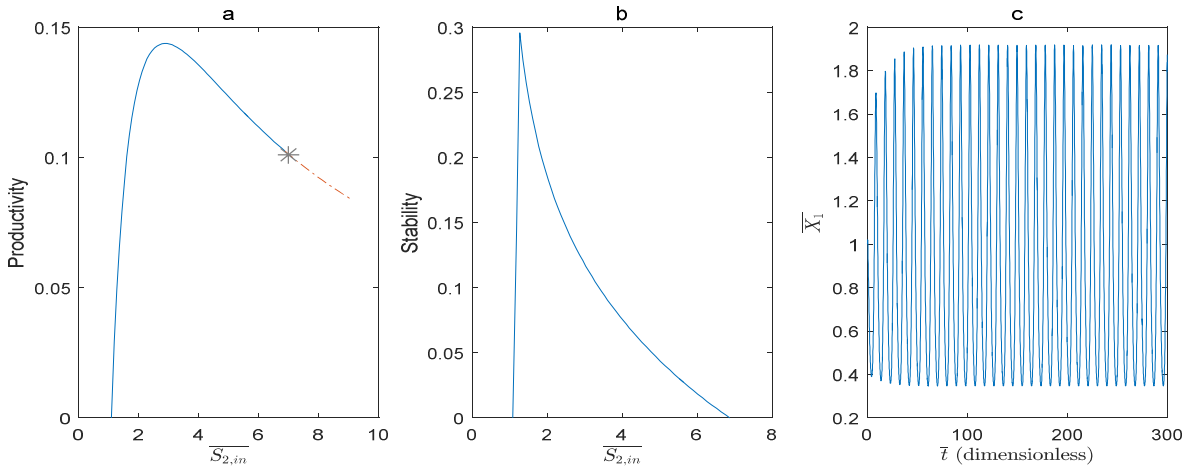


Figure 5. 9. Oscillation in case 2-5. $\bar{\phi}_2$ = 2, $\bar{\phi}_4$ = 0.2. (a) Productivity, solid line represents stable steady state, dashed line represents oscillation region, '*' represents onset of stable oscillation. (b) Stability at stable steady state. (c) Numerical simulation of the stable oscillation, $\overline{s_{2,in}}$=8.

5.3.4 Layer-3 cases: addition of a third species

Here we study the two cases where a third species is introduced to the bioreactor for improving its performance, which however also brings some unwanted side effect to the microbial community. Same as for layer 2, parameter set “a” has been used in the analysis. The complete set of parameter values for each of the layer-3 cases, together with their mathematical models, is given in Appendix B.

As described earlier, case 3-1 augments the base system and involves a third species (X_3), which is introduced to utilise a second (and otherwise unused) intermediate product (S_3), which at the same time undesirably exhibits inhibition to the existing X_2. From Figure 5.10, one can see that the change in the strength in the new inhibition alters the cost-benefit ratio. When the inhibition cost is sufficiently low, the new system can be more productive than the two-species counterpart, as expected. However, the introduction of the third species appears to reduce stability, regardless of the inhibition strength, and the change in stability with the inlet substrate concentration does not always synchronise with that in productivity. Structurally, case 3-1 resembles the combination of the base system and case 1-2; the latter was shown earlier (in Section 5.3.2) to exhibit oscillation when the inhibition from X_1 on X_2 is sufficiently high. In the three species system, this behaviour is again preserved, as shown in Figure 5.11, which is similar to the layer-2 cases that contain the same structure.

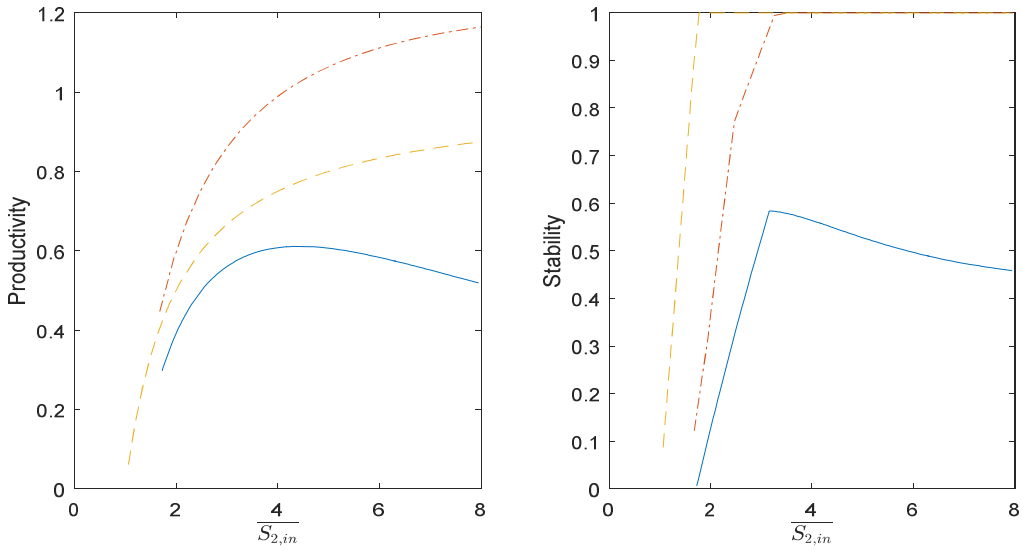


Figure 5. 10. Three-species system, case 3-1. Solid line represents three species systems with different parameter settings; dashed line represents the base system. The dotted line corresponds to a dimensionless strength of 0.2 of the new inhibition; the solid line corresponds to a strength of 1.

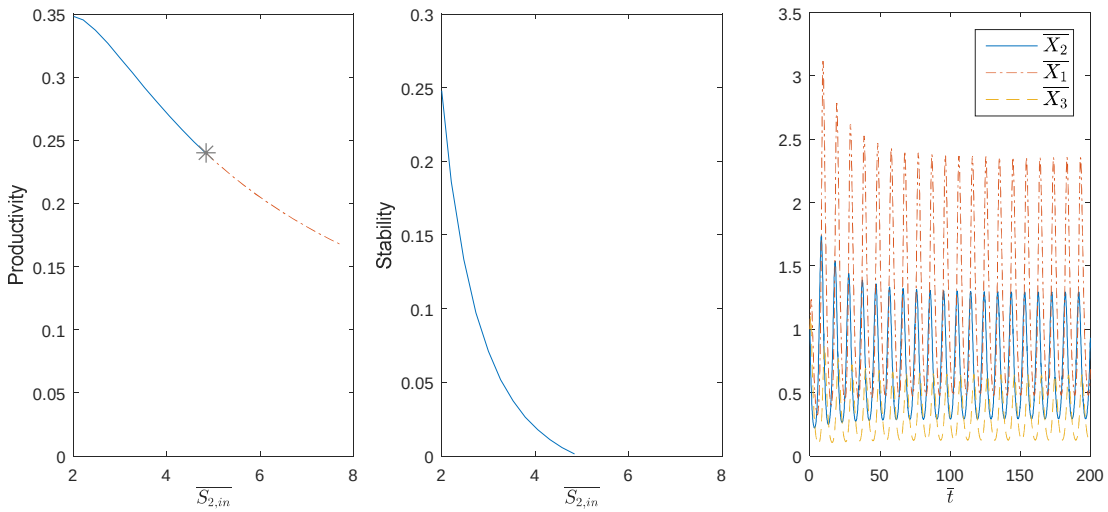


Figure 5. 11. Occurrence of stable oscillation in three-species system, case 3-1. The dimensionless strength of the new inhibition is 6. (a) Productivity, solid line represents stable steady state, dashed line represents oscillation region, '*' represents onset of stable oscillation. (b) Stability at stable steady state. (c) Numerical simulation of the stable oscillation, $\overline{s_{2,in}}$ = 6.

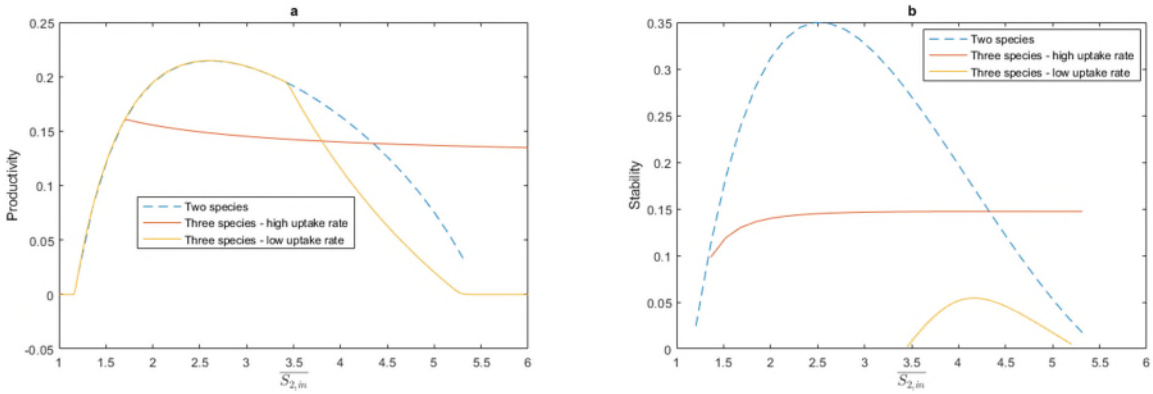


Figure 5. 12. Three-species system, case 3-2. (a) Productivity; (b) Stability.

In case 3-2, a third species (X_3) is introduced to case 2-3 to ease the inhibition from X_2 on X_1 by consuming the inhibiting metabolite S_3, but at the cost of X_3 competing with X_2 for S_1 (see Appendix B for parameter settings). As shown in Figure 5.12, with a high uptake rate by X_3 on S_3, this new species cannot co-survive with X_1 when the inlet substrate concentration is very low (below 1.5, see Figure 5.12, right plot). However, co-existence of all the three species becomes possible when the inlet substrate concentration increases, but the competition between X_3 and X_1 for metabolite S_1 results in a drop of productivity compared to the two-species system. The benefit of introducing the new species manifests only after the inlet substrate concentration becomes sufficiently high (over 4.5), where the gain by the removal of the inhibiting metabolite S_3 begins to overweigh the loss due to the competing consumption of the intermediate feed S_1.

However, if the uptake rate on S_3 by X_3 is low, under no circumstances does the introduction of the third species demonstrate any advantages in productivity, as the cost always overweighs the benefit.

Finally, regardless of the level of the update rate on S_3, this second case of three species shows constantly a reduced stability compared to its two-species counterpart, same as observed from the first case in layer 3.

5.4. Discussion

The dynamics and stability of complex ecosystems have attracted considerable theoretical interests [22-25], with seemingly conflicting results presented on important issues such as the relationship between diversity and stability [30]. Different from these past studies which have commonly focused on natural systems involving potentially a large number of species, this work focuses on engineered microbial communities with rather limited sizes which are realistic for controlled construction. In particular, variations of the commonly encountered pattern of a two-species forward-feeding chain have been investigated by introducing one or two inhibitions and further by considering the introduction of a third species for complementarity (in resource utilisation) and facilitation. In terms of stability that indicates how fast a system re-stablises to its previous steady state following a disturbance, all the two-species and three-species variations have shown a reduced stability compared to the simple forward-feeding chain. Some of the inter-species relations studied in this work can be regarded as competitive in the sense that the presence of one species imposes a negative impact on the growth of another species, e.g. the inhibition between X_1 and X_2. In [24], it was shown that the increase of competitive relationships in a (large) microbial community tends to increase the system's stability. Such a trend has not been observed in the systems studied in this work which are of much smaller sizes. For these systems, which are very much likely to be encountered in the design of bioreactors, the results of this work suggest a tendency of reduced stability as the complexity increases, which

deserves attention in engineering decisions. Besides, a very recent work [29] proved the stability of several structures of microbial communities which are not covered by the present work. Nevertheless, it suggested that the stability of a community could be affected not only by its structure but also by the strength of the inter-species relationships, which is consistent with the finding of the current study.

In all the two-species cases studied in this work, inhibitions are shown to also reduce productivity. In the two three-species cases, intended gains in productivity have been observed only when the cost of introducing the third species is relatively moderate. Besides, the operating condition (represented by the inlet substrate concentration) that yields a higher productivity does not necessarily lead to a higher stability: the trends of productivity and stability responding to the increase in inlet substrate concentration may be different (such as in cases 1-2 and 3-1), and the highest productivity may correspond to an operating condition where bi-stability occurs (as in case 2-3). These all suggest a possible conflict between the two engineering goals.

As well as reducing the stability in the co-existence region, three out of the four types of individual inhibition have shown a negative impact on the size of the productive operating window. Besides, the involvement of one or two inhibitions, when forming an indirect negative or positive feedback loop, has shown the *possibility* of giving rise to oscillation or bi-stability, respectively, similar to the behaviours well-known at the intracellular level due to biomolecular interactions [26, 27]. When a parameter setting favourable for productivity falls in the operating window where stable

oscillation sustains, the bioreactor will be expected to operate in the naturally occurring cycles as opposed to a stable state. However, the presence of such inter-species structures does not *necessarily* lead to a sophisticated behaviour. For example, we mentioned earlier that an indirect negative feedback loop exists in case 2-2, yet no oscillation was identified in this case. This is possibly because the inhibition strength from X_2 on X_1 needs to be rather low in order to make co-existence feasible, and a low strength of the inhibition does not render a sufficiently strong negative feedback needed for the oscillation to occur. This suggests that the occurrence of a complex behaviour in a synthetic microbial community (such as stable oscillation) may be determined by both its structural feature and the feasible parameter space.

There are existing studies on understanding the possibility of predicting the behaviour of a complex microbial community based on that of its basic elements. Using binary interactions to predict the dynamics of a system involving many species has previously been considered highly unreliable [18]. On the other hand, a recent work showed a significant level of accuracy in applying a simple rule to predict the co-existence of a microbial community with a large number of species based on the binary patterns [28]. In the present work, we have shown that the structural relationship between two separate inhibitions (being either “additive” or “opposing”, cf. Section 3.3) can well predict the effect of combining the two inhibitions on productivity and stability. Besides, the inheritance of a special structural feature (e.g. indirect negative feedback) by a two-inhibition case from a one-inhibition case well predicts the preservation of a

corresponding behaviour (e.g. oscillation). On the other hand, if a new (and important) structural feature occurs from the combination (e.g. a positive feedback, as in case 2-3), a new behaviour could emerge which cannot be predicted from the behaviour of the simpler systems. This suggests that for the synthetic microbial communities such as those studied in this work, the behaviour of a more complex system could be predicted, at least qualitatively, from that of its constituents, when there is no important new structure being formed as the complexity increases. However, even in such cases, detailed modelling analysis of the more complex system may still be needed for the purpose of engineering design, because quantitative (as opposed to qualitative) prediction could be important in order to shed light on issues such as productivity-stability trade-off, determination of the productive operational window, and cost-benefit analysis for introducing additional species.

Finally, since this work is confined to the analysis of bioreactors with a synthetic community that involves a small number of defined species, the dynamic evaluation of the community structure, often an interesting aspect of studying a community with a relatively large number of species, such as that in an anaerobic digester [32], has not been an important consideration here. Furthermore, the bioreactors have been assumed to be well-mixed; stability features arising from spatial heterogeneity [33] remain a subject of future work.

5.5. Conclusions

In this work, a number of representative structures of synthetic microbial community with two or three species have been evaluated in the context of a continuous bioreactor. Through bifurcation analysis and numerical simulation, it was shown that the introduction of inhibitions to a simple two-species cross-feeding chain generally reduces both productivity and stability of the bioreactor. When introducing a third species for complementary resource utilisation or for facilitating existing species, productivity gain can occur when the benefit of the intended mechanism overweighs the cost of any negative side effect of the additional species, but accompanied with a reduction in stability. Inhibitions were also shown to be able to reduce the productive operating window that corresponds to the co-existence region of the operating space. Complex behaviours such as stable oscillation and bi-stability may occur with certain structural features such as feedback loops, combined with a suitable range of parameter values. It was also learned that the behaviour of a system that combines simpler structures may be qualitatively predicted based on the knowledge of its simpler constituents, particularly when no new important structural feature arises from the combination. However, detailed studies of the more complex systems, including both analytical stability analysis and numerical simulation, are warranted for engineering purposes in order to obtain quantitative understandings needed to address issues such as the balances between productivity and stability of a bioreactor and between costs and benefits of introducing new species into an engineered microbial community.

References

[1] Großkopf, T., Soyer, O. S., 2014. Synthetic microbial communities. *Current Opinion in Microbiology*, 18, 72–77.

[2] Song, H., Ding, M. Z., Jia, X. Q., Ma, Q., & Yuan, Y. J. 2014. Synthetic microbial consortia: from systematic analysis to construction and applications. *Chemical Society Reviews*, *43*(20), 6954-6981.

[3] Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N., Silver, P. A., 2015. Better together: engineering and application of microbial symbioses. *Current Opinion in Biotechnology*, 36, 40–49.

[4] Lindemann, S. R., Bernstein, H. C., Song, H., Fredrickson, J. K., Fields, M. W., Shou, W., Johnson, D. R., Beliaev, A. S., 2016. Engineering microbial consortia for controllable outputs. *The ISME Journal*, 10, 2077–2084.

[5] Zhou, K., Qiao, K., Edgar, S., Stephanopoulos, G., 2015. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature biotechnology*, 33(4), 377-383.

[6] Konopka, A., Lindemann, S., Fredrickson, J., 2015. Dynamics in microbial communities: unravelling mechanisms to identify principles. *The ISME Journal*, 9, 1488–1495.

[7] Sari, T., Harmand, J., 2016. A model of a syntrophic relationship between two microbial species in a chemostat including maintenance. *Math. Biosci*. 275, 1-9.

[8] Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F., Hairston Jr, N. G., 2003. Rapid evolution drives ecological dynamics in a predator-prey system. *Nature*, 424, 303-306.

[9] Tilman, D., 1980. Resources: a graphical-mechanistic approach to competition and predation. *The American Naturalist*. 116, 362-393.

[10] Davison, B.H., Stephanopoulos, G., 1986. Effect of pH oscillations on a competing mixed culture. *Biotechnol. Bioeng.* 28, 1127–1137.

[11] Pavlou, S., Kevrekidis, I. G., Lyberatos, G., 1990. On the coexistence of competing microbial species in a chemostat under cycling. *Biotechnol. Bioeng*. 35, 224-232.

[12] Dikshitulu, S., Baltzis, B. C., Lewandowski, G. A., Pavlou, S., 1993. Competition between two microbial populations in a sequencing fed-batch reactor: Theory, experimental verification, and implications for waste treatment applications. *Biotechnol. Bioeng*. 42, 643-656.

[13] Lenas, P., Pavlou, S., 1994. Periodic, quasi-periodic, and chaotic coexistence of two competing microbial populations in a periodically operated chemostat. *Math. Biosci*. 121, 61-110.

[14] Kambam, P. K. R., Henson, M. A., Sun, L., 2008. Design and mathematical modelling of a synthetic symbiotic ecosystem. *IET Syst. Biol.* 2, 33-38.

[15] Thomopoulos, N. A., Vayenas, D. V., Pavlou, S., 1998. On the coexistence of three microbial populations competing for two complementary substrates in configurations of interconnected chemostats. *Math. Biosci.* 154, 87-102.

[16] Wade, M. J., Pattinson, R. W., Parker, N. G., Dolfing, J., 2016. Emergent behaviour in a chlorophenol-mineralising three-tiered microbial 'food web'. *J. Theor. Biol.* 389, 171-186.

[17] Weedermann, M., Wolkowicz, G. S., Sasara, J., 2015. Optimal biogas production in a model for anaerobic digestion. *Nonlinear Dynamics*, 81, 1097-1112.

[18] Dolinšek, J., Goldschmidt, F., Johnson, D. R., 2016. Synthetic microbial ecology and the dynamic interplay between microbial genotypes. *FEMS microbiology reviews*, *40*(6), 961-979.

[19] Lu, H., Lee, P. K., 2015. Effects of cellulose concentrations on the syntrophic interactions between Clostridium cellulovorans 743B and Rhodopseudomonas palustris CGA009 in coculture fermentation for biohydrogen production. *International Journal of Hydrogen Energy*, *40*(35), 11800-11808.

[20] Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C. H., Ahn, J. H., Foster, C. E., ... & Lin, X. N., 2013. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings of the National Academy of Sciences*, *110*(36), 14592-14597.

[21] Zuroff, T. R., Xiques, S. B., Curtis, W. R., 2013. Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture. *Biotechnology for biofuels*, *6*(1), 59.

[22] May, R. M., 1972. Will a large complex system be stable? *Nature*, 238, 413–414.

[23] Allesina, S., Tang, S., 2012. Stability criteria for complex ecosystems. *Nature*, *483*(7388), 205.

[24] Coyte, K. Z., Schluter, J., Foster, K. R., 2015. The ecology of the microbiome: networks, competition, and stability. *Science*, *350*(6261), 663-666.

[25] Liu, Y., Sumpter, D., 2017. Insights into resource consumption, cross-feeding, system collapse, stability and biodiversity from an artificial ecosystem. *Journal of The Royal Society Interface*, *14*(126), 20160816.

[26] Del Vecchio D, Dy A. J., Qian Y., 2016. Control theory meets synthetic biology. *J. R. Soc. Interface* 13, 20160380.

[27] Leloup, J.-C., Goldbeter, A., 2003. Toward a detailed computational model for the mammalian circadian clock. *PNAS*, 100(12), 7051–7056.

[28] Friedman, J., Higgins, L. M., Gore, J., 2017. Community structure follows simple assembly rules in microbial microcosms. *Nature ecology & evolution*, *1*(5), 0109.

[29] Butler, S, O'Dwyer, J. P., 2018. Stability Criteria for Complex Microbial Communities. *Nature Communications v(9)*, 2970.

[30] Haruta, S., Saito, Y., Futamata, H., 2016. Editorial: Development of Microbial Ecological Theory: Stability, Plasticity, and Evolution of Microbial Ecosystems. *Frontiers in microbiology*, *7*, 2069.

[31] Sari, T, Wade, M. J., 2017, Generalised approach to modelling a three-tiered microbial food-web. *Mathematical Biosciences*, 291, 21–37.

[32] Fernández A, Huang S, Seston S, et al., 1999. How stable is stable? Function versus community composition. *Appl Environ Microbiol*. 65(8):3697-704.

[33] Kim, H, J., Boedicker, J. Q., Choi, J. W., Ismagilov, R. F., 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proceedings of the National Academy of Sciences*, 105 (47) 18188-18193.

[34] Sierocinski, P., Milferstedt, K., Bayer, F., et al., 2017. A single community dominates structure and function of a mixture of multiple methanogenic communities. *Current Biology*, *27*(21), 3390-3395.

Appendix A. Additional equations of the dimensionless model for the two-species system

$$\frac{d\overline{X_1}}{d\bar{t}} = -\overline{X_1} + \overline{r_1}$$

$$\frac{d\overline{X_2}}{d\bar{t}} = -\overline{X_2} + \overline{r_2}$$

$$\frac{d\overline{S_1}}{d\bar{t}} = \left(\overline{S_{1,in}} - \overline{S_1}\right) + \alpha_1\overline{r_2} - \overline{r_1}$$

$$\frac{d\overline{S_2}}{d\bar{t}} = \left(\overline{S_{2,in}} - \overline{S_2}\right) - \alpha_2\overline{r_2}$$

$$\overline{r_1} = \theta_1 \frac{\overline{S_1}}{\overline{S_1}+1}\overline{X_1 I_{1,J}}$$

$$\overline{r_2} = \theta_2 \frac{\overline{S_2}}{\overline{S_2}+\gamma}\overline{X_2 I_{1,J}}$$

$$\overline{s_1} = s_1/K_1$$

$$\overline{s_2} = s_2/K_1$$

$$\bar{t} = t/HRT$$

$$\overline{s_{1,in}} = s_{1,in}/K_1$$

$$\overline{I_{i,J}} = \begin{cases} \frac{1}{1+\overline{\phi_{i,j}}\times(\overline{X_j} \text{ or } \overline{S_j})}, if\ inhibition\ is\ present \\ 1, otherwise \end{cases}$$

$$\overline{\phi_{i,J}} = \begin{cases} \frac{\phi_{i,j}K_1}{Y_{1,1}}, for\ inhibitions\ introduced\ in\ cases\ 1-2\ and\ 1-3 \\ \phi_{i,j}K_1, for\ other\ inhibitions \end{cases}$$

Appendix B. Mathematical models and parameter values for the three-species system

B.1 Model for case 3-1

$$\frac{dX_1}{dt} = {}^{1}/_{HRT}\,(0 - X_1) + \mu_{max,1}\frac{S_1}{S_1+K_1}X_1$$

$$\frac{dX_2}{dt} = {}^{1}/_{HRT}\,(0 - X_2) + \mu_{max,2}\frac{S_2}{S_2+K_2}X_2 I_{2,3}$$

$$\frac{dX_3}{dt} = {}^{1}/_{HRT}\,(0 - X_3) + \mu_{max,3}\frac{S_3}{S_3+K_3}X_3$$

$$\frac{dS_1}{dt} = {}^{1}/_{HRT}\left(S_{1,\,in} - S_1\right) + Y_{1,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2 - Y_{1,1}\mu_{max,1}\frac{S_1}{S_1+K_1}X_1$$

$$\frac{dS_2}{dt} = {}^{1}/_{HRT}\left(S_{2,\,in} - S_2\right) - Y_{2,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2$$

$$\frac{dS_3}{dt} = {}^{1}/_{HRT}\left(S_{3,\,in} - S_3\right) + Y_{3,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2 - Y_{3,3}\mu_{max,3}\frac{S_3}{S_3+K_3}X_3$$

$$I_{2,3} = \frac{1}{1+\phi_{2,3}X_3}$$

Note, $I_{2,3}$ represents inhibition to X_2 by X_3, and $\phi_{2,3}$ is the corresponding inhibition constant.

Converting the above model to a dimensionless form:

$$\theta_1 = \mu_{max,1} \times HRT$$

$$\theta_2 = \mu_{max,2} \times HRT$$

$$\theta_3 = \mu_{max,3} \times HRT$$

$$\overline{X_1} = Y_{1,1}X_1/K_1$$

$$\overline{X_2} = Y_{1,1}X_2/K_1$$

$$\overline{X_3} = Y_{1,1}X_3/K_1$$

$$\gamma_2 = K_2/K_1$$

$$\gamma_3 = K_3/K_1$$

$$\overline{S_1} = S_1/K_1$$

$$\overline{S_2} = S_2/K_1$$

$$\overline{S_3} = S_3/K_1$$

$$\bar{t} = t/HRT$$

$$\alpha_{1,2} = Y_{1,2}/Y_{1,1}$$

$$\alpha_{2,2} = Y_{2,2}/Y_{1,1}$$

$$\alpha_{3,2} = Y_{3,2}/Y_{1,1}$$

$$\alpha_{3,3} = Y_{3,3}/Y_{1,1}$$

$$\overline{S_{1,in}} = S_{1,in}/K_1$$

$$\overline{S_{2,in}} = S_{2,in}/K_1$$

$$\overline{I_{2,3}} = \frac{1}{1 + \overline{\phi_{2,3}}\overline{X_3}}$$

$$\overline{\phi_{2,3}} = \frac{\phi_{2,3}K_1}{Y_{1,1}}$$

The differential equations now become:

$$\frac{d\overline{X_1}}{d\bar{t}} = -\overline{X_1} + \theta_1 \frac{\bar{S_1}}{\bar{S_1} + 1}\overline{X_1}$$

$$\frac{d\overline{X_2}}{d\bar{t}} = -\overline{X_2} + \theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2 I_{2,3}}$$

$$\frac{d\overline{X_3}}{d\bar{t}} = -\overline{X_1} + \theta_3 \frac{\bar{S_3}}{\bar{S_3} + \gamma_3}\overline{X_1}$$

$$\frac{d\bar{S_1}}{d\bar{t}} = \left(\overline{S_{1,in}} - \bar{S_1}\right) + \alpha_{1,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2} - \theta_1 \frac{\bar{S_1}}{\bar{S_1} + 1}\overline{X_1}$$

$\frac{d\bar{S_2}}{d\bar{t}} = \left(\overline{S_{2,in}} - \bar{S_2}\right) - \alpha_{2,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2}+\gamma_2}\overline{X_2}$ $\frac{d\bar{S_3}}{d\bar{t}} = \left(\overline{S_{3,in}} - \bar{S_3}\right) + \alpha_{3,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2}+\gamma_2}\overline{X_2} -$

$\alpha_{3,3}\theta_3 \frac{\bar{S_3}}{\bar{S_3}+\gamma_3}\overline{X_3}$

B.2 Model for case 3-2

$$\frac{dX_1}{dt} = {}^{1}/_{HRT}\,(0 - X_1) + \mu_{max,1}\frac{S_1}{S_1+K_1}X_1 I_{1,3}$$

$$\frac{dX_2}{dt} = {}^{1}/_{HRT}\,(0 - X_2) + \mu_{max,2}\frac{S_2}{S_2+K_2}X_2$$

$$\frac{dX_3}{dt} = {}^{1}/_{HRT}\,(0 - X_3) + \mu_{max,3,3}\frac{S_3}{S_3+K_{3,3}}X_3 + \mu_{max,3,1}\frac{S_1}{S_1+K_{3,1}}X_3$$

$$\frac{dS_1}{dt} = {}^{1}/_{HRT}\left(S_{1,\,in} - S_1\right) + Y_{1,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2 - Y_{1,1}\mu_{max,1}\frac{S_1}{S_1+K_1}X_1 - Y_{1,3}\mu_{max,3,1}\frac{S_1}{S_1+K_{3,1}}X_3$$

$$\frac{dS_2}{dt} = {}^{1}/_{HRT}\left(S_{2,\,in} - S_2\right) - Y_{2,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2$$

$$\frac{dS_3}{dt} = {}^{1}/_{HRT}\left(S_{3,\,in} - S_3\right) + Y_{3,2}\mu_{max,2}\frac{S_2}{S_2+K_2}X_2 - Y_{3,3}\mu_{max,3,3}\frac{S_3}{S_3+K_{3,3}}X_3$$

$$I_{1,3} = \frac{1}{1+\phi_{1,3}S_3}$$

$\mu_{max,3,3}$ and $\mu_{max,3,1}$ are maximum growth rate of X_3 by up-taking S_3 and S_1, respectively, and $K_{3,3}$ and $K_{3,1}$ are the corresponding half-saturation constants. In this case, we assume the growth of X_3 is contributed by uptaking S_1 and S_3 at the same time, in other words, uptaking either substrates are independent and without priority.

Converting the above model to a dimensionless form:

$$\theta_1 = \mu_{max,1} \times HRT$$

$$\theta_2 = \mu_{max,2} \times HRT$$

$$\theta_{3,1} = \mu_{max,3,1} \times HRT$$

$$\theta_{3,3} = \mu_{max,3,3} \times HRT$$

$$\overline{X_1} = Y_{1,1}X_1/K_1$$

$$\overline{X_2} = Y_{1,1}X_2/K_1$$

$$\overline{X_3} = Y_{1,1}X_3/K_1$$

$$\gamma_2 = K_2/K_1$$

$$\gamma_{3,1} = K_{3,1}/K_1$$

$$\gamma_{3,3} = K_{3,3}/K_1$$

$$\overline{S_1} = S_1/K_1$$

$$\overline{S_2} = S_2/K_1$$

$$\overline{S_3} = S_3/K_1$$

$$\bar{t} = \mathrm{t}/HRT$$

$$\alpha_{1,2} = Y_{1,2}/Y_{1,1}$$

$$\alpha_{2,2} = Y_{2,2}/Y_{1,1}$$

$$\alpha_{3,2} = Y_{3,2}/Y_{1,1}$$

$$\alpha_{3,3} = Y_{3,3}/Y_{1,1}$$

$$\alpha_{1,3} = Y_{1,3}/Y_{1,1}$$

$$\overline{S_{1,in}} = S_{1,in}/K_1$$

$$\overline{S_{2,in}} = S_{2,in}/K_1$$

$$\overline{I_{1,3}} = \frac{1}{1 + \overline{\phi_{1,3}}\overline{S_3}}$$

$$\overline{\phi_{1,3}} = \phi_{1,3}K_1$$

The differential equations now become:

$$\frac{d\overline{X_1}}{d\bar{t}} = -\overline{X_1} + \theta_1 \frac{\bar{S_1}}{\bar{S_1} + 1}\overline{X_1 I_{1,3}}$$

$$\frac{d\overline{X_2}}{d\bar{t}} = -\overline{X_2} + \theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2}$$

$$\frac{d\overline{X_3}}{d\bar{t}} = -\overline{X_1} + \theta_{3,3} \frac{\bar{S_3}}{\bar{S_3} + \gamma_{3,3}}\overline{X_1} + \theta_{3,1} \frac{\bar{S_1}}{\bar{S_1} + \gamma_{3,1}}\overline{X_3}$$

$$\frac{d\bar{S_1}}{d\bar{t}} = \left(\overline{S_{1,in}} - \bar{S_1}\right) + \alpha_{1,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2} - \theta_1 \frac{\bar{S_1}}{\bar{S_1} + 1}\overline{X_1} - \alpha_{1,3}\theta_{3,1} \frac{\bar{S_1}}{\bar{S_1} + \gamma_{3,1}}\overline{X_3}$$

$$\frac{d\bar{S_2}}{d\bar{t}} = \left(\overline{S_{2,in}} - \bar{S_2}\right) - \alpha_{2,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2}$$

$$\frac{d\bar{S_3}}{d\bar{t}} = \left(\overline{S_{3,in}} - \bar{S_3}\right) + \alpha_{3,2}\theta_2 \frac{\bar{S_2}}{\bar{S_2} + \gamma_2}\overline{X_2} - \alpha_{3,3}\theta_3 \frac{\bar{S_3}}{\bar{S_3} + \gamma_{3,3}}\overline{X_3}$$

B.3 Parameter values adopted

Case	Parameters	value
Case 3-1	γ_2, γ_3	1
	θ_1	3
	θ_2	3
	θ_3	1
	$\alpha_{1,2}$	1
	$\alpha_{2,2}$	1
	$\alpha_{3,3}$	1
	$\alpha_{2,3}$	0.5
	$\bar{\phi}_{2,3}$	various, see text
Case 3-2	γ_2, $\gamma_{3,1}$, $\gamma_{3,3}$	1
	θ_1	3
	θ_2	3
	$\theta_{3,1}$	1
	$\theta_{3,3}$	1
	$\alpha_{1,2}$	1
	$\alpha_{2,2}$	1
	$\alpha_{3,2}$	1
	$\alpha_{1,3}$	1
	$\alpha_{3,3}$	1
	$\bar{\phi}_{1,3}$	0.3

Table B1. Parameter values adopted by the two layer-3 cases.

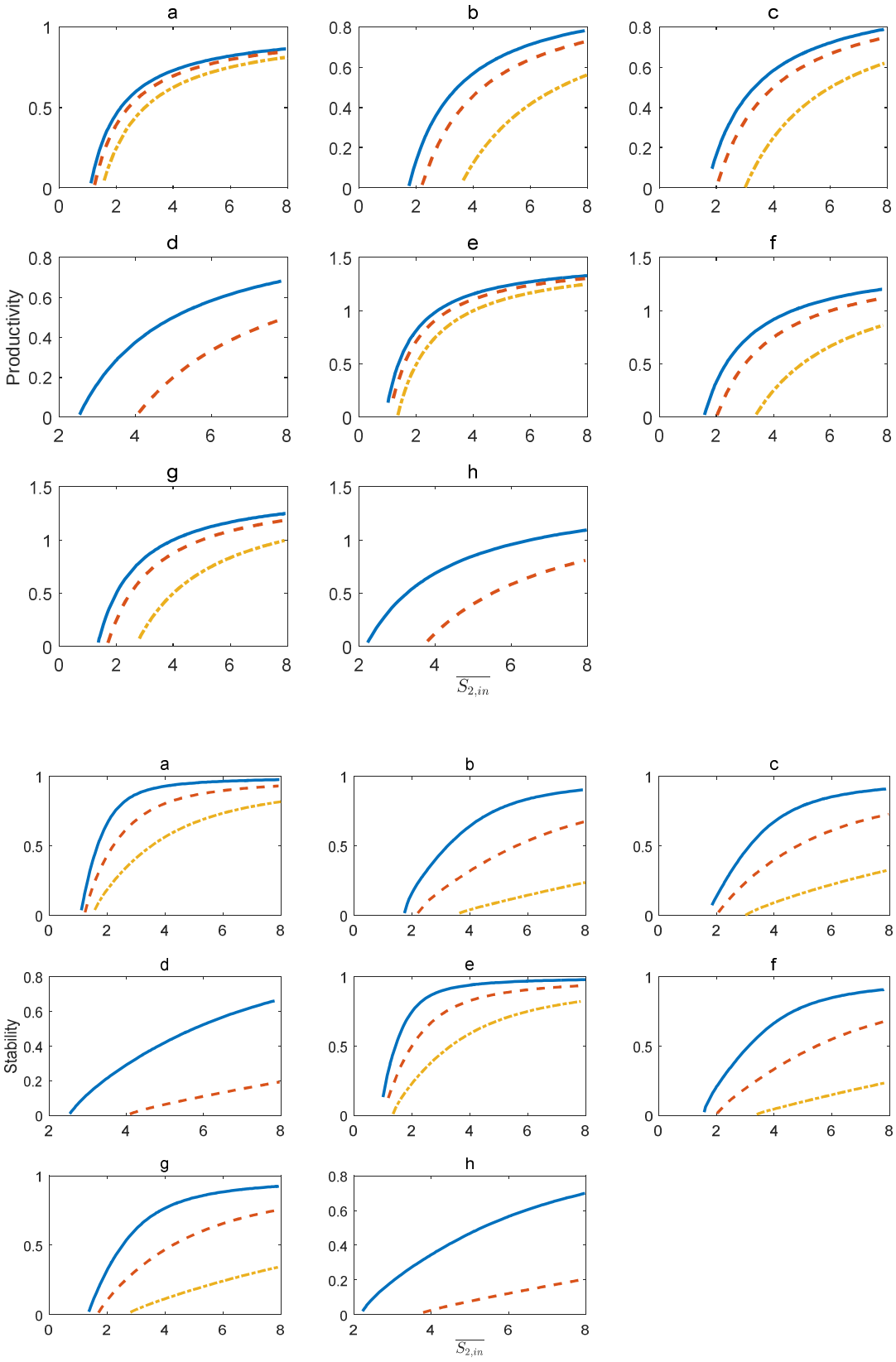


Figure C1: Case 1-1.

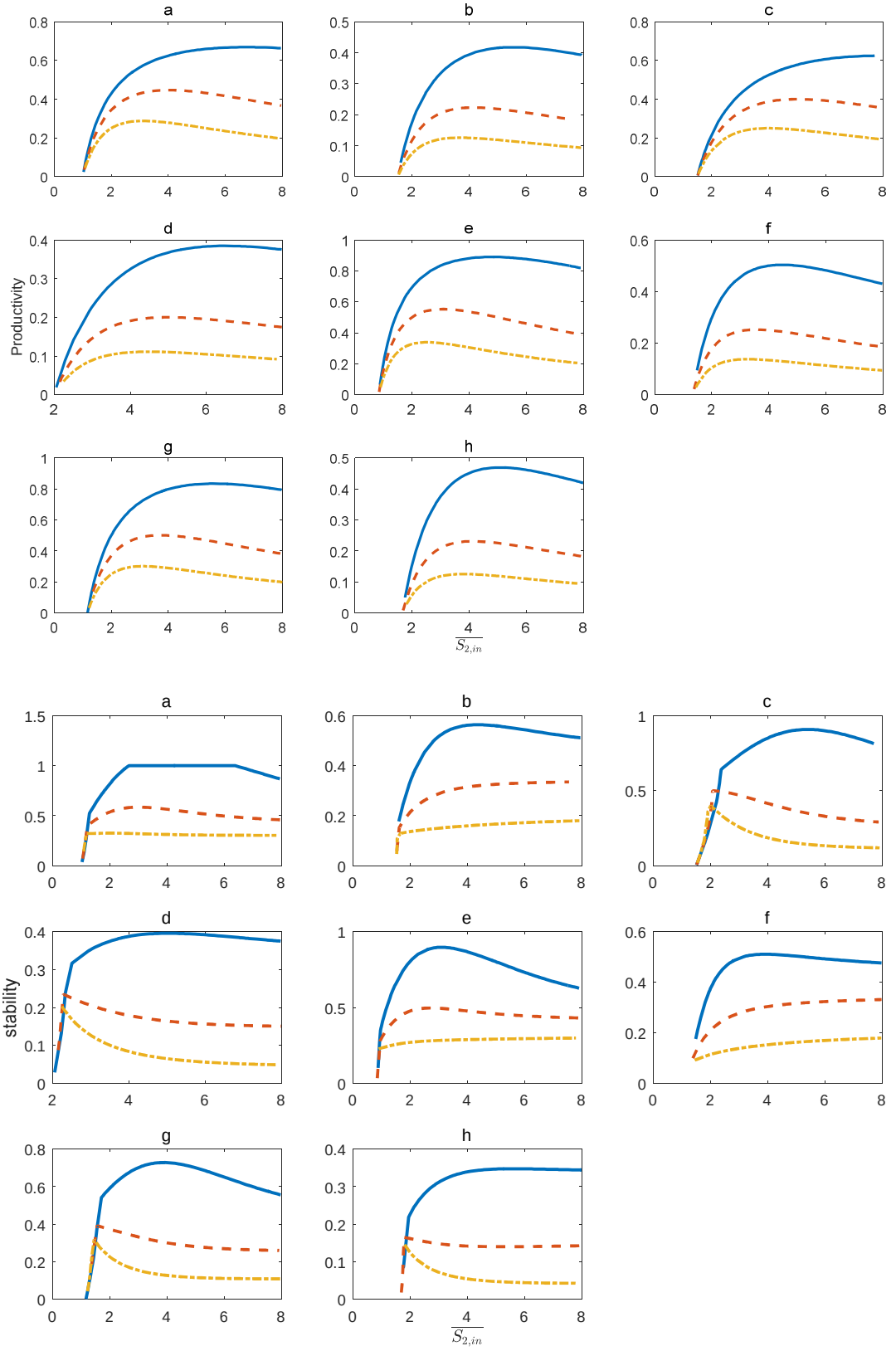


Figure C2: Case 1-2.

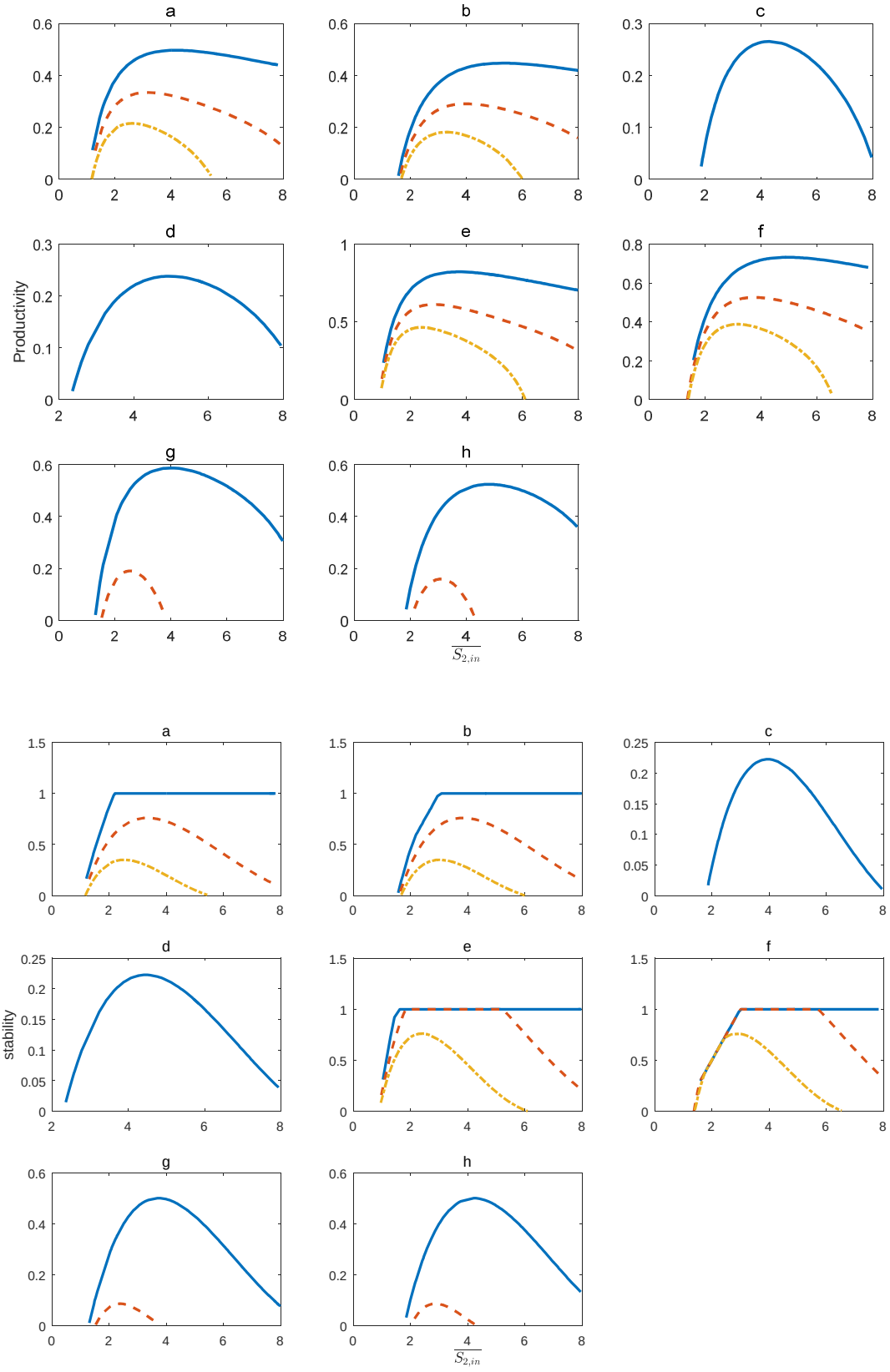
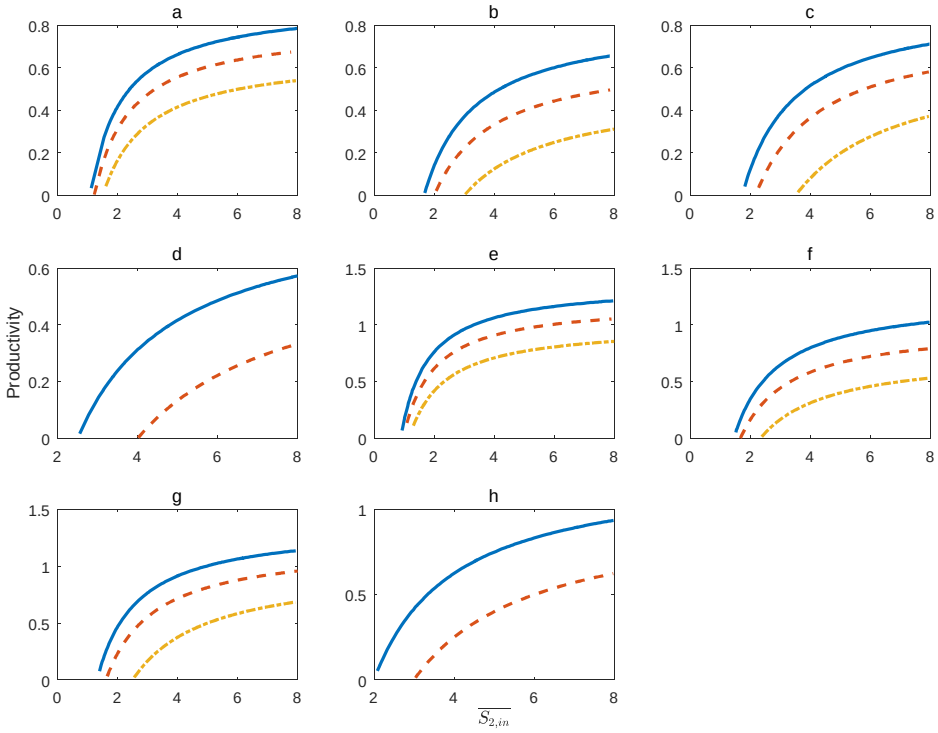


Figure C3: Case 1-3.



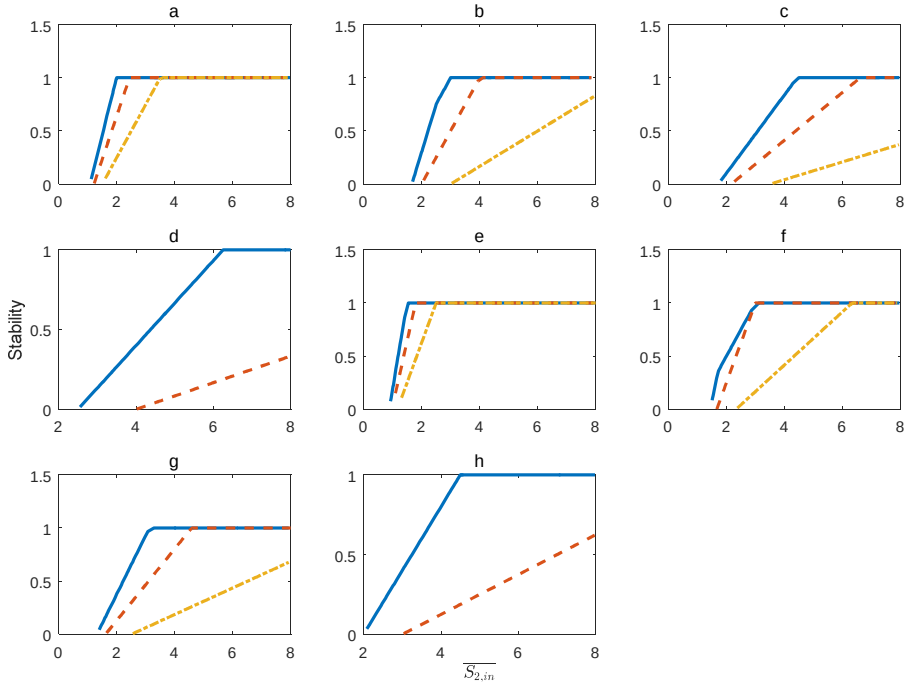


Figure C4: Case 1-4.

Chapter 6. Ecological network inference in microbial community with multiple inter-species interactions through chemicals

Summary

In previous chapters, the systems analysed contain a small number of microbial species with defined traits and ecological roles, as expected in bioreactors utilising an artificially constructed community. This chapter moves on to investigate systems that involve a large number of species which are not necessarily clearly identified individually, as in the cases of naturally occurring communities. Typical examples such as anaerobic digestion and human gut communities have been known to be relevant for engineering or medical purposes. For this type of systems, choosing appropriate statistical modelling tools to infer the pairwise interactions of the species within the community, i.e. the community network structure, is an important research challenge, as the inferred network structure has great potential to guide the understanding and manipulation of such complex microbial systems. Recognising the fact that most microbial interactions are mediated by chemicals (substrates, signalling molecules, and other metabolites), this work aims to systematically evaluate an existing network inference model in comparison with a new model proposed in this study, to shed light on the potential of these models in elucidating the structure of chemical-mediated microbial interactions within a complex community.

This work uses data generated from numerical simulation based mechanistic (ground-truth) models for anaerobic digestion and a number of artificial microbial communities with varying complexities. The generalized Lotka Volterra (gLV) model, used by several recent studies for microbial network inference, and a model proposed by this work which explicitly captures chemical-mediated interactions (CMI), are evaluated thoroughly with data generated with different perturbation strategies and noise levels. Lasso (Least Absolute Shrinkage and Selection Operator) based regularization is applied to obtain sparse solutions from both models to infer network structures, with ROC (Receiver Operating Characteristic) curves to measure the quality of inferred networks.

The results show that, in general, network inference benefits from data generated with stronger perturbation and lower measurement noise, as expected. In all cases with modest perturbation, gLV performs stably in inferring the food webs presented in the networks with a reasonable quality, whereas CMI performs generally worse than gLV due to the accumulation of errors in the construction of microbe-microbe interactions from microbe-chemical interactions. However, CMI does show the potential to infer both the food web and inhibitions in a microbial community, when proper perturbation strategies are applied. In comparison, gLV has shown difficulties in inferring inhibitions in all tested cases.

Overall, this work reveals important factors for the widely adopted gLV model to achieve successful network inference for microbial communities, shows both the

disadvantages as well as the potential of the newly proposed CMI model, and calls for further research to improve CMI-like models that establish the structure of a microbial community based on microbe-chemical links. Besides, this work highlights the importance of data, particularly in terms of the richness of information, as a key to the inference of complex networks.

List of symbols:

D: dilution rate

$KI_{i,k}$: inhibition factor of chemical species k on microbial species i

K_j: constant to represent the positive or negative impact of chemical species j on the growth of microbial species i

r_i: growth rate of species i

$S_{j,in}$: chemical species j inlet concentration

S_j: chemical species j concentration

X_i: abundance of species i

$X_{i,in}$: inlet concentration of microbial species i

$Y_{i,j}$ yield coefficient for the production (when positive) or consumption (when negative) of chemical species j by the growth of microbial species i

$\beta_{i,j}$: effect of species j on species i

μ_i: specific growth rate of species i

$\mu_{i,max}$: maximum growth rate of microbial species i

σ: perturbation level

6.1. Introduction

Naturally occurring microbial communities are typically complex and dynamic ecosystems which may play essential roles in various fields such as waste degradation [1] and human health [2]. Modern high-throughput sequencing based techniques such as 16S rRNA gene profiling, provide potential to elucidate the complex network structures of natural microbial communities [3]. As an important tool for analysing complex microbial communities, a microbial network comprises nodes representing microbial species, with the links between the nodes representing inter-species interactions, such as feeding, inhibition and/or competition [4].

With a successfully constructed network, one is able to identify the key microbial species in the community (centrality), as well as key clusters or subsets of species more densely connected with each other than with species outside a cluster [5]. A known network structure can also play an important role in designing control strategies. In 2011, Liu and his team proposed a 'controllability' theory for general networks with linear dynamics, which deals with the selection of a suitable set of inputs for driving a network from an initial state to a desired state. This work has more recently been extended to networks with nonlinear dynamics applicable to microbial communities [6] [7]. A parallel approach to the control of networks using structural information is based on the concept of "feedback vertex set", which can be manipulated so that the underlying system is steering to a specific steady state or an attractor of another type

[8] [9]. The latter approach has not been applied to the control of microbial communities, but holds potential to do so given its generic nature.

Both direct and indirect microbial community networks can be constructed from meta-genomic sequencing data; Table 1 lists a number of modelling approaches or tools. Friedman and Alm introduced a technique, called SparCC (Sparse Correlations for Compositional data), for inferring correlations between microbial species from compositional data [10]. Similarly, Fang et al. proposed a model called CCLasso (Correlation inference for Compositional data through Lasso) based on l_1-*norm* penalty on the off-diagonal variance-covariance matrix after log ratio transformation for compositional data [11]. Instead of inferring the correlation matrix directly, Kurtz et al. applied Graphical Lasso, to learn the structure of networks describing the interactions between microbial species and between microbial species and environmental factors from the precision matrix [12]. Schwager et al. proposed a Bayesian approach to quantify uncertainties for the structure of such learned networks [13].

These correlation-based methods rely on pairwise associations and result in undirected networks. A different category of approaches has been developed to construct directed networks from the abundance of microbial species using time series data. Such networks have directed links, each pointing from one species to other species to represent the influence of the former on the latter. Ridenhour et al. used an ARIMA (Autoregressive integrated moving average) model with elastic-net regularization to construct a network model for the gut microbiome of white-throated woodrats [14].

Gibbons et al. applied sparse vector autoregressive model (sVAR) to model the gut community [15]. Chen et al. proposed a linear state space model (SSM) to capture the dynamics of the vaginal microbiome community, with a proposed regularised expectation maximization algorithm to determine the network structure [16].

Beside the auto-regression based models, the generalized Lotka Volterra model (gLV) is another model that can be built from time series data to capture the dynamics of microbial species resulting from inter-species connections, which was proposed by Stein et al [17]. With this model, a number of algorithms have been developed to obtain a sparse and reliable (directed) microbial community network. For instance, LIMITS (Learning Interactions from Microbial Time Series) was proposed that combines a linear regression with a bagging strategy [18], Shaw et al. developed MetaMIS (Metagenomic Microbial Interaction Simulator) with an abundance-ranking strategy to obtain a sparse network [19]. Bucci et al. presented Microbial Dynamical Systems Inference Engine (MDSINE), an open-source software toolkit to model microbiome dynamics and infer the network structure from time series data with the extended gLV equations as underlying dynamical system model [20]. A summary of these models is provided in table 1.

Most network inference tools aim to construct networks to capture interactions between microbial species, with only a few studies have also considering the impact of important environmental factors, e.g. diet information in human gut microbiomes [14] [20]. However, most interactions between microbial species are mediated by chemicals,

such as the exchange of metabolites along a food chain or network [23], inhibition [31] and facilitation [22] to growth by metabolites, and the exchange of chemical signals (e.g. quorum sensing [21]). Momeni et al. reported that Lotka-Volterra pairwise modelling does not guarantee the accurate capture of microbial dynamics arising from interactions between microbial species to chemicals [24]. Our previous work also demonstrates the complexity of microbial communities dynamics mediated by chemicals [25]. These suggest that it is still far from clear how an approach using gLV, one of the most widely used models, would perform for inferring the structure of a microbial community involving microbe-chemical interactions of various levels of complexity.

	Modelling approach or tool	Description	ref.
Undirected network	SparCC	Correlation based for log transformation of compositional data	[10]
	CCLasso	Correlation based inference through Lasso regularization	[11]
	SPIEC-EASI	Sparse inverse covariance estimation (Graphical Lasso)	[12]
	BAnOCC	Bayesian approach for estimation log-basis precision matrix	[13]
Directed Network	ARIMA	ARIMA model with elastic-net regularization	[14]
	sVAR	Sparse autoregressive model with multiple time lags (3 time lag)	[15]
	gLV	Linearsization and discretization of Lotka-Volterra equations	[17]
	LIMIT	gLV model with bagging strategy to obtain sparse solution	[18]
	MetaMIS	gLV model with abundance-ranking based strategy to get sparse network	[19]
	MDSINE	A suite of algorithms for inferring microbial community networks	[20]

Table 6. 1. Summary of existing network inference models or tools.

In this work, we aim to answer two questions:

(1) How well can network inference based on gLV (which directly models microbe-microbe interactions) perform when dealing with a chemical-mediated complex microbial community?

(2) How well can an alternative model that captures microbe-chemical interactions perform in inferring the structure of a microbial community, in comparison with gLV?

To answer these questions, we make use of a variety of data sets generated by numerical simulation, to allow thorough and systematic tests that span a good range of complexities (in terms of, for example, number of species, number of layers of the networks, and sparsity) which would be difficult to achieve by using experimental data. Simulation-generated data have also been used in several previous studies (e.g. [26], [27]). However, to our best knowledge these existing studies directly generated microbial species abundance data using gLV or similar models without explicitly modelling microbe-chemical interactions.

6.2. Methodology

6.2.1. Inference models

6.2.1.1. Generalized Lotka Volterra (gLV) model

Extending the original Lotka-Volterra model, a time dependent deterministic model for a microbial community of L number of species is given by a system of coupled first order ordinary differential equations [17]:

$$\frac{dX_i}{dt} = \mu_i X_i + X_i \sum_{j=1}^{L} \beta_{i,j} X_j \tag{6.1}$$

where X_i represents the abundance of species i, μ_i is the specific growth rate of species i, $\beta_{i,j}$ is the effect of species j on species i. Discretizing equation (1) gives:

$$\frac{\ln(X_{i,t+1}) - \ln(X_{i,t})}{\Delta t} = \mu_i + \sum_{j=1}^{L} \beta_{i,j} X_{j,t} \tag{6.2}$$

The model parameters are then determined by a system of linear equations. Once the parameters are fitted, a directed network can be obtained: $\beta_{i,j}$ represents the link from species j to i, where a positive (negative) $\beta_{i,j}$ means a positive (negative) effect where the increase of species j at time t would lead to an increase (a drop) of species i at time $t+1$. If $\beta_{i,j}$ is 0, it means species i is independent of species j, and hence no link from species j to species i.

In this work, we interpret a positive $\beta_{i,j}$ as microbial species *i* feeding microbial species *j*, and a negative $\beta_{i,j}$ as microbial species *i* inhibiting microbial species *j*. This is to maintain consistency with how the alternative model (introduced below) is used for network inference, which is explained in Section 6.2.1.2.2.

6.2.1.2. Chemical-mediated interaction (CMI) model

6.2.1.2.1. The model equations

In a real microbial community ecosystem, most microbial species (M) interact with each other through chemical species, such as feeding, inhibition and facilitation. Reflecting this chemically mediated nature, here we propose a new model which retains linearity (as of gLV) but which describes interactions between microbial species (M-M) through the combination of (1) the effect of microbial species on chemical species (M-C) and (2) that of chemical species on microbial species (C-M). For simplicity, we refer to the microbe-chemical interactions in both directions as M-C. Consider a system of equations:

$$\frac{dX_i}{dt} = X_i\left(\mu_i + \sum_j K_{i,j}S_j\right) - DX_i \tag{6.3}$$

$$\frac{dS_j}{dt} = \sum_i (Y_{i,j}r_i) + D(S_{j,in} - S_j) \tag{6.4}$$

where X_i denotes microbial species *i*, μ_i is the specific growth rate of species *i*, S_j denotes chemical species *j*, *D* is the dilution rate (per unit time) (ignoring microbial

species decay), K_j is a constant to represent the positive or negative impact of chemical species j on the growth of microbial species *i*. $Y_{i,j}$ is the yield coefficient for the production (when positive) or consumption (when negative) of chemical species *j* by the growth of microbial species *i* at a rate represented by $r_i = X_i(\mu_i + \sum_j K_{i,j}S_j)$. $S_{j,in}$ is the inlet substrate concentration for chemical species *j*.

Discretizing equation (6.3) gives:

$$\ln(X_{i,t+1}) - \ln(X_{i,t}) = \left(\mu_i - D + \sum_j K_{i,j}S_{j,t}\right)\Delta t \qquad (6.5)$$

For equation (6.4), consider re-arranging equation (6.3):

$$\frac{dX_i}{dt} + DX_i = X_i\left(\mu_i + \sum_j K_{i,j}S_j\right) = r_i \qquad (6.6)$$

Substituting r_i in equation (6.4) with equation (6.6) gives

$$\frac{dS_j}{dt} + DS_{j,t} - DS_{j,in} = \sum_i \left(Y_{i,j}\left(\frac{dX_i}{dt} + DX_i\right)\right) \qquad (6.7)$$

Discretizing:

$$\frac{S_{j,t+1} - S_{j,t}}{\Delta t} + DS_{j,t} - DS_{j,in} = \sum_i Y_{i,j}\left(\frac{X_{i,t+1} - X_{i,t}}{\Delta t} + DX_i\right) \qquad (6.8)$$

Finally,

$$S_{j,t+1} - \tau S_{j,t} - (1 - \tau S_{j,in}) = \sum_i Y_{i,j}(X_{i,t+1} - \tau X_{i,t}) \qquad (6.9)$$

where $\tau = 1 - D\Delta t$ and is dimensionless.

In CMI, equations (6.5) and (6.9) are linear whose parameters are to be fitted from data. In equation (6.5), $\ln(X_{i,t+1}) - \ln(X_{i,t})$ is the dependent variable in the linear model (commonly denoted as Y), $S_{j,t}$ is independent variable (X), and $K_j\Delta t$ and $(\mu - D)\Delta t$ are the parameter (β), respectively. In equation (6.9), $S_{j,t+1} - \tau S_{j,t} - \left(1 - \tau S_{j,in}\right)$ becomes the dependent variable (Y), $\left(X_{i,t+1} - \tau X_{i,t}\right)$ is independent variable (X), and $Y_{i,j}$ is the parameter (β).

It is worth to highlight that, equation (6.4) is the primary equation to determine a food web. Equation (6.3) is used as a supplementary equation, and only negative $K_{i,j}$ is considered when constructing the network, interpreted as inhibition from chemical species *j* to microbial species *i*.

6.2.1.2.2. Microbial species interactions (M-M) from interactions between microbial and chemical species (M-C)

Among possible types of M-M interactions, equation (6.9) captures the food chain (i.e. the relationship of feeding) through the production and consumption of chemical species by upstream and downstream microbial species (through the yield coefficient $Y_{i,j}$). A second type of M-M interactions considered by this work is inhibition of microbial species *i* and on microbial species *k* mediated by chemical species *j*, through combining a negative value of $K_{k,j}$ in equation (6.5) and a positive value of $Y_{i,j}$ in equation (6.9). The derivation of these two types of M-M interactions is illustrated in

Figure 6.1. In principle, a positive M-M interaction would also be manifested in equation (6.5) between two microbes when they are upstream and downstream microbes in a food chain (reflected by equation (6.9)), but this redundancy is ignored and a positive M-M interaction is determined in this work solely from the food chain, since the latter alone is sufficient for this purpose. This means that equation (6.5) is used only for inferring negative M-M interactions which, as shown later, originate from chemical inhibitions in the numerical simulation for generating data sets. Therefore, negative M-M interactions, identified by equation (6.5), are interpreted as inhibition. In principle, a negative M-M interaction could be of a nature of competition instead of inhibition, which however is not considered here because competitions between species that can co-survive in a microbial community have not been common in the cases evaluated in this work. On the other hand, a positive M-M interaction may arise from other circumstances than feeding, e.g. when one microbial species removes a chemical species that inhibits another microbial species. Such cases are not considered here, because the positive M-M interactions arising from such cases are often week compared with that between a cross-feeding pair.

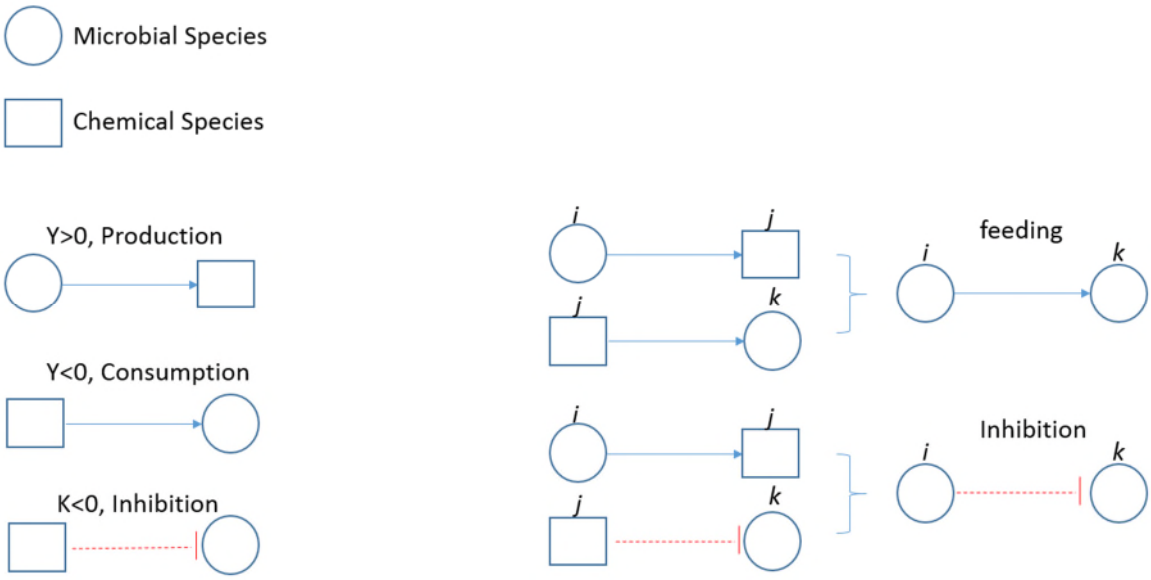


Figure 6. 1. Illustration of deriving M-M interactions from M-C interactions.

With the CMI model, we need to be aware that the proposed approach to infer M-M interactions from M-C interactions can potentially suffer from error accumulation: errors in identified M-C interactions will impact on the accuracy of the further inferred M-M interactions, a point to be revisited later in this chapter.

6.2.2. Sparse Solution and Lasso regression

When applying linear regression on gLV and CMI models, a sparse solution is assumed to avoid overfitting, that is, most of the parameters are 0 leaving only a small number of significant ones with a non-zero value to denote M-M (in the case of gLV) or M-C (in the case of CMI) interactions in a microbial network. To achieve this purpose, a Lasso regularization [29] is used:

$$\hat{\beta} = argmin\left(\frac{1}{2}\|Y - X\beta\|_F^2 + \lambda\|\beta\|_1\right) \quad (6.10)$$

where $Y \in R^{n\times m}$, $X \in R^{n\times p}$ and $\beta \in R^{p\times m}$. n is the number of data points, m is the dimension of Y (dependent variable) and p is the dimension of X (independent variable). $\|A\|_F^2 = \sum_i \sum_j A_{i,j}^2$, the Frobenius norm and $\|A\|_1 = \sum_i \sum_j |A_{i,j}|$, the l_1 norm. λ is the regularization factor, the magnitude of which determines the degree of penalisation on the lack of sparsity. Typically, λ is selected by using a cross-validation dataset, however, we applied the ROC curve to measure the performance of the proposed models, which does not involve the selection of λ. The above linear regression problem was solved by using MATLAB 2019b fitlinear solver.

6.2.3. Data generation from numerical simulation

In this work, all data for testing network inference methods are generated from numerical simulation which uses mechanistic models (ODE based) aimed to represent the dynamics of real microbial communities. Two different cases have been considered, as detailed below.

6.2.3.1. Anaerobic digestion: inference of a food web

Anaerobic digestion (AD) is a collection of processes by which microbial species break down biodegradable materials in the absence of oxygen. The overall process is used for converting waste materials to biogas, and is selected in this work as an example of a widely studied, real microbial community. In particular, we have adapted the ADM1 model developed by Rosen and Jeppsson [28]. The original model considers hydrolysis, acidogenesis, acetogenesis, and methanogenesis. For simplicity, the hydrolysis step is

left out in our simulation which directly sets sugar, amino acid and fatty acid as inlet substrates. The overall food web is shown in figure 6.2. Furthermore, several types of inhibition are modelled in ADM1, which are not excluded from this study. This is because, from our preliminary tests, the presence of strong inhibitions (particularly those from pH) appears to completely dominate the dynamics of the system, which makes it impossible to infer other types of interactions. In the simulation runs, the concentrations of inlet substrates have been kept low in order not to cause strong inhibition. All model equations and parameters are adopted from [28].

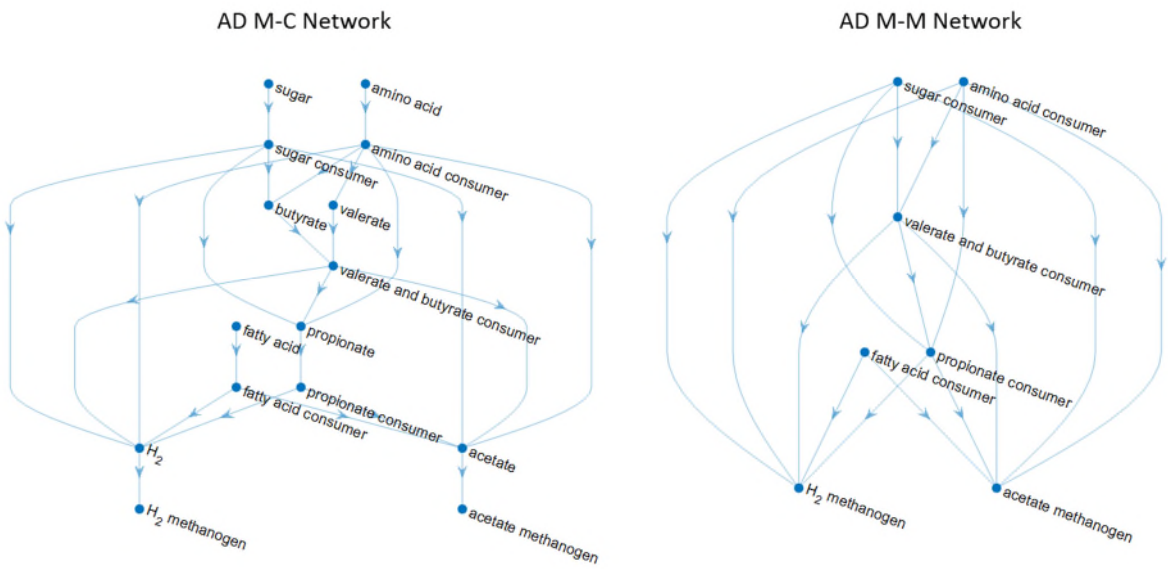


Figure 6. 2. The AD networks (food web only) adapted from ADM1.

6.2.3.2. Artificial microbial communities

Complementing the AD food web, in the second case we use a generic mechanistic model that implements typically assumed microbial growth kinetics (i.e. Monod) to

simulate a range of artificial microbial networks involving both feeding and inhibition and with varying degrees of complexity.

6.2.3.2.1. The mathematical model

Artificially generated microbial communities follow equations (6.11) and (6.12):

$$\frac{dX_i}{dt} = D\left(X_{i,in} - X_i\right) + r_i \tag{6.11}$$

$$\frac{dS_j}{dt} = D\left(S_{j,in} - S_j\right) + \sum_j Y_{i,j}\, r_i \tag{6.12}$$

where $r_i = \mu_{i,max} X_i \prod_j f(S_j) \prod_k I(S_k)$, $f(S_j) = \frac{S_j}{S_j + K_{i,j}}$, $I(S_k) = \frac{KI_{i,k}}{S_k + KI_{i,k}}$. X_i is the concentration of microbial species *i*, $X_{i,in}$ is inlet concentration of microbial species *i*. Similarly $S_{j,in}$ and S_j are chemical species *j* inlet concentration and concentration in the system, respectively. $Y_{i,j}$ represents the yield factor, similar to that introduced earlier in equation (6.4). *D* is the dilution rate, $K_{i,j}$ is the half saturation constant for the support of chemical species *j* to microbial species *i*, and $KI_{i,k}$ is the inhibition factor of chemical species *k* on microbial species *i*. $\mu_{i,max}$ is maximum growth rate of microbial species *i*.

This set of ODE equations are frequently used to simulate the dynamics of microbial communities. The default model parameter values adopted are listed in Table 6.2.

$Y_{i,j}$	Uniform (-0.5, 0.5)
S_{in}	0.5
K	Uniform (0.25, 0.5)
KI	0.01 (high level), 0.1 (low level)
μ_{max}	Uniform (0.5, 1)
D	$0.1 \times \min(\mu_{i,max})$

Table 6. 2. Parameter values for the artificial microbial community model.

6.2.3.2.2. Generated networks

In a microbial community, a key feature is the maximum length of the food chains within the community. In our simulation, this feature is reflected by the number of layers over which the microbes are (randomly) distributed. Between the neighbouring layers of microbes, the feeding relationships occur via a number of chemical species, with the microbes at the top layer receiving externally supplied substrates. Between a layer with N microbial species and a neighbouring downstream layer with M chemical species, there can be maximum N x M feeding links; our simulation randomly selects a subset of these links; the size of the subset is according to a sparseness factor that can be adjusted to provide a sparser or a denser network. Inhibition is generated in a similar way from the maximum N x M links from chemical species to microbial species, controlled by another sparseness factor. The number of microbial species, the number

of chemical species, the number of layers, and the sparseness factors together govern the complexity of each resulting food web. In this work, we have evaluated 4 types of networks at different levels of complexity, which are shown in Table 6.3. Note that real microbial communities may contain hundreds or even thousands of microbial species. When applying network inference on such communities, often only the microbial species with large abundance are considered; in previous studies such as [1] and [15] the number of abundant species were in the range of 30-50, which formed the basis for the chosen size of the "large" networks studied in this work.

No.	Number of microbial species	Number of layers	Number of inlet substrates	Sparsity	Name
1	20	2	5	Sparse	2-layer network
2	20	3	3	Sparse	3-layer network
3	35-45	4	5	Sparse	large sparse network
4	35-45	4	5	Dense	large dense network

Table 6. 3. Properties of artificially generated networks.

For each type of network listed in Table 6.3, 5 (random) network instances have been generated, network inference is carried out for every instance, and the collective results from the five instances have been used to produce robust conclusions for the type of network represented by the five instances. Throughout this work, we use "1-4" to represent the type of networks, and "a-e" to represent the specific network instances. All the used network instances are listed in the appendix. Illustrative examples of these types of network are shown in Figure 6.3. It is worth to mention that the number of

chemical species in each type of networks is not strictly constrained to allow feasible simulation of stable communities, which however does not vary significantly between different network instances of the same type.

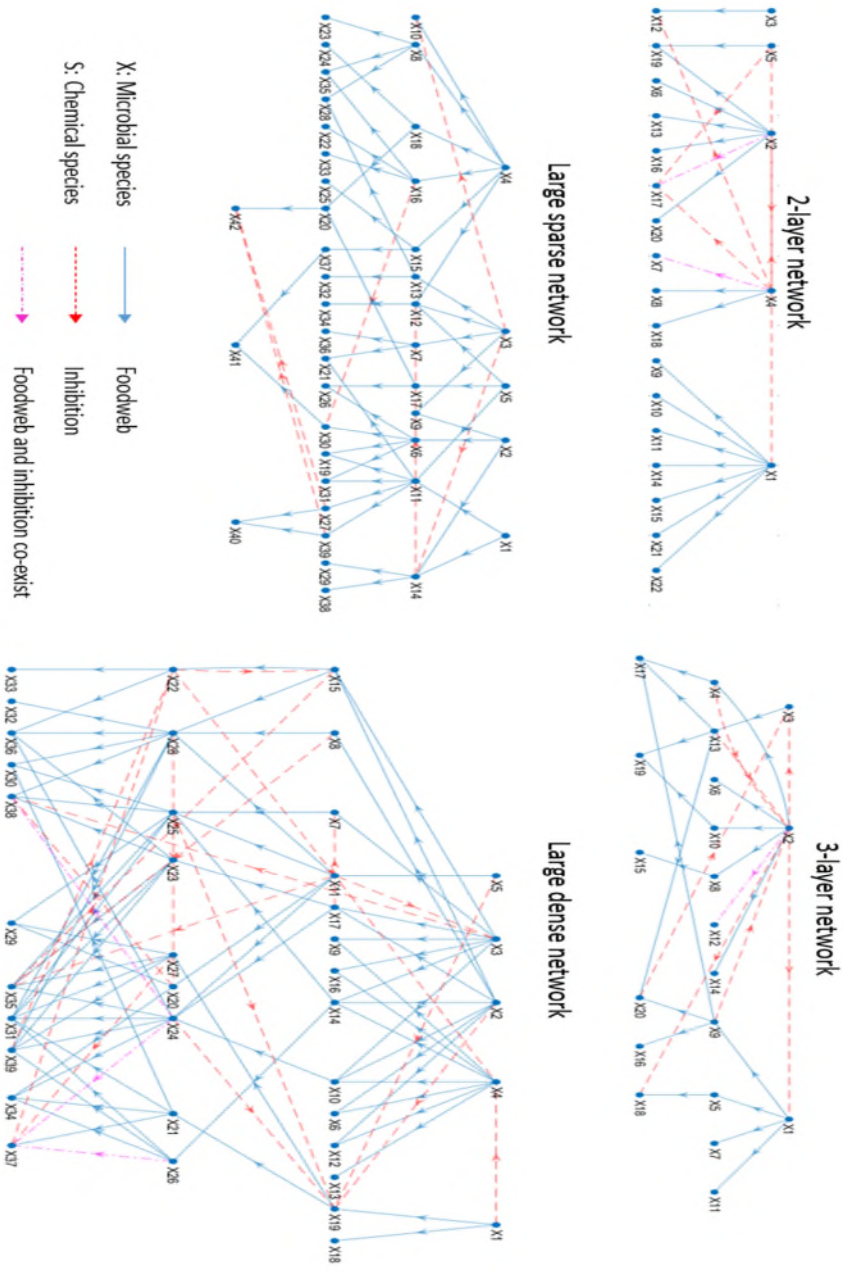


Figure 6. 3. Examples of the generated networks.

6.2.3.3. Data generation: perturbation strategy and measurement noise

In this work, we collect data from transient responses of the system stimulated with pulse perturbation to the steady state values of the differential variables or to the nominal values of the input variables by varying the concentration of the chosen chemical species, as shown in Figure 6.4. The data collection frequency is set to 1 day^{-1} and 2 day^{-1} for the AD system and the artificial communities, respectively, according to their different dynamical characteristics. The perturbation can be supplied on a rolling basis (with one chosen substrate perturbed per time) or a simultaneous basis (with all chosen substrates perturbed at the same time). Pulse perturbation has been used in past studies on a simultaneous basis [32], which in this work is to be compared with the newly introduced rolling perturbation strategy. At each numerical experiment, the perturbation strength in terms of the increase to the concentration of chemical species *i* is set by:

$$\Delta S_i = |\varepsilon|, \quad \varepsilon \sim N(0, \sigma_i) \tag{6.13}$$

where σ_i is the standard deviation of the zero-mean normal distribution adopted for generating the perturbation. For inlet substrates, the standard deviations are set according to the following:

$$\sigma_i = S_{in,i}\sigma \tag{6.14}$$

to make the perturbation proportional to the nominal inlet concentrations. All other chemical species tend to have much lower steady-state concentrations of similar magnitudes, therefore an identical standard derivation σ is applied to their perturbation. In the simulation runs, two levels of σ have been adopted (listed in the corresponding case studies), to analyse the impact of perturbation strength on network inference.

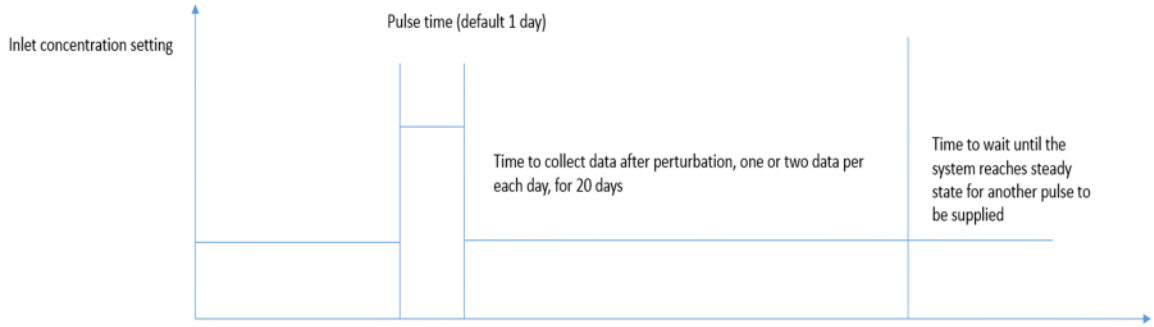


Figure 6. 4. Illustration of perturbation strategy

6.2.4. Evaluation of the inference results

Network inference results are measured by ROC (Receiver Operating Characteristic) curves. An ROC curve [29], illustrated in Figure 6.5, is a graphical plot that quantifies the performance of a binary classifier, which in this work is the network inference model. The curve comprises points each of which represents the result at a particular discrimination threshold, which in the modelling approach adopted in this work (see Section 6.2.2) is the regularization factor λ in equation 6.10. A binary classifier evaluated by an ROC curve is in general meant to make predictions about "positive" and "negative" relationships. In the context of this work, a "positive" relationship refers to the (predicted or actual) existence of an M-M or M-C interaction, while a "negative"

relationship refers to the non-existence of an M-M or M-C interaction. Furthermore, a correctly (wrongly) predicted relationship is termed true (false). The counts of true and false positive interactions in the network inference result (obtained at a particular value of λ, which itself is not shown on the plot) form a particular point on the ROC curve. A network inference model that produces "useful" results will have its ROC curve located above the diagonal line (showing that it tends to make more correct predictions than mistakes), and is regarded as more superior as the distance between the ROC and the diagonal line is larger.

		True Interaction	
		Positive (P)	Negative (N)
Predicted Interaction	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

$$TPR\ (true\ positive\ rate) = \frac{TP}{P} = \frac{TP}{TP + FN}$$

$$FPR\ (false\ positive\ rate) = \frac{FP}{N} = \frac{FP}{FP + TN}$$

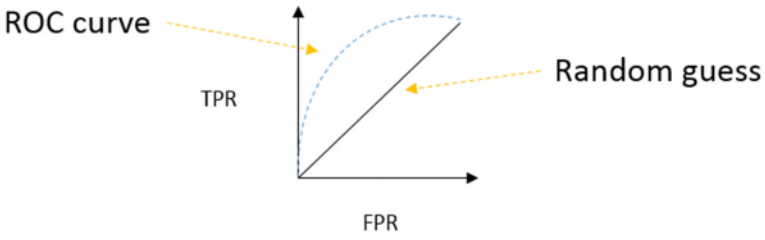


Figure 6. 5. An illustrative ROC curve.

6.3. Results and discussion

6.3.1. The AD System

In the AD system case, we carried out comprehensive analysis on network inference accuracy using CMI and gLV with different noise levels, perturbation levels, and perturbation strategies. Relevant parameters and explanations are listed in Table 6.4. Overall, the gLV model demonstrated stable performance in inferring the M-M structure over a wide range of conditions. In contrast, the CMI model appeared to be particularly sensitive to the perturbation strategy adopted, with rolling (as opposed to simultaneous) perturbation being essential for achieving useful network inference. Generally, both models were shown to benefit from data with a higher perturbation level and a lower noise level, as expected. Please note, 1 epoch contains data generated from one set of perturbation. With rolling perturbation, this includes perturbations separately applied to all the intended individual substrates (e.g. sugar, amino acid, etc.). For simultaneous perturbation, 1 epoch contains data collected following the simultaneous perturbation of all selected substrates. The number of epochs adopted in each case is set such that the total number of simulated dynamics (triggered by the perturbation to either one substrate or multiple substrates), as an indication of the amount of data used for network inference, remains comparable between different cases. The purpose for setting different number of epochs for different perturbation strategies under the rolling basis is to keep the total number of perturbations same (or very similar)

for all perturbation strategies. A summary of perturbation strategies and noise levels for AD case is shown in table 6.4.

		Standard deviation for perturbation to inlet concentrations (kg COD m^{-3})			
		sugar	amino acid	fatty acid	valerate, butyrate, propionate and acetate
Perturbation level	low level:	0.263	1	0.238	0.2
	high level	1.575	6	1.425	1
Classification of substrates	Sugar, amino acid and fatty acid are inlet substrates. Perturbations on 5 substrates include inlet substrates and valerate and butyrate. Perturbations on 7 substrates involve all substrates apart from soluble hydrogen.				
Perturbation strategy	1 day for pulsing, 20 days for collecting data after the perturbation with 1 data point per day				
Noise level			**No. epochs for different perturbation strategies under rolling basis**		
High level:	1%		Only inlet (3) substrates		50
Low level:	5%		5 substrates		30
			7 substrates		21
			No. epochs for simultaneous perturbation		150

Table 6. 4. Perturbation strategy and noise level for the AD case.

6.3.1.1 Performance of the gLV model

When gLV was used with data from rolling perturbation, perturbation on different substrates does not change network inference accuracy significantly. There are 4 factors

contributing to the impact of perturbations on different substrates, which are discussed in section 6.3.1.3. In the current case (with rolling perturbation), the attempt of introducing perturbation on intermediates did not provide sufficient additional information which can compensate for the reduced total number of perturbations on each individual substrate. The performance of the gLV model is shown in figure 6.6

When perturbation is generated simultaneously, the gLV model still provided robust network inference accuracy. Similar to the case of the rolling strategy, noise level and perturbation level plays very important role. Different perturbed substrates (inlet, intermediate or all substrates) did not lead to significant difference in terms of network inference. However, it is interesting that with the low noise level, an increase of perturbation level leads to a decrease in network inference accuracy. One possible reason is that such high level perturbation can lead to strong multi-collinearity, which is known to jeopardise linear regression [29]. Although the same problem could arise also with rolling perturbation, it is more severe in this case because all species are stimulated at the same time.

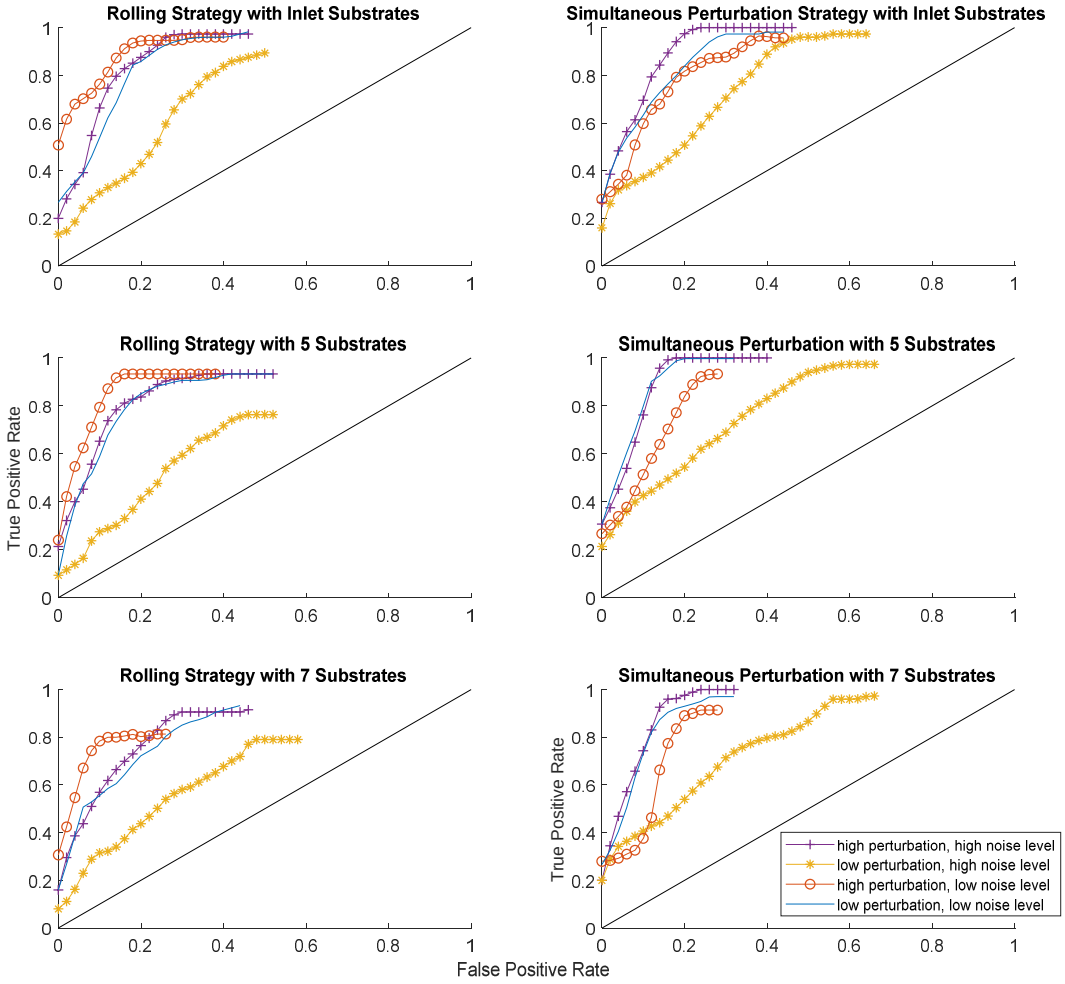


Figure 6. 6. ROC curves for network inference using the gLV model. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance between different runs is not significant, thus the error bar is not shown.

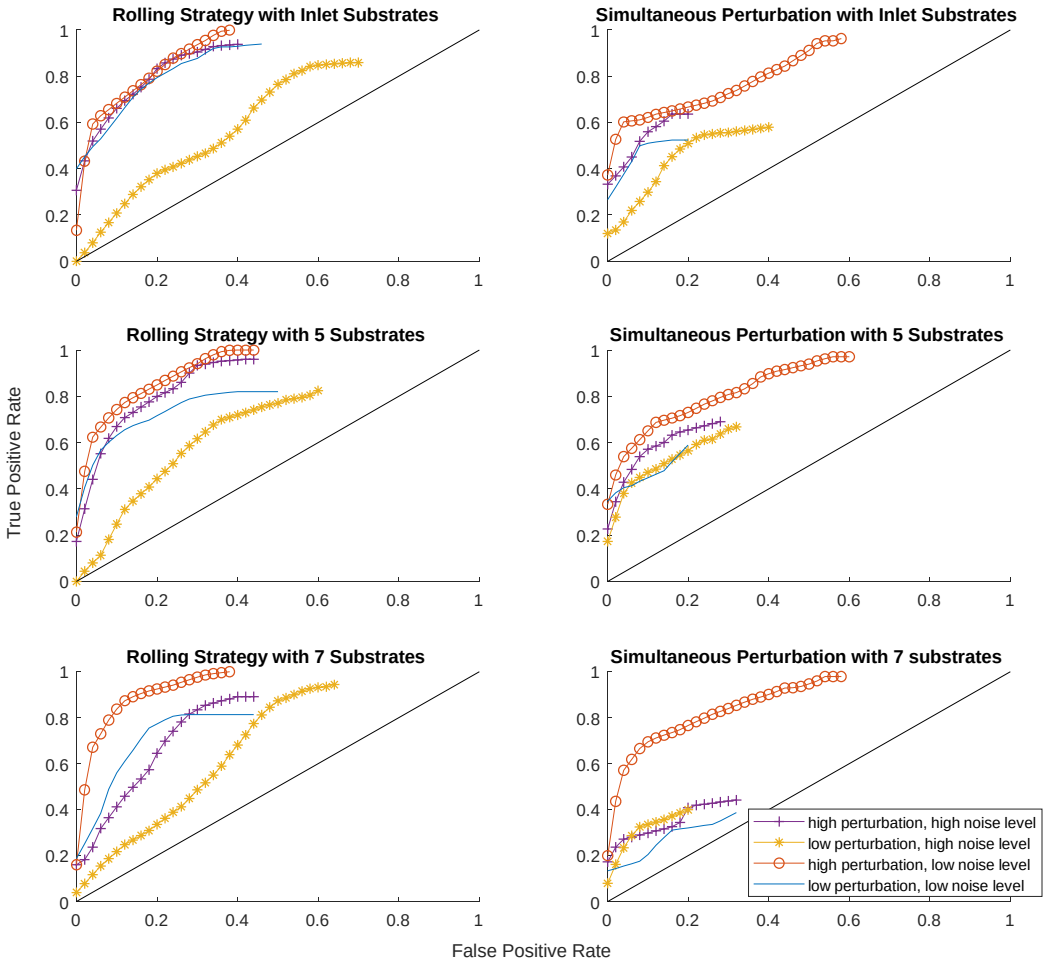


Figure 6. 7. ROC curves for M-M network inference using the CMI model. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance of different runs is not significant, thus the error bar is not shown.

6.3.1.2 Performance of CMI

With the CMI model and using data generated with rolling perturbation, network inference with high perturbation under low level of noise gave rather high accuracy when perturbation was on all chemical species (rolling perturbation with 7 substrates).

However, in the other cases, perturbation on all substrates gave slightly worse network inference quality. This is likely because, when perturbations were introduced to more substrates (5 substrates or all substrates), the number of perturbations on individual selected substrates decreased, hence if perturbations on intermediates were not able to provide additional information as compensation, the inference accuracy would decrease. For instance, if the total number of epochs is 30 and the perturbations are introduced to 3 substrates, then each of the substrate would be perturbed 10 times. However if the perturbations are introduced to 5 substrates, each of the substrate is only perturbed 6 times. Therefore, there is a trade-off between the number of substrates to introduce perturbations on and the number of perturbations per substrate, given a fixed total number of epochs. Figures 6.7 and 6.8 shows respectively the performance for M-M network inference and the M-C network inference of the CMI model.

It is worth to note that inference of production links is worse than consumption links. The reason very likely lies in the following: since the AD network starts with the consumption of inlet substrates by the first-layer microbial species (mainly acidogens), the production links tend to be at “deeper” layer than the consumption links, as illustrated in Figure 6.9, and the associated dynamic responses thus tend to be more indirect than those associated with the consumption links. This indirectness may hinder the inference. It is worth to notice that, in AD system, a wrong consumption interaction and a wrong production interaction can lead to a correct feeding process, as long as the

upstream and downstream species are identified. However, we highlight that this scenario is special only at the AD system.

With the simultaneous perturbation strategy, the CMI model failed to yield good results under most conditions. Only with high perturbation under the low noise level, it was able to give reasonable network inference. The most important reason lies in the poor ability to infer production, which is due to the fact that all chemical and microbial species were “excited” at the same time. When the perturbation is less significant and/or noise level is high, the data would contain information not sufficiently differentiating as to which microbial species producing which chemical.

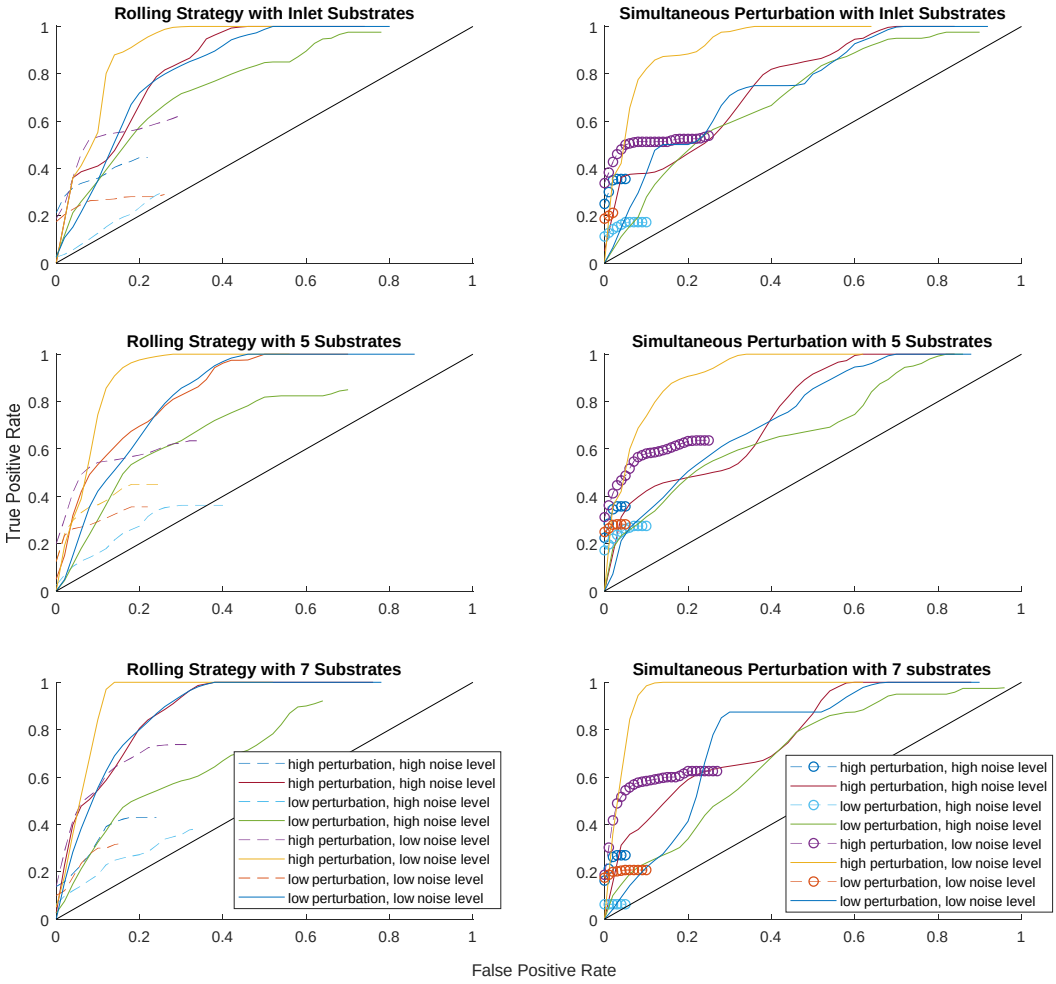


Figure 6. 8. ROC curves for CMI model M-C network inference. Each result is averaged by 5 runs with different perturbations generated from same perturbation level distribution. Note the variance of different runs is not significant, thus the error bar is not shown. In the left-hand-side figures, dash line represents production and solid line represents consumption; in the right-hand-side figures, dash line with circle represents production and solid line represents consumption.

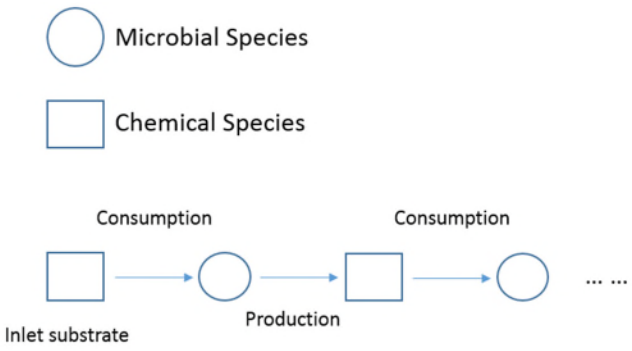


Figure 6. 9. How consumption inference is better than production.

6.3.1.3 Impact of the selection of chemicals for perturbation

The impact of selecting the chemicals for perturbation (inlet only, inlet with intermediate, and all chemicals) on the network inference accuracy appears to be complex. It is worth to note the following:

(1) With more chemicals perturbed under a rolling basis while the total number of epochs remains the same, the number of epochs from the perturbation of each single chemical decreases, which potentially "dilute" the information content and possibly introduces disadvantages for network inference.

(2) Perturbation on intermediates can provide more insight information that is not able to be captured by only perturbing inlet substrates.

(3) For the gLV model, perturbation on some chemicals at the downstream of the food web (i.e. around the last layer of microbial species, such as acetate-consuming methanogen and H_2-consumingmethanogen) cannot provide much useful information. This is because few or even none microbial species are stimulated. When there is

stimulation to microbial species, it is not through M-M interactions, and the information generated as such would “confuse” gLV model.

(4) A sudden increases in intermediate chemicals does not necessarily lead to better network inference. In the AD system, an exceeding and sudden increase in volatile fatty acid can inhibit many species (including the acid consumer), leading to rather supressed response and hence data with limited variation which is not ideal for network inference.

The above aspects often interact with each other and collectively determine the impact of choosing perturbed chemicals on the network inference accuracy.

6.3.1.4. Key results summary

Table 6.5 offers a compilation of the observed trends arising from all the tests carried out with the AD case as to how the performance of network inference by gLV and CMI is affected by various factors.

			N:high, P:high	**N:low, P:high**	**N:high, P:low**	**N:low, P:low**
AD system	Rolling basis perturbation	Inlet substrate	gLV:B, CMI:B	gLV:A, CMI:B	gLV:D, CMI:D	gLV:B, CMI:B
		Inlet and intermediate	gLV:B, CMI:B	gLV:A, CMI:B+	gLV:D, CMI:D	gLV:B, CMI:B
		All substrate	gLV:B, CMI:B	gLV:A, CMI:A+	gLV:D, CMI:D	gLV:B, CMI:B
	Simultaneous Perturbation	Inlet substrate	gLV:A, CMI:C	gLV:B, CMI:B	gLV:C, CMI:D	gLV:B, CMI:C
		Inlet and intermediate	gLV:A, CMI:C	gLV:B, CMI:B	gLV:D, CMI:C	gLV:A, CMI:C
		All substrate	gLV:A, CMI:D	gLV:B, CMI:B	gLV:D, CMI:D	gLV:A, CMI:D

Table 6. 5. Key results summary. Performance is marked from A to E, representing network inference accuracy from the best to the worst.

6.3.2. Artificially generated networks

The four types of networks introduced earlier in Table 6.6 have been studied. In the following, we first present the results of studying the 2-layer network and the 3-layer network; the parameters used for generating data are listed in Table 6.6. After that, results on the larger networks with two levels of sparseness are presented. For all the artificially generated networks, each network type has been studied with 5 (randomly generated) network instances to improve the robustness of the results. On perturbation, applying the rolling strategy (which in the study of the AD system was shown to be more advantageous in general compared to simultaneous perturbation) to the inlet substrates (which is more realistic for implementation, compared to perturbing more chemical species) was taken as the default option. In all studies, data collection was assumed to be 2 data points per day.

Inhibition level	high level	KI=0.01
	low level	KI=0.1
Perturbation level	high level	σ=100
	low level	σ=30
Noise level	high level	5%
	low level	1%
Perturbation epochs	2 layer network	30
	3 layer network	50

Table 6. 6. Parameters used in the 2-layer and the 3-layer networks.

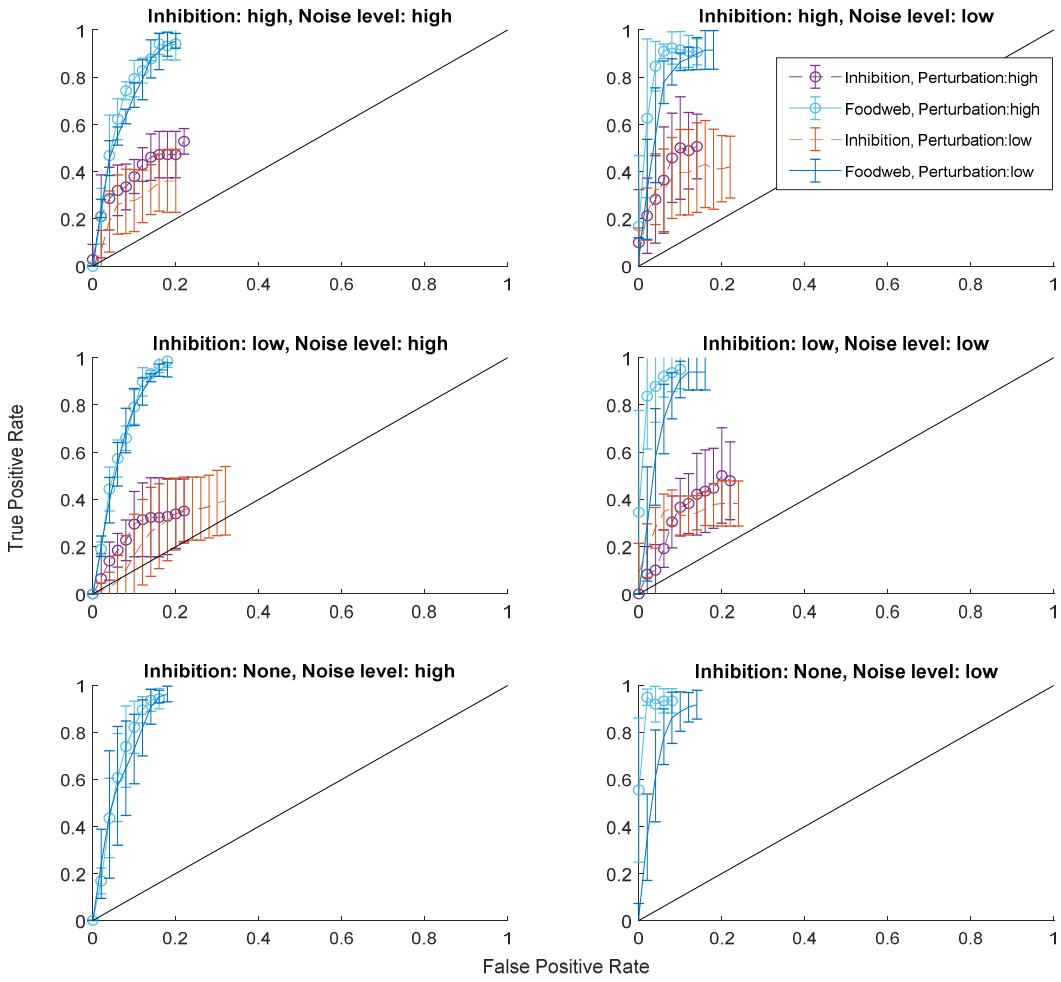


Figure 6. 10. ROC curve for inference of the 2-layer network by the gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent “mean +/- standard deviation”.

6.3.2.1. 2-layer network

6.3.2.1.1 Results of the gLV model

From Figure 6.10, one can see gLV performed well and stably for inferring the food web under different levels of inhibition, noise levels and perturbation levels. However, its ability of inferring inhibition was significantly worse. This contrast is most likely

because the dynamics of the microbial community is more dominated by the food web than the inhibition links, which makes it difficult for the network inference to separate accurately the effect of inhibition from that of the food web. As one can see in Figure 6.11, from a regression point of view, when the regularization factor λ (introduced earlier in equation 6.10) decreases from a relatively high level, the frequency (i.e. count) of positive elements (i.e. regression parameters with a positive value, predicting a feeding connection between two microbes) increases rather rapidly once a threshold is passed. However, the frequency of the negative elements (predicting an inhibition link) starts to increase much later, only when λ (hence the residual, MSE) becomes much smaller. In Lasso regression, this can be interpreted as follows: when the residual is large, the dependent variable (of which the residual is calculated) has higher correlation with the independent variables with a positive parameter than those with a negative parameter. In other words, the feeding interactions (which form the food web) have a better ability than the inhibition links in explaining the dynamics of the microbial community.

From Figure 6.10, one can see that when the inhibition level increases, gLV has an improved ability to infer inhibition interactions. This suggests that in such cases more dynamics of the system is created by inhibition, which thus becomes easier for gLV to grasp. It is also worth to note that the ROC curves for inhibition tend to have longer error bars than those for the food web, especially when the perturbation level is low. There are two possible reasons to explain this difference. Firstly, there are much fewer

inhibition interactions than food-web links, thus the success or failure to infer any individual inhibition link can significantly increase or decrease the true positive rate, hence introducing a larger variance in the ROC plots. Secondly, as discussed above, inhibition (even at a higher level) appears to contribute less to the overall dynamics of the system, compared to the food-web, which could hinder not only the accuracy but also the precision (i.e. variance) of its inference.

In all cases, when the noise level decreases and/or the perturbation level decreases, the accuracy of inference tends to decrease for both the food web and inhibition, as expected.

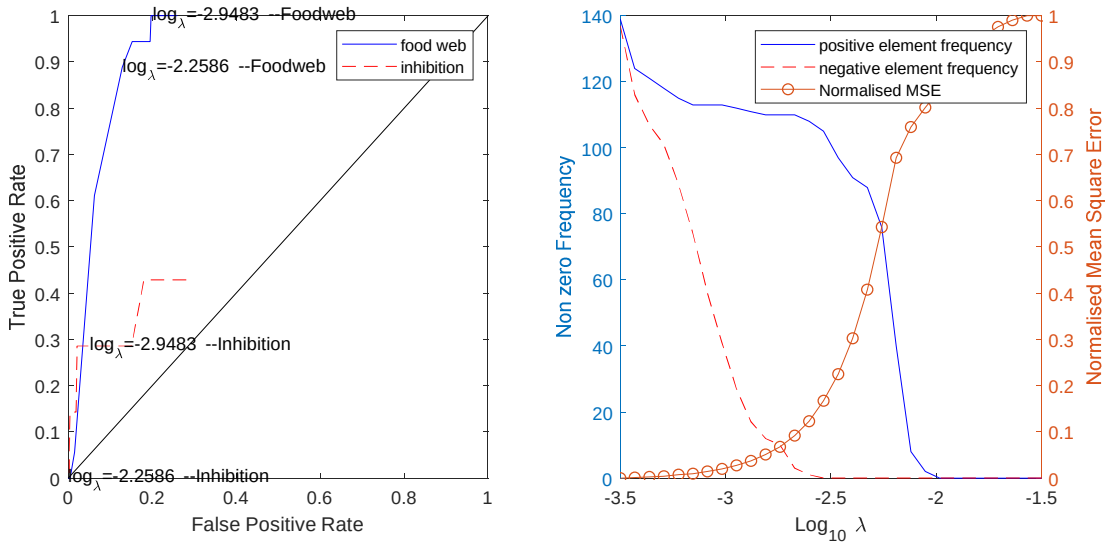


Figure 6. 11. Impact of λ on inferring inhibition by gLV (data: network instance 1a; Noise level = 0.05, Inhibition level = 0.01, perturbation level = 50). Left: ROC curve for food web and inhibition inference; Right: relations between λ and inferred frequency of positive and negative elements.

6.3.2.1.2 Results of the CMI model

The inference results using CMI are summarised in Figure 6. 12. Comparing with Figure 6.10 (for gLV), one can see the network inference ability by CMI is worse than gLV, under the same levels of inhibition, noise and perturbation. Only when the noise level is low and the perturbation level is high, CMI starts to show the ability to infer a network with a reasonable quality. However, it is worth to note that, when the inhibition level is high, CMI has a very good ability to infer inhibitions, which is significantly better than gLV does, However, at the same time, CMI's ability in inferring the food web decreases. This trade-off of CMI between its prediction accuracy of inhibition and that of the food web when the inhibition level varies can be more clearly observed at the low noise level (see the two top-right subplots in Figure 6.12). With the gLV model, the performance was shown earlier to be different: a decrease of the inhibition level can slightly increase the network inference accuracy on the food web; at a high inhibition level, it does not reduce as significantly as in the case of the CMI model. The CMI model infers M-M interactions from M-C links. As shown in Figure 6.13 and in comparison with Figure 6.12, M-C interactions are inferred with a better quality than M-M interactions. Also, within M-C interactions, CMI has a better ability in inferring consumption links than production and inhibition interactions, same as in the case of the AD system.

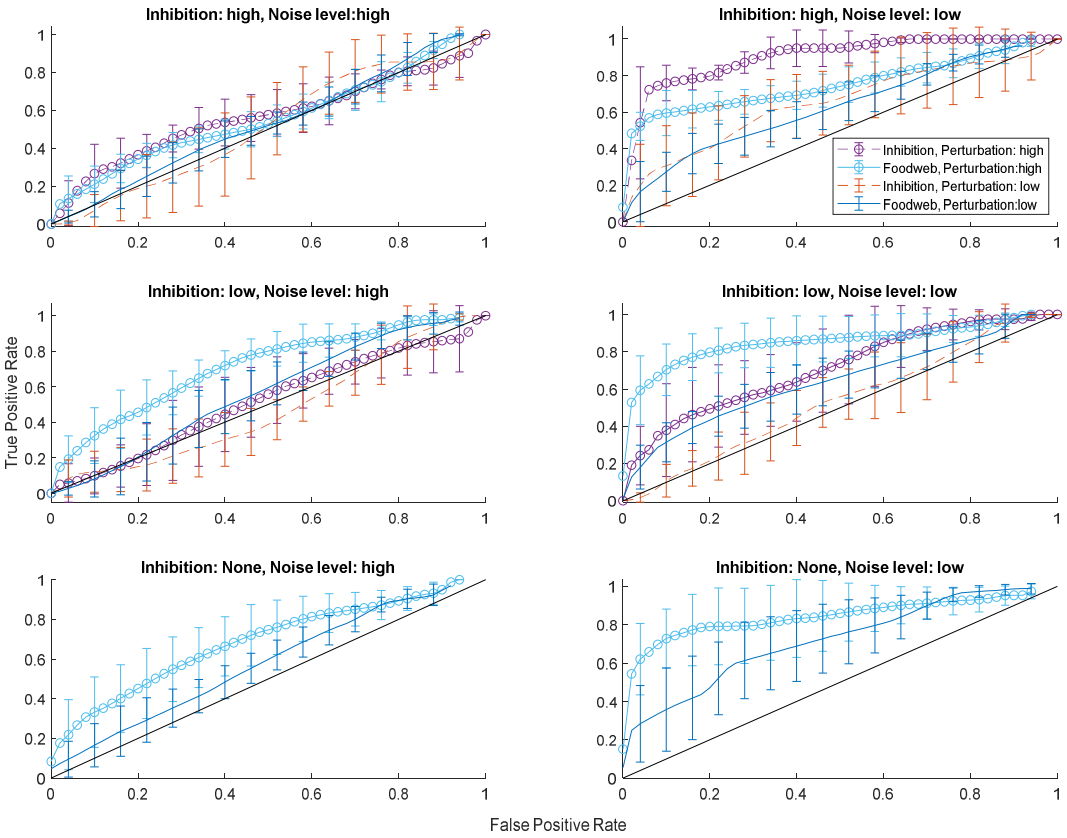


Figure 6. 12. ROC curve for inference of the 2-layer network by the CMI model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent “mean +/- standard deviation”.

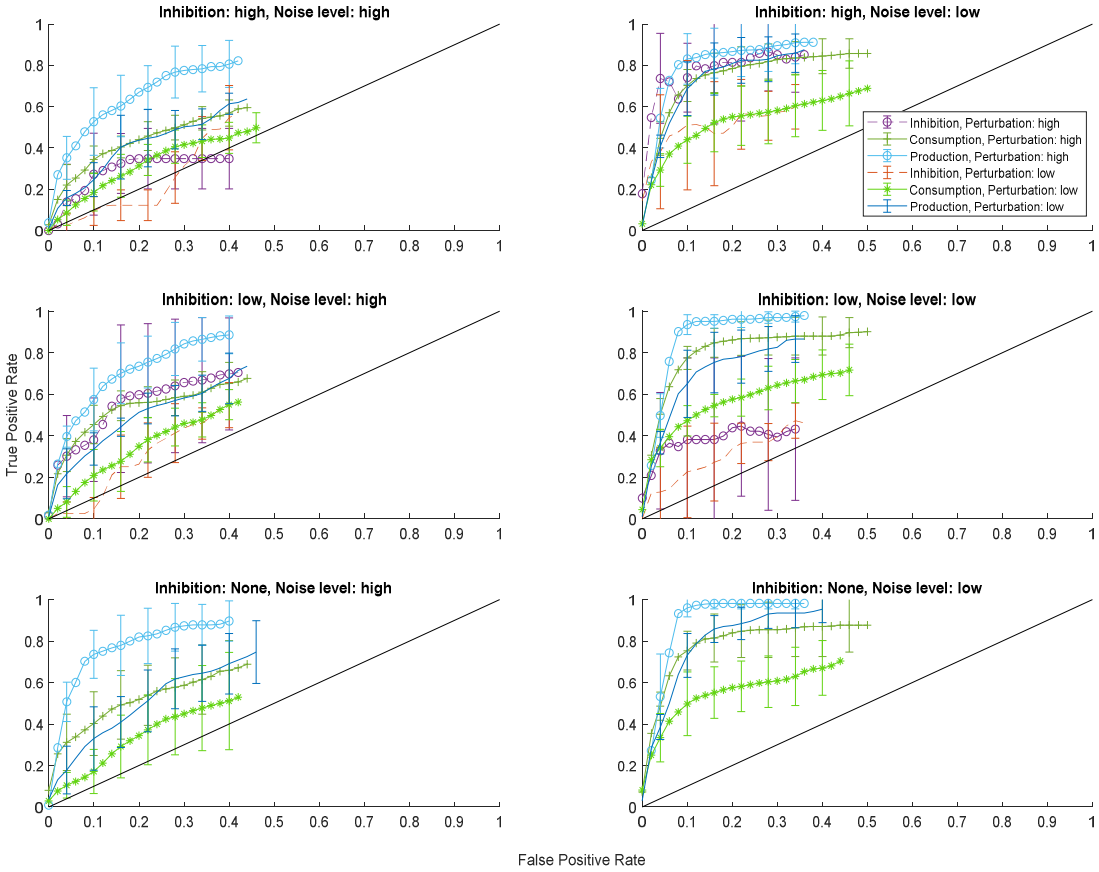


Figure 6. 13. ROC curve for inference of the 2-layer network M-C interactions by the CMI model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent “mean +/- standard deviation”.

As gLV outperforms CMI in most cases with the current perturbation strategy, we hypothese that a changed perturbation strategy is able to improve the performance of CMI. To verify it, we conducted an additional numerical experiment on network 1a, as an example. Rolling perturbation was introduced on both inlet and intermediate substrates (S_1 – S_{23}) to generate 8 epochs of data, and both gLV and CMI were used to test the network inference accuracy. Figure 6.15 shows the ROC curves for the performance of both models. Noise level and perturbation level were chosen to be 1%

and 5% respectively, under high level of inhibition. One can see that CMI achieved good network inference on both food web and inhibition at the same time. With gLV, although the inference on the food web still remained very good, it showed a much worse ability in inferring inhibition. Additionally, one can see CMI has the potential to correctly construct all feed web and inhibition links at the same time. As the CMI model captures the food web and other M-C interactions separately with two different equations (6.5 and 6.9), theoretically it is possible to infer multiple interactions (feeding and inhibition) between a microbial pair. With gLV model, the relationship between a microbial pair is characterised by a single binary parameter ($\beta_{i,j}$ in equation 6.1), which makes it impossible to capture multiple types of links between the pair.

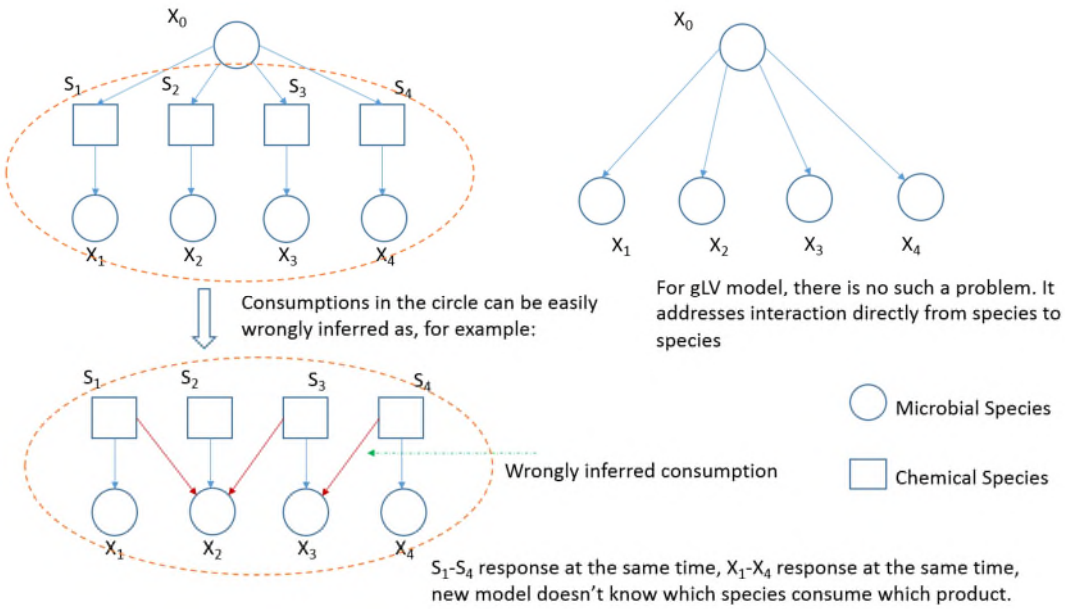


Figure 6. 14. Errors for the CMI model in consumption inference.

There are two other aspects of the results that are worth to mention. First of all, M-C interactions have shorter error bars than M-M inferred interactions, which is well-

expected since M-M interactions were derived by combining M-C interactions which would accumulate uncertainties. Secondly, even M-C consumption links were inferred (by CMI) with a quality slightly worse than that of the food web by gLV. The reason is explored in Figure 6.14. In this illustrative example, S_1-S_4 are stimulated at the same time (through a change in X_0), and X_1-X_4 start to be affected at the same time. When the noise level is high or when perturbation is not sufficient, it is hard to distinguish which species uptake which substrates. Please note, when the microbial community has more layers (e.g. 3 layers, explored in the next section), gLV also suffers the problem of lacking distinguishability (in terms of which species feeds which species).

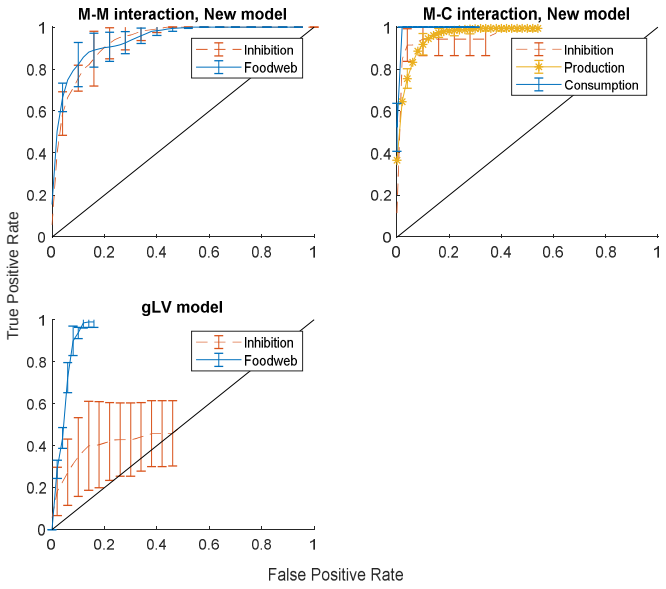


Figure 6. 15. Potential for CMI to infer multiple M-M interactions in a 2-layer network. Results are from testing network 1a and averaged with 5 runs with different actual perturbations generated from the same perturbation level. The error bars represent "mean +/- standard deviation".

6.3.2.2. 3-layer network

The number of microbial species in 3-layer networks was kept roughly the same as in 2-layer networks (~20 species per network). When the default perturbation strategy for artificial networks (i.e. rolling perturbation on inlet substrates) was applied, the CMI model failed to produce reasonable inference at all conditions, and the results are not shown here. The reason appears to lie in the weak stimulation to the deeper layers of the networks, which also affected the performance of gLV although the latter produced reasonable results. Here we present and analyse the performance of the gLV model on 3-layer networks. We also show how well CMI could perform in inferring multiple M-M interactions, investigated earlier with 2-layer networks, this time with networks of a "deeper" structure.

6.3.2.2.1. Performance of gLV with rolling perturbation on inlet substrates

As shown in Figure 6.16, the gLV model still performed stably for inferring the food web in the case of 3-layer networks, although slightly worse than with 2-layer networks. However, the inference on inhibition was much worse when the networks had one more layer, although as with the 2-layer networks, a higher inhibition level improved the inference of inhibition links. The favourable impact of stronger perturbation and a lower noise level, shown earlier with the 2-layer networks, was also observed in this case.

Figure 6. 16. ROC curve for gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation".

6.3.2.2.2. Analysis of the inference performance when dealing with a "deeper" network

When there are more layers in a network, more network links between species (microbial species in gLV; microbial and chemical species in CMI) are inferred with data not containing the direct effect of perturbation, when perturbation is carried out only to inlet substrates. In order to investigate how this might affect the performance of network inference, we show in Figure 6.17 the networks generated by gLV with the same number of links as in the ground true network, obtained from selecting appropriate levels of regularization (λ). The true network, also shown in Figure 6.17, is network 2a. It was simulated with a high inhibition level, a high perturbation level, and a low noise level, which as shown earlier would give the best opportunity for accurate network inference, to allow us to focus on the impact of the existence of a "deeper" layer on the results. From Figure 6.17, one can see that gLV had a very good ability to infer the food web between the first and the second layer, when the original strategy (perturbing only inlet substrates) was applied. However, it performed worse in inferring interactions between the second layer and the third layer. Other types of mistakes were also made, e.g. inferred feeding links between the first layer and the third layer, between species

in the same layer, and in wrong (downstream to upstream) directions, but failures in predicting links between 2nd-3rd layers represented dominating inference errors.

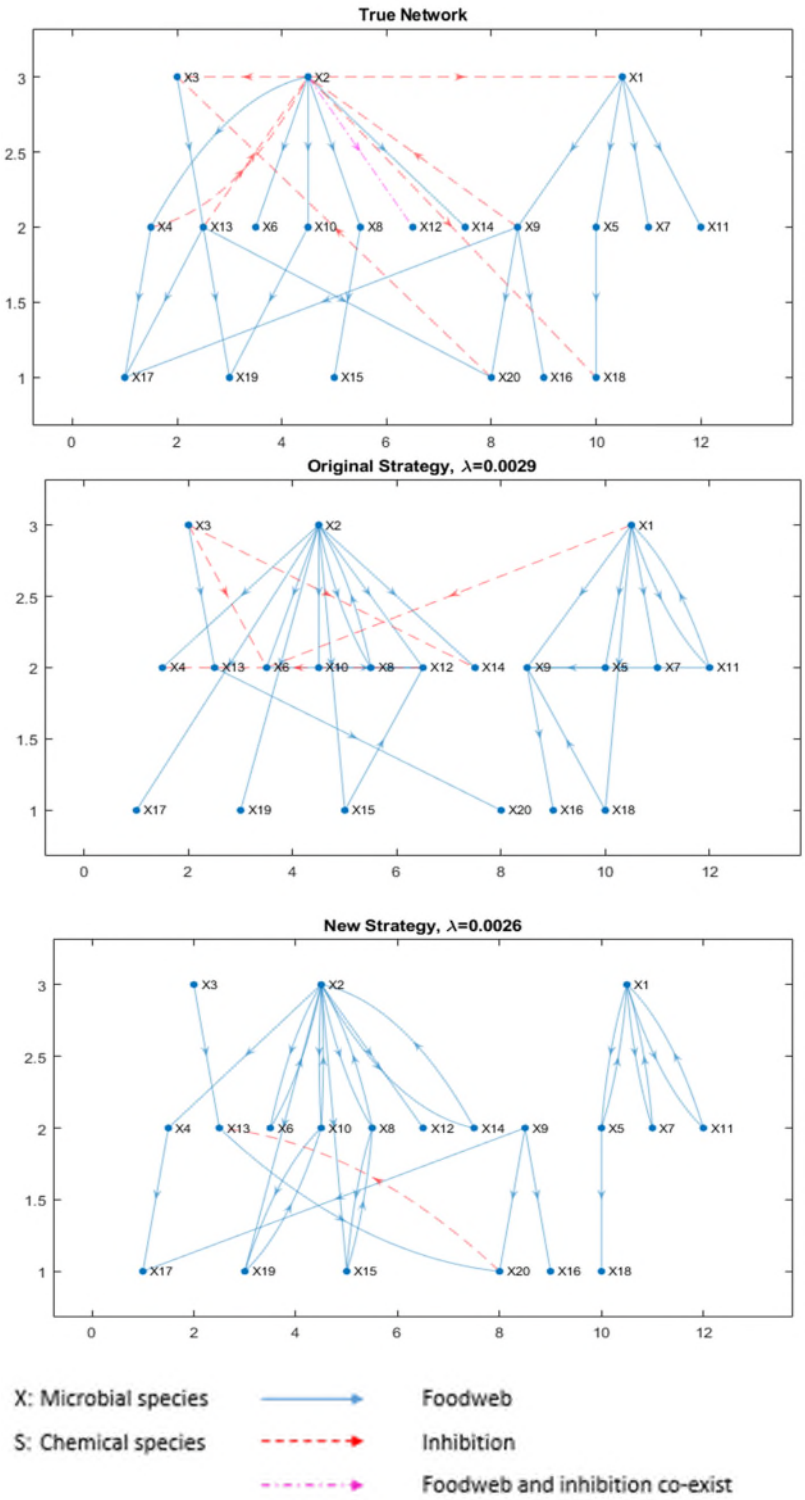

True Network
X3
X2
X1
X4
X13
X6
X10
X8
X12
X14
X9
X5
X7
X11
X17
X19
X15
X20
X16
X18
Original Strategy, λ=0.0029
New Strategy, λ=0.0026
X: Microbial species
S: Chemical species
Foodweb
Inhibition
Foodweb and inhibition co-exist

Figure 6. 17. Comparison between inferred and true 3-layer networks by gLV (with rolling perturbation on only inlet substrates and inlet and intermediate substrates). Network instance 2a was used as an example.

We hypothesized that, the poor inference of "deeper", downstream links by gLV was due to the lack of direct stimulation around these links when only the inlet substrates were perturbed, and that a change of the perturbation strategy could make an improvement. To verify this, we conducted a separate numerical experiment with perturbation on inlet and second-layer chemical species (S_1-S_{16}), while keeping the total number of perturbations (150) and other conditions (low noise level, high perturbation level and high inhibition level) identical to those used in producing results shown in Figure 6.16 original strategy. The new results are shown in Figure 6.17. With this new perturbation strategy, which introduced direct stimulation to the network links between the second and the third layer, the inference of these links was significantly improved. At the same time, the amount of wrongly inferred food web links between same layers (especially in layer 2) also reduced.

This experiment demonstrates the importance of different perturbation strategies, which ultimately affects the useful information content in the generated data. Figure 6.18 demonstrates the comparison between the two strategies by ROC curves. One can see that the new strategy significantly improved the performance of food web inference. However, the ability to infer inhibition decreased. As indicated in AD network inference (Section 6.3.1), a change of the perturbation strategy can have opposite

impacts on different aspects. In the current case, additional information from perturbation with the new strategy was apparently a major factor to increase the inference ability on the food web. For the network chosen in this exercise (network 2a), most inhibitions exist between species in layer 1 and layer 2. Given a fixed amount of total number of epochs, fewer perturbations are subjected to layer 1, which makes the inhibition inference slightly worse. Please note, this is special only for this network instance.

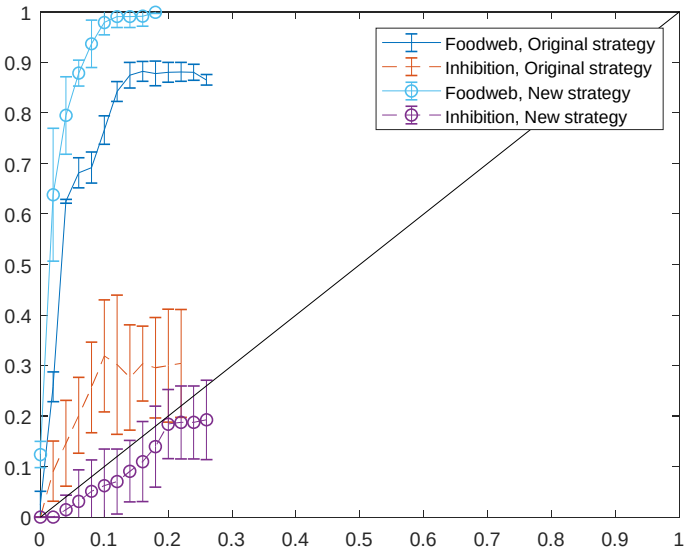


Figure 6. 18. Network inference performance with network 2a using gLV and different perturbation strategies. Each result is averaged with 5 runs with different actual perturbations generated from the same perturbation level distribution. The error bars represent “mean +/- standard deviation”.

6.3.2.2.3. Inference of multiple M-M interactions

With an appropriate perturbation strategy, the CMI model was still able to infer inhibition and feeding connections at the same time in 3 layer networks with decent accuracy, as shown in Figure 6.19 which presents the results on network 2e. As with two-layer networks (Section 6.3.2.1.2), this part of the investigation still adopted rolling perturbation on all inlet and intermediate substrates (S_1-S_{30}), with a low noise level and a high inhibition level, generating 8 epochs of data. The perturbation level applied however was 3 times that for the same investigation on the 2-layer networks, which was found necessary to get satisfactory results. In figure 6.19, CMI is shown to be able to obtain good inference on both food web and inhibition. Similar to what was shown in section 6.3.2.1.2, CMI has demonstrated the potential to infer multiple M-M interactions. In network 2e, there are 50% inhibitions overlapping with the food web (in terms of M-M pairs), and CMI was able to identify all food web and inhibition links (TPR=1 for both food web and inhibition). However, one needs to recognise that with networks having more layers and greater complexity, the level of perturbation required for satisfactory inference may become unrealistically high.

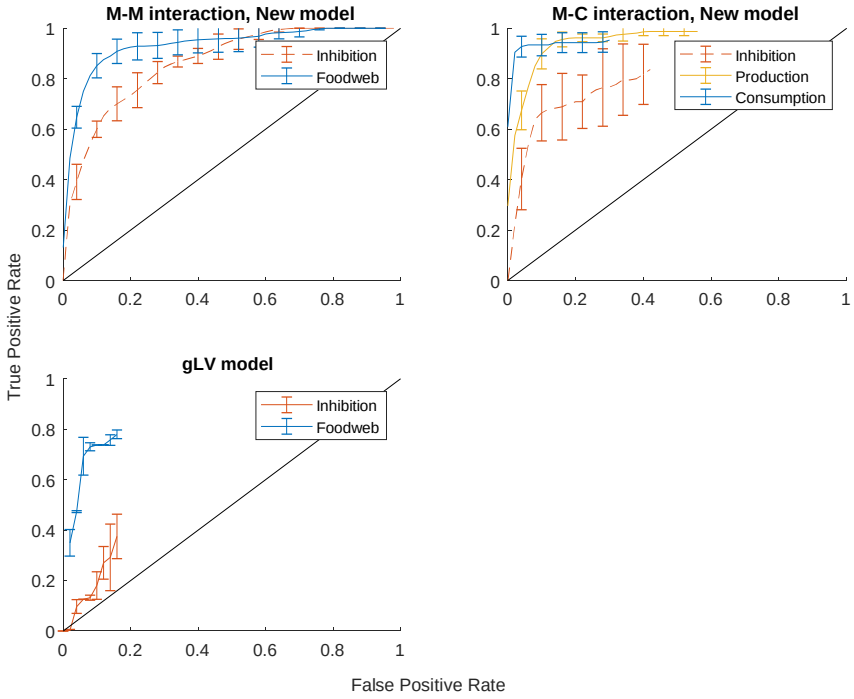


Figure 6. 19. Potential for CMI to infer multiple M-M interactions in a 3-layer network. Results are from testing network 2e and averaged with 5 runs with different actual perturbations generated from same perturbation level. The error bars represent “mean +/- standard deviation”.

6.3.2.3. Large scale networks and impact of sparsity

In this section, we applied gLV model to both sparse and dense large scale networks. CMI was not used because it showed increasing difficulties from previous sections in inferring larger and more complex networks. All networks tested here were with a high inhibition level. The reason is that when the inhibition level is low, the contribution of inhibition to the dynamics decreases, as shown from previous sections. Especially in large and complex networks, the overall inference accuracy is expected to decrease, inhibition at a low level would become more impossible to get successfully inferred. Secondly, for a given large network structure, a slight change to the strength of interaction (e.g. increase or decrease of inhibition level or yield) can lead to a change of the network structure, i.e. some microbial species become unable to survive (even with a decreased inhibition level) due to the complex interactions in the microbial community.

The results are shown in Figure 6.20. First of all, for large scale networks, network inference accuracy remains similar compared to smaller (in terms of the number of species) and shallowed (in terms of the number of layers) networks. The inference accuracy for dense networks is significantly worse than for sparse networks, suggesting that gLV struggled to capture the real structure when the number of interactions increased. As commonly observed in the tests reported in earlier sections, higher perturbation and lower noise remain favourable factors for achieving more accurate inference.

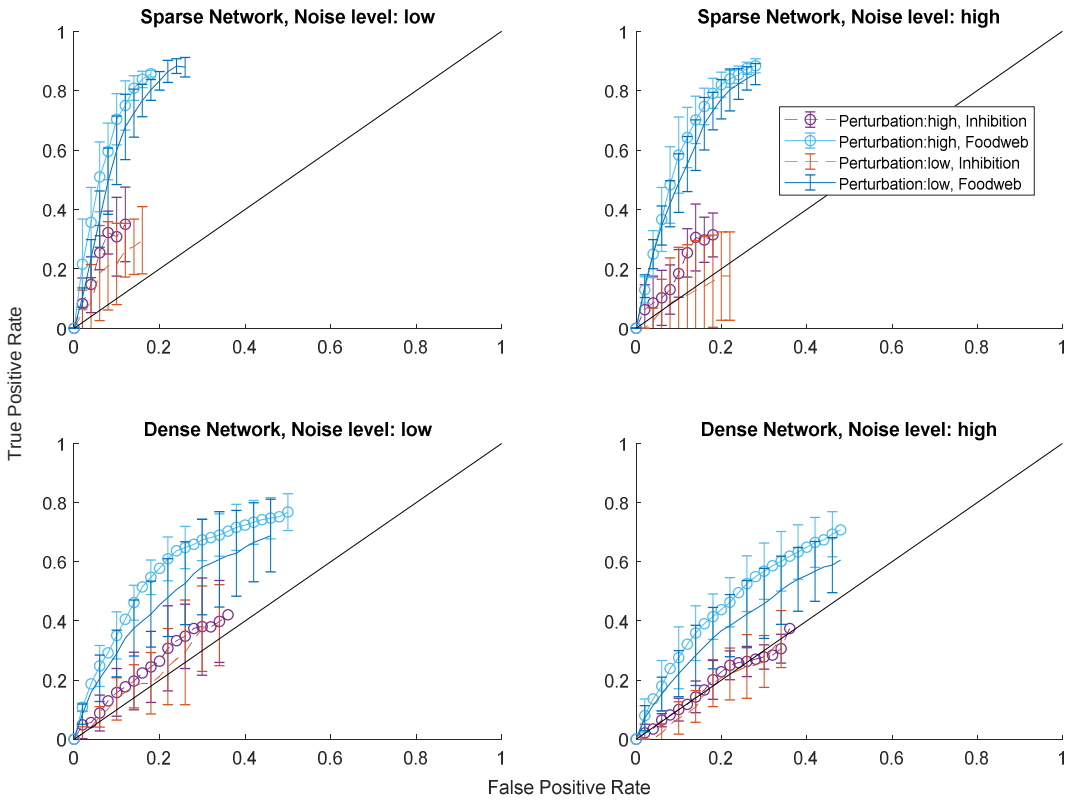
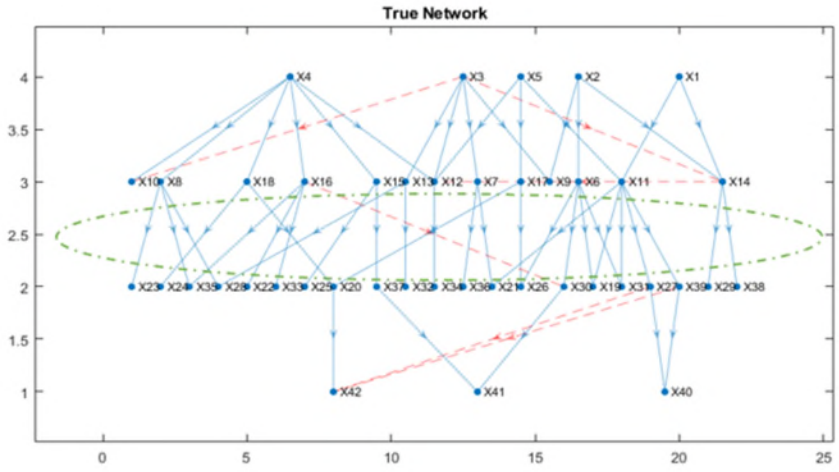
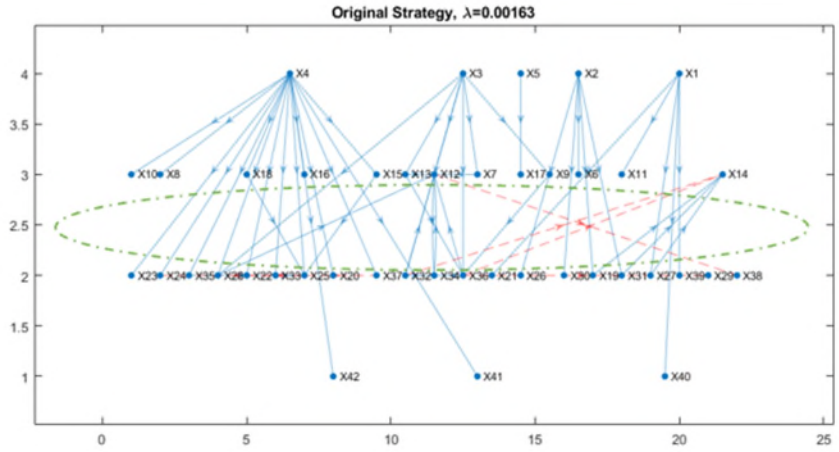


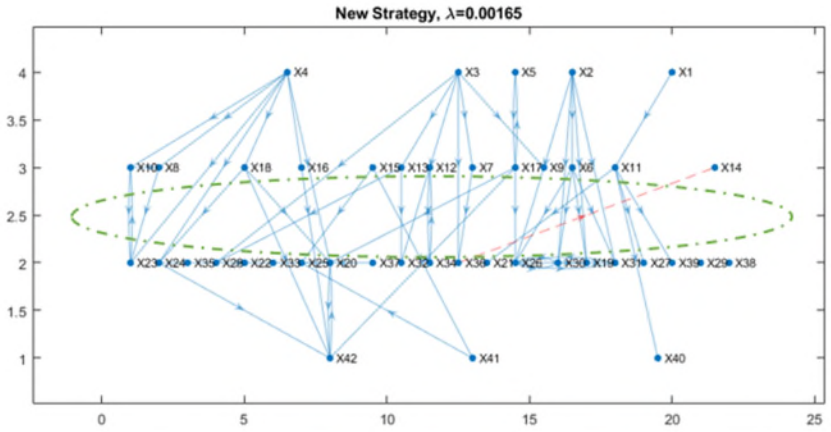
Figure 6. 20. Network inference of large-scale dense and sparse networks by the gLV model. Each curve is the average (mean) over the results from inferring 5 random network instances; the error bars represent "mean +/- standard deviation".

We also tested the impact of replacing the default ("original") perturbation strategy, i.e. perturbing inlet substrates only, with a more aggressive ("new") strategy with perturbation on inlet and first layer intermediate substrates (S_1-S_{24}), while keeping the overall number of perturbations (150) unchanged. The results are shown in Figure 21. We selected proper regularization (λ) so that the amount of inferred links is the same as the ground truth network. With the original strategy, one can see that two types of mistakes, (1) the first layer species were wrongly inferred to feed species on the third and the fourth layer, and (2) the inaccuracy of the inferred links between the second

layer and the third layer; similar mistakes were also observed earlier with the 3-layer networks (Section 6.3.2.2.2). With the new strategy, one can see from Figure 6.21 that both types of mistakes were reduced (shown in the green dash circle). Figure 6.22 shows the comparison between the two strategies by ROC curves. One can see that the new strategy significantly improved the performance of food web inference, with a slightly decreased ability to infer inhibition, similar to what was shown in section 6.3.2.2.2. Additional information from perturbation with the new strategy still becomes the major factor to increase the inference ability on the food web. As for inhibition, perturbation on intermediates (as in the new strategy) may decrease the likelihood to infer the two inhibitions linking from X_3 to X_{10} and X_{14}. However, unlike the case in section 6.3.2.2.2, there are only 2 inhibitions linking from species in the first layer, thus the degree of reduced ability on inferring inhibitions becomes less significant.







X: Microbial species — Foodweb

S: Chemical species — Inhibition

Foodweb and inhibition co-exist

Figure 6. 21. Comparison between inferred and true 3-layer networks by gLV (with rolling perturbation on only inlet substrates and inlet and intermediate substrates). Network instance 3e is used as the example

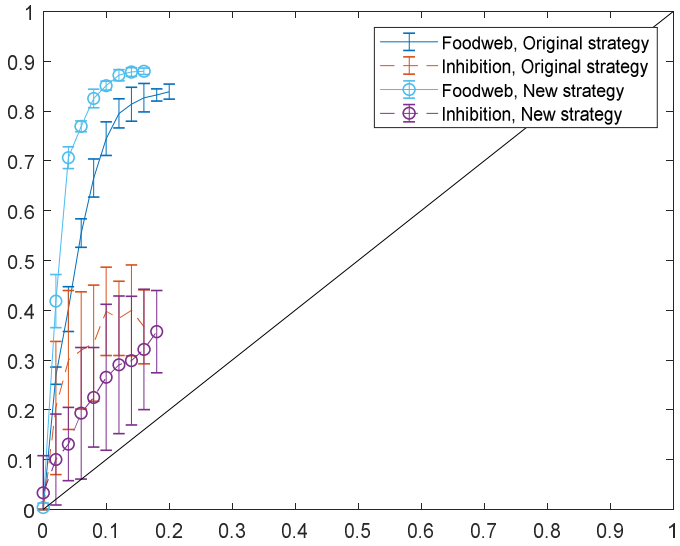


Figure 6. 22. Network inference performance with network 3e using gLV and different perturbation strategies. Results are from testing network 3e and averaged with 5 runs with different perturbations generated from same perturbation level. The error bars represent “mean +/- standard deviation”.

6.3.2.4 Summary of key results

Table 6.7 offers a compilation of the observed trends arising from all the tests carried out in this section as to how the performance of network inference by gLV and CMI is affected by various factors.

			N:high, P:high	N:low, P:high	N:high, P:low	N:low, P:low
2-layer network	gLV model	Inhibition high	F:A-, I:C	F:A+, I:B	F:B+, I:C-	F:A-, I:C
		Inhibition low	F:A-, I:C	F:A+, I:C+	F:B+, I:D	F:A-, I:C
		Inhibition None	F:A-, I:None	F:A+, I:None	F:B+, I:None	F:A-, I:None
	CMI model	Inhibition high	F:E, I:E	F:C, I:A	F:E, I:E	F:D, I:D
		Inhibition low	F:D, I:E	F:B, I:D	F:E, I:E	F:D, I:E
		Inhibition None	F:D, I:None	F:B, I:None	F:E, I:None	F:D, I:None
		Potentials to infer multiple interactions				
3-layer network	gLV model	Inhibition high	F:B, I:D	F:A I:D+	F:B-, I:D	F:B, I:D
		Inhibition low	F:B, I:E	F:A, I:E	F:B-, I:E	F:B, I:E
	CMI model	Potentials to infer multiple interactions				
Large Sparse Network		Sparse	F:B, I:D+	F:B+ I:B	F:B-, I:E	F:B, I:C
		Dense	F:D, I:E	F:D+, I:E	F:D, I:E	F:D, I:E

Table 6. 7. Key results summary. F: foodweb, I: inhibition. Performance is marked from

A to E representing network inference accuracy from the best to the worst.

6.4. Conclusion

In this work, we conducted a comprehensive analysis on network inference for microbial species networks mediated by chemical species. The widely adopted gLV model and the newly proposed chemical-mediated-interaction (CMI) model were assessed with the food web of anaerobic digestion and a range of artificially generated networks that included both food web and inhibition links with different levels of complexity. Across all the tested cases, network inference generally benefited from a higher quality of data, in terms of a low noise level and a greater degree of variation (resulting from strong perturbations to the system for data collection). Compared to the CMI model, the gLV model demonstrated a more stable inference ability from simple to complex network structures particularly for the food web links. This observed advantage of gLV is somehow at odd with the fact that the CMI model, which explicitly captures microbial-chemical interactions, can utilise information about not only the microbes but also the chemical species, while a microbial interaction model like gLV can only make use of the former. A key factor leading to the generally inferior performance of the CMI model was considered to be the accumulation of errors when constructing the microbial network from chemical-microbial interactions. On the other hand, the gLV model demonstrated an unsatisfactory performance for inferring inhibition links under all the tests, while the CMI model was able to produce good inference on both the food web and the inhibition links, when sufficient perturbation was applied to generate data. This points to the potential of future improvements over

modelling approaches capable of making the best use of the chemical information in inferring microbial networks.

References

[1]. Shaw, G. T. W., Liu, A. C., Weng, C. Y., Chou, C. Y., & Wang, D. (2017). Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. PLoS one, 12(7), e0181395.

[2]. Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., ... & Chan, A. T. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nature microbiology, 3(3), 356.

[3]. Streit, W. R., & Schmitz, R. A. (2004). Metagenomics–the key to the uncultured microbes. Current opinion in microbiology, 7(5), 492-498.

[4]. Faust, K., & Raes, J. (2012). Microbial interactions: from networks to models. Nature Reviews Microbiology, 10(8), 538.

[5]. Newman, M. (2010). Networks: an introduction. Oxford university press.

[6]. Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2011). Controllability of complex networks. Nature, 473(7346), 167.

[7]. Angulo, M. T., Moog, C. H., & Liu, Y. Y. (2019). A theoretical framework for controlling complex microbial communities. Nature Communications, 10.

[8]. Fiedler, B., Mochizuki, A., Kurosawa, G., & Saito, D. (2013). Dynamics and control at feedback vertex sets. I: Informative and determining nodes in regulatory networks. Journal of Dynamics and Differential Equations, 25(3), 563-604.

[9]. Mochizuki, A., Fiedler, B., Kurosawa, G., & Saito, D. (2013). Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. Journal of theoretical biology, 335, 130-146.

[10]. Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. PLoS computational biology, 8(9), e1002687.

[11]. Fang, H., Huang, C., Zhao, H., & Deng, M. (2015). CCLasso: correlation inference for compositional data through Lasso. Bioinformatics, 31(19), 3172-3180.

[12]. Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. PLoS computational biology, 11(5), e1004226.

[13]. Schwager, E., Mallick, H., Ventz, S., & Huttenhower, C. (2017). A Bayesian method for detecting pairwise associations in compositional data. PLoS computational biology, 13(11), e1005852.

[14]. Ridenhour, B. J., Brooker, S. L., Williams, J. E., Van Leuven, J. T., Miller, A. W., Dearing, M. D., & Remien, C. H. (2017). Modeling time-series data from microbial communities. The ISME journal, 11(11), 2526.

[15]. Gibbons, S. M., Kearney, S. M., Smillie, C. S., & Alm, E. J. (2017). Two dynamic regimes in the human gut microbiome. PLoS computational biology, 13(2), e1005364.

[16]. Chen, I., Kelkar, Y. D., Gu, Y., Zhou, J., Qiu, X., & Wu, H. (2017). High-dimensional linear state space models for dynamic microbial interaction networks. PloS one, 12(11), e0187822.

[17]. Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., ... & Xavier, J. B. (2013). Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS computational biology, 9(12), e1003388.

[18]. Fisher, C. K., & Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PloS one, 9(7), e102451.

[19]. Shaw, G. T. W., Pao, Y. Y., & Wang, D. (2016). MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. Bmc Bioinformatics, 17(1), 488.

[20]. Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., ... & Olle, B. (2016). MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome biology, 17(1), 121.

[21]. Kambam, P. K. R., Henson, M. A., & Sun, L. (2008). Design and mathematical modelling of a synthetic symbiotic ecosystem. IET systems biology, 2(1), 33-38.

[22]. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B 12 through a symbiotic relationship with bacteria. Nature, 438(7064), 90.

[23]. Wade, M. J., Pattinson, R. W., Parker, N. G., & Dolfing, J. (2016). Emergent behaviour in a chlorophenol-mineralising three-tiered microbial 'food web'. Journal of theoretical biology, 389, 171-186.

[24]. Momeni, B., Xie, L., & Shou, W. (2017). Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. Elife, 6, e25051.

[25]. Di, S., & Yang, A. (2019). Analysis of productivity and stability of synthetic microbial communities. Journal of the Royal Society Interface, 16(150), 20180859.

[26]. Ren B, Schwager E, Tickle TL, Huttenhower C. SparseDOSSA: Sparse data observations for simulating synthetic abundance. 2016.

[27]. Schaffter T, Marbach D, Floreano D (2011) GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics (Oxford, England) 27: 2263–2270.

[28]. Rosén, C., & Jeppsson, U. (2006). Aspects on ADM1 Implementation within the BSM2 Framework. Department of Industrial Electrical Engineering and Automation, Lund University, Lund, Sweden, 1-35.

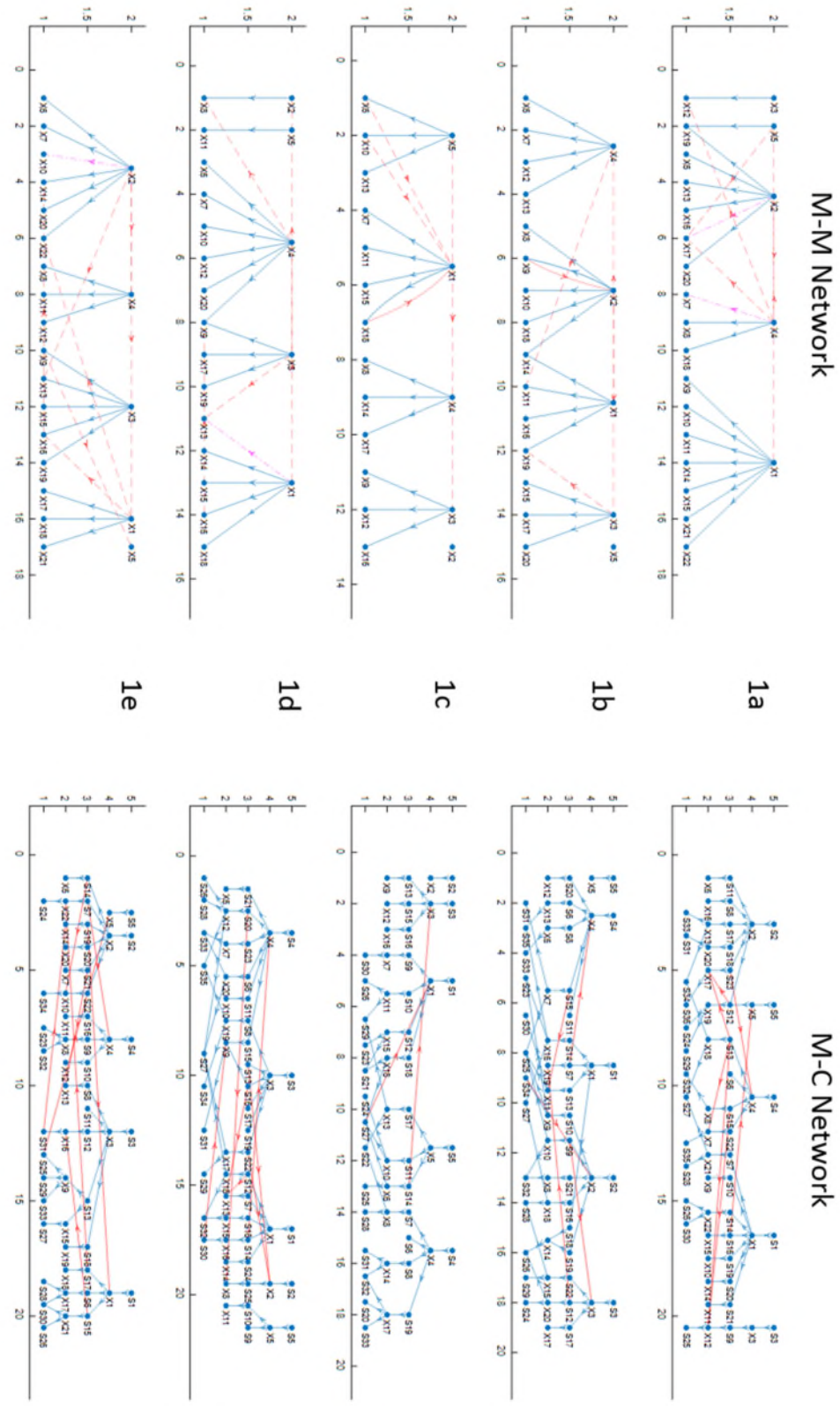
[29]. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

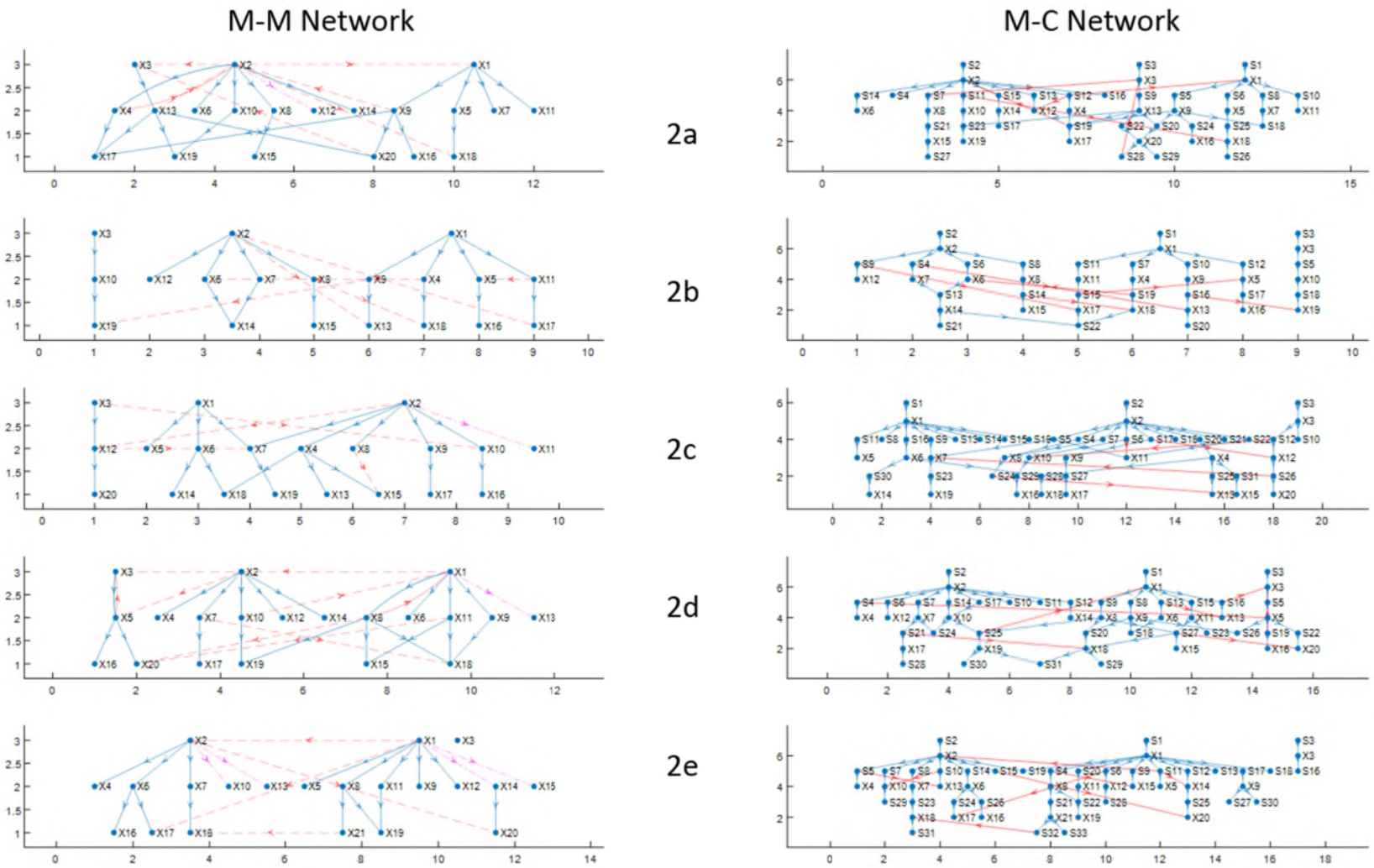
[30]. Angulo, M. T., Moreno, J. A., Lippner, G., Barabási, A. L., & Liu, Y. Y. (2017). Fundamental limitations of network reconstruction from temporal data. Journal of the Royal Society Interface, 14(127), 20160966.

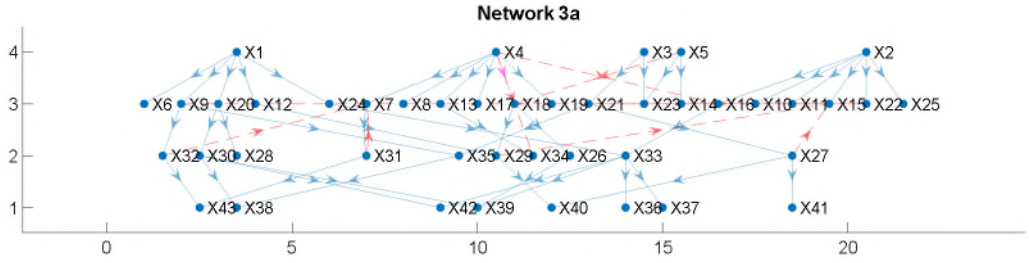
[31]. Lu, H., & Lee, P. K. (2015). Effects of cellulose concentrations on the syntrophic interactions between Clostridium cellulovorans 743B and Rhodopseudomonas palustris CGA009 in coculture fermentation for biohydrogen production. International Journal of Hydrogen Energy, 40(35), 11800-11808.

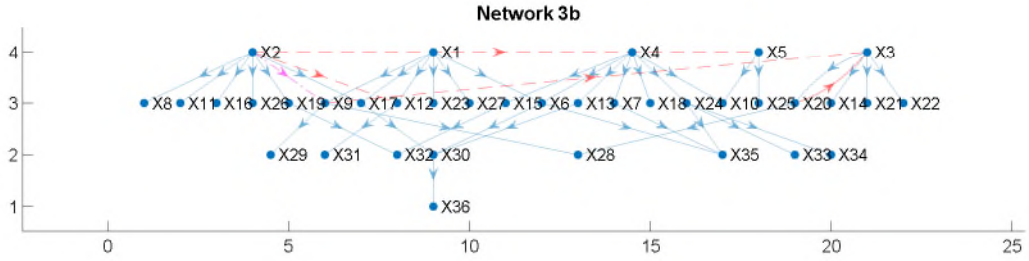
[32]. Cao, H. T., Gibson, T. E., Bashan, A., & Liu, Y. Y. (2017). Inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons. BioEssays, 39(2).

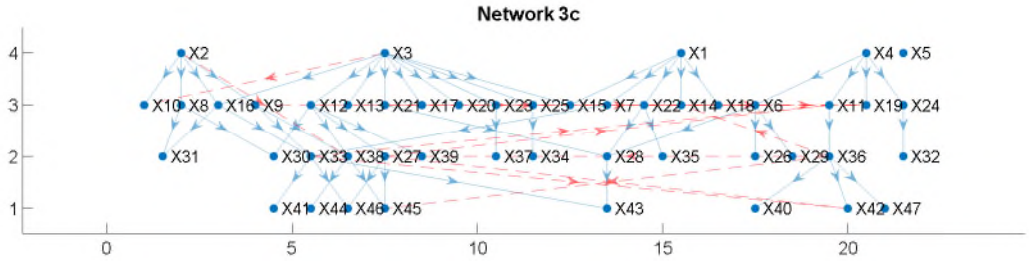
Appendix: Network instances studied

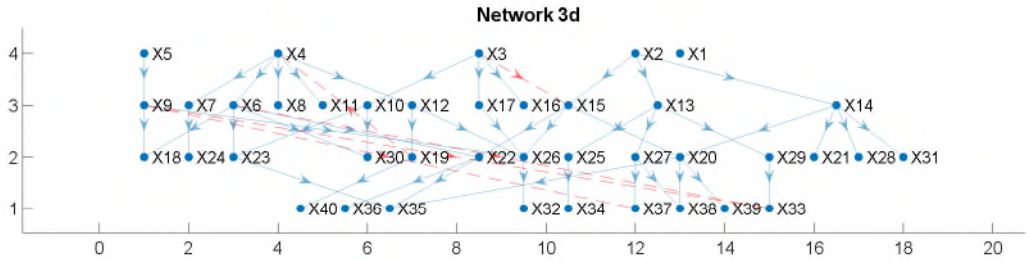


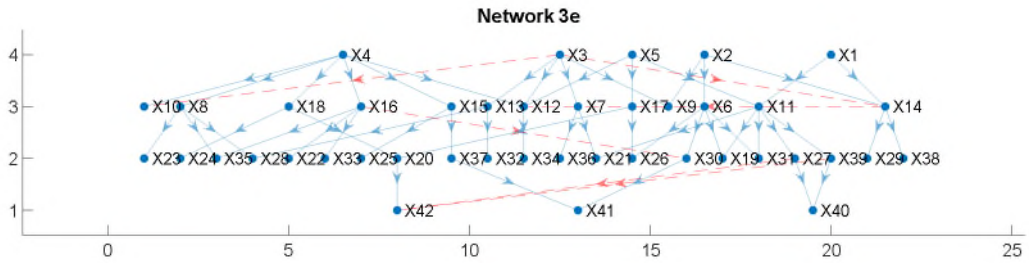

M-M Network
M-C Network
2a
2b
2c
2d
2e

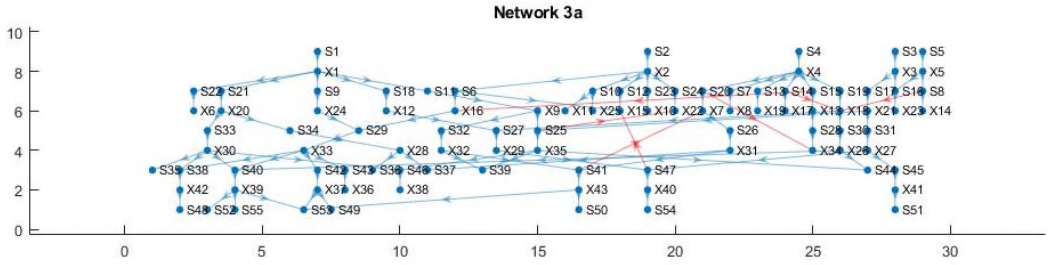

Network 3a
X1 X4 X3 X5 X2
X6 X9 X20 X12 X24 X7 X8 X13 X17 X18 X19 X21 X23 X14 X16 X10 X11 X15 X22 X25
X32 X30 X28 X31 X35 X29 X34 X26 X33 X27
X43 X38 X42 X39 X40 X36 X37 X41
0 5 10 15 20

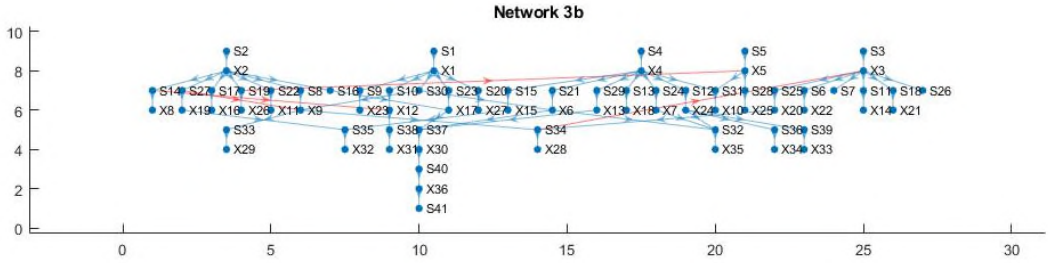

Network 3b
X2 X1 X4 X5 X3
X8 X11 X16 X26 X19 X9 X17 X12 X23 X27 X15 X6 X13 X7 X18 X24 X10 X25 X20 X14 X21 X22
X29 X31 X32 X30 X28 X35 X33 X34
X36
0 5 10 15 20 25

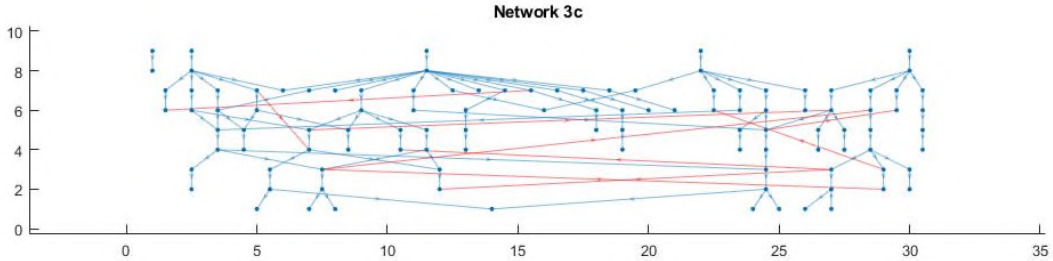

Network 3c
X2 X3 X1 X4 X5
X10 X8 X16 X9 X12 X13 X21 X17 X20 X23 X25 X15 X7 X22 X14 X18 X6 X11 X19 X24
X31 X30 X33 X38 X27 X39 X37 X34 X28 X35 X26 X29 X36 X32
X41 X44 X46 X45 X43 X40 X42 X47
0 5 10 15 20

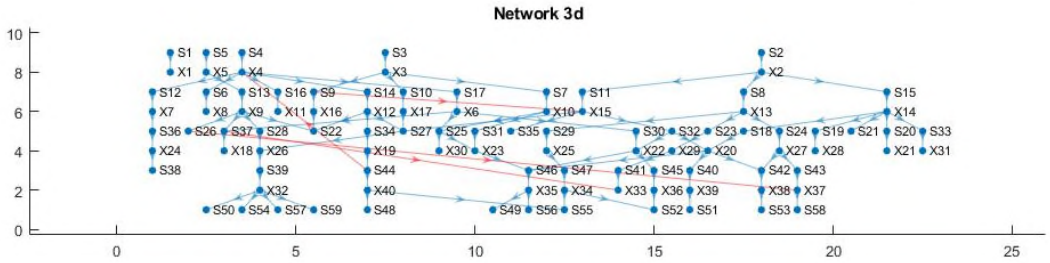

Network 3d
X5 X4 X3 X2 X1
X9 X7 X6 X8 X11 X10 X12 X17 X16 X15 X13 X14
X18 X24 X23 X30 X19 X22 X26 X25 X27 X20 X29 X21 X28 X31
X40 X36 X35 X32 X34 X37 X38 X39 X33
0 2 4 6 8 10 12 14 16 18 20

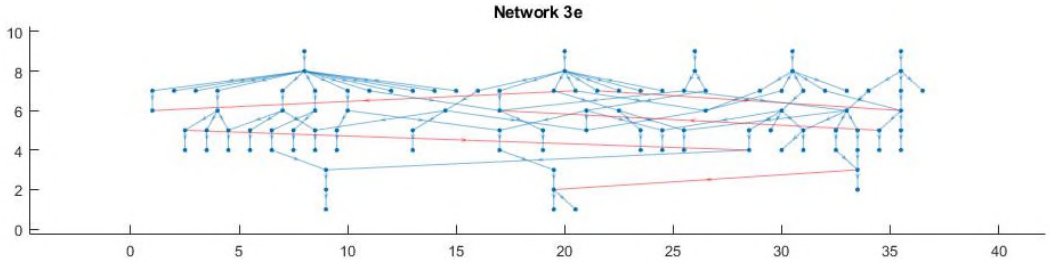

Network 3e
X4 X3 X5 X2 X1
X10 X8 X18 X16 X15 X13 X12 X7 X17 X9 X6 X11 X14
X23 X24 X35 X28 X22 X33 X25 X20 X37 X32 X34 X36 X21 X26 X30 X19 X31 X27 X39 X29 X38
X42 X41 X40
0 5 10 15 20 25

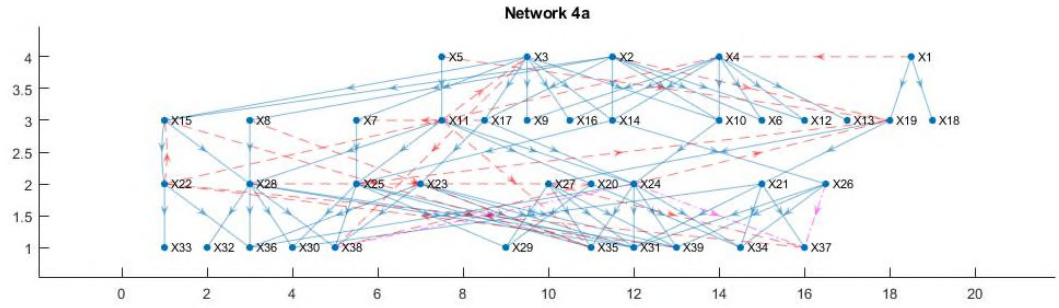

Network 3a

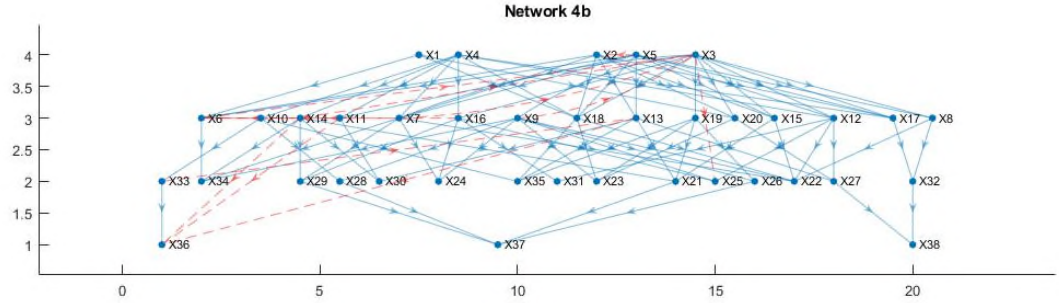

Network 3b

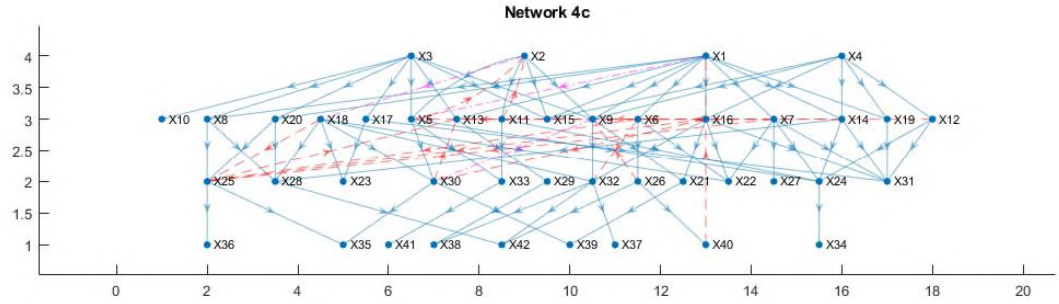

Network 3c

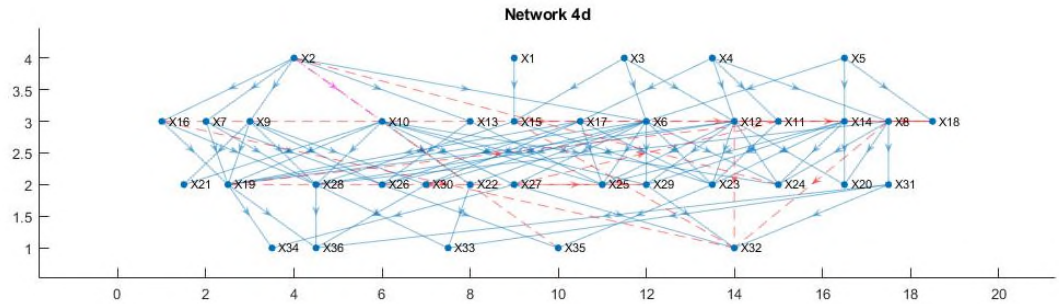

Network 3d

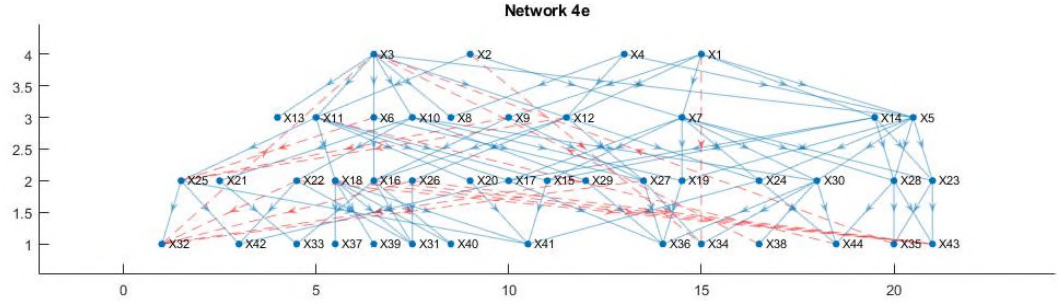

Network 3e

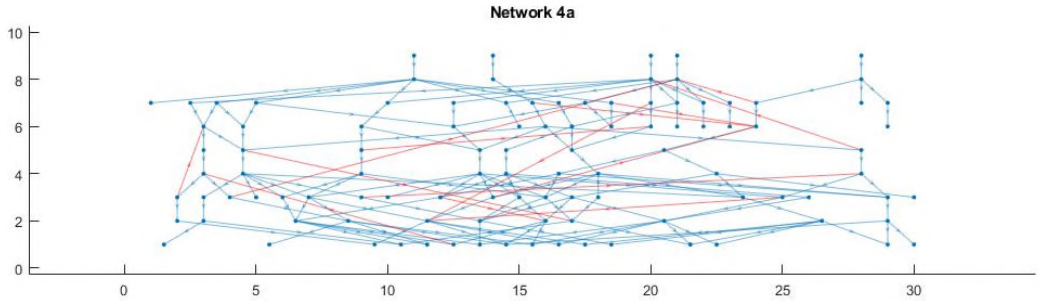

Network 4a
X5 X3 X2 X4 X1
X15 X8 X7 X11 X17 X9 X16 X14 X10 X6 X12 X13 X19 X18
X22 X28 X25 X23 X27 X20 X24 X21 X26
X33 X32 X36 X30 X38 X29 X35 X31 X39 X34 X37
0 2 4 6 8 10 12 14 16 18 20
1 1.5 2 2.5 3 3.5 4

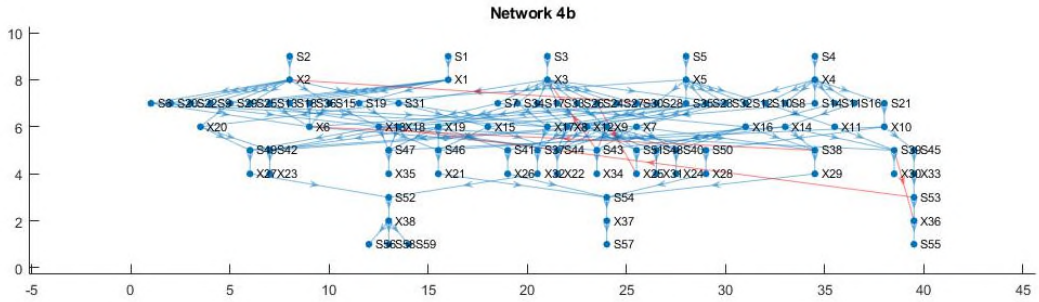

Network 4b
X1 X4 X2 X5 X3
X6 X10 X14 X11 X7 X16 X9 X18 X13 X19 X20 X15 X12 X17 X8
X33 X34 X29 X28 X30 X24 X35 X31 X23 X21 X25 X26 X22 X27 X32
X36 X37 X38
0 5 10 15 20
1 1.5 2 2.5 3 3.5 4

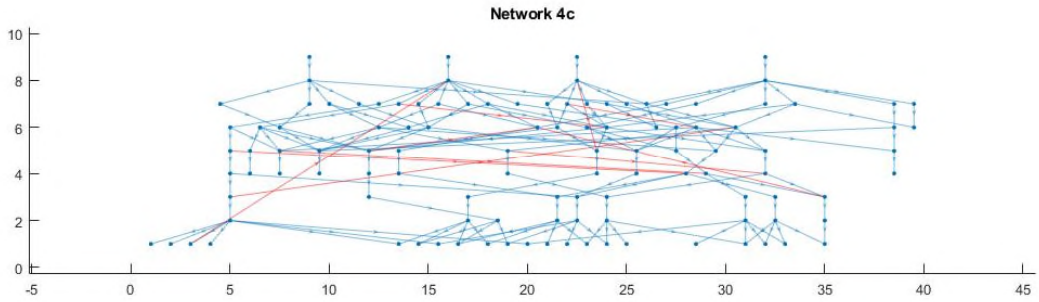

Network 4c
X3 X2 X1 X4
X10 X8 X20 X18 X17 X5 X13 X11 X15 X9 X6 X16 X7 X14 X19 X12
X25 X28 X23 X30 X33 X29 X32 X26 X21 X22 X27 X24 X31
X36 X35 X41 X38 X42 X39 X37 X40 X34
0 2 4 6 8 10 12 14 16 18 20
1 1.5 2 2.5 3 3.5 4

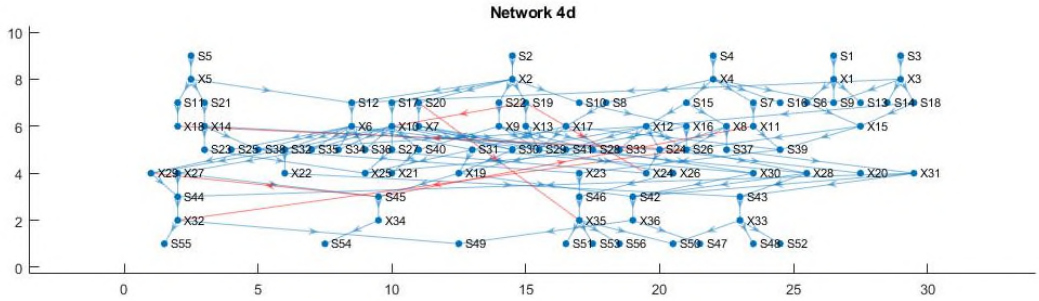

Network 4d
X2 X1 X3 X4 X5
X16 X7 X9 X10 X13 X15 X17 X6 X12 X11 X14 X8 X18
X21 X19 X28 X26 X30 X22 X27 X25 X29 X23 X24 X20 X31
X34 X36 X33 X35 X32
0 2 4 6 8 10 12 14 16 18 20
1 1.5 2 2.5 3 3.5 4

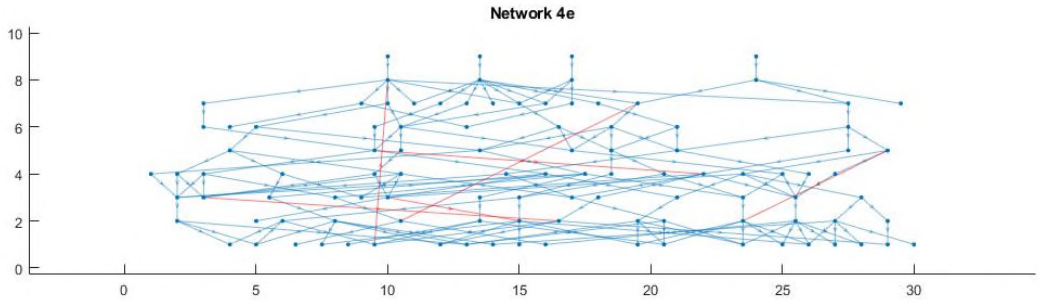

Network 4e
X3 X2 X4 X1
X13 X11 X6 X10 X8 X9 X12 X7 X14 X5
X25 X21 X22 X18 X16 X26 X20 X17 X15 X29 X27 X19 X24 X30 X28 X23
X32 X42 X33 X37 X39 X31 X40 X41 X36 X34 X38 X44 X35 X43
0 5 10 15 20
1 1.5 2 2.5 3 3.5 4


Network 4a


Network 4b


Network 4c


Network 4d


Network 4e

Chapter 7. Conclusions and future work

7.1. Summary of research and key conclusions

Microbial community based systems have been widely applied for engineering purposes, ranging from waste organic streams to biomedicine and bioprocessing. This project conducted systematic, model-based studies of microbial communities of two different types. For the first type, i.e. synthetic communities with a small number of well-defined species, systems with two important ecological relationships, complementarity and facilitation, were studied through two case studies, namely poly-culture of microalgae and multi-species based biohydrogen production. A further, generalised study on synthetic communities was also carried out. All these studies were for revealing important factors that affect the productivity and the stability of bioreactors employing such microbial communities, by means of numerical simulation and bifurcation analysis. For the second type of systems, i.e. (often naturally occurring) microbial communities that involve a large number of species which might not be clearly defined, this project focused on the challenge of inferring the network structures of such communities, and provided a thorough evaluation of an existing network inference model and a newly proposed model, to generate insights on their applicability and the factors affecting their performance. The learnings from both types of systems have the potential to guide future design and operations of bioreactors that employ a microbial community.

The studies and key conclusions in each chapter are summarised as follows:

- Chapter 3 conducted a study on the complementarity in light utilization in poly-culture of microalgae. Different microalgal species have different compositions of pigments which thus have different abilities to absorb different parts of light, hence leading to the potential of overyielding of polyculture compared to mono culture. Numerical simulations demonstrated that complementary light utilization has the potential of overyielding, however, there is no guarantee; operation settings and species traits are also important factors. It was also indicated that co-existence, a prerequisite for overyielding, is affected by different species traits, such as growth rate and light use efficiency. In general, a lower level of cell density and a higher level of light use complementarity is more likely for overyielding to happen. The model revealed the importance of a trade-off between light absorption and light utilization which dictates that an exceedingly high light intensity applied to a complementary co-culture can be contra-productive, because the species with a better ability to absorb light may not be able to efficiently use the light to grow, which results in a low efficiency of overall light utilization.
- Chapter 4 studied two microbial community systems for dark-photo fermentation based hydrogen production, a two-species system with *C.cellulovorans* and *R.palustris* and a three-species one with a hypothetic *E.coli*

introduced into the two species system for the purpose to further improve productivity. In all cases, the three-species system was predicted to offer improved productivity in terms of hydrogen production. With bifurcation analysis and numerical simulation, stability was analysed which revealed complex behaviours such as bi-stability and oscillation, arising from multiple inter-species relations, especially in the three-species system. Key operation and biological parameters affecting the behaviours were found, including pH tolerance, inlet substrate concentration and hydraulic retention time, and their importance were demonstrated. This work highlighted special attentions needed in designing a synthetic microbial community in order to avoid unwanted outcomes.

- Chapter 5 offered a generalized study extending the work in Chapter 4, on a set of artificially defined representative microbial community structures involving 2 or 3 species, with sequential cross-feeding as the backbone of a sophisticated set of ecological interactions. Bifurcation analysis and numerical simulation were applied for analysing those systems. Inhibition was demonstrated to decrease both productivity and stability in general. In three-species systems, productivity gain occurred by introducing the third species but with a risk to reduce the stability. Complex behaviours such as oscillation and bi-stability were found with specific system structural features such as feedback loops with specific ranges of parameters. Our results also demonstrated the potential to

qualitatively predict the behaviour of a system that combines simpler structures based on the knowledge of its simpler constituents.

- Chapter 6 investigated systems that involve a large number of species which are not necessarily clearly identified. In this chapter, the goal was to identify the inter-species relations in the community and to reconstruct the network that represents the community structure. Both an existing model, the generated Lotka Volterra model (gLV), and a new model that captures chemical-mediated interactions (CMI) between microbes, were evaluated against numerically generated data for anaerobic digestion and a set of artificially created communities with varying complexity. It was found that the gLV model performs better than CMI in general because of error accumulation in the application of CMI. With gLV, the inference of the food web structure was consistently more accurate than that of the inhibitions, due to the much larger influence of the former than the later on the system dynamics. However, CMI exhibited excellent performance in predicting both the food web and inhibitions, when more intensive perturbation was applied to generate the data. This work also showed the additional challenges in the inference of more dense networks, and the importance of sufficient perturbation to the system for generating data rich in information, particularly when the networks to be inferred were of a "deeper" structure (i.e. with more layers).

7.2. Limitations

In this mathematical modelling based project, experimental evidences or references have been drawn from literature where possible. However, no direct experimental validation has been carried out. More specifically, in chapter 3, the overyielding potential on a pair of microalgal species with actual light absorption spectra was demonstrated by means of numerical simulation. However, this finding is not validated by experiments. Similarly, in chapter 4, interesting dynamics such as bi-stability and oscillation, predicted by the mathematical models, have not been experimentally validated. Moreover, a hypothetical *E. coli* was proposed to construct a 3-species system, however, there is no experimental validation on the feasibility.

In all the bioreactor models used in this work, well-mixed systems were assumed, without considering the heterogeneity in the spatial distribution of microbial species and chemicals in bioreactors. How the spatial distribution would affect the outcome of ecological interactions remains a topic for future research.

Many of the mathematical models adopted in this work included various kinetic parameters such as maximum growth rate, half saturation constant and yield factors. These parameters were assumed to be constant. However, biological kinetics could change considerably according to the environmental factors (pH, temperature) and growth phases of microbes. More detailed models capturing such properties are needed to obtain more accurate results.

Finally, there is a limitation on the types of interactions between microbial species considered in Chapter 6. In reality, there are many different interactions between members in a microbial community, either positive (e.g. food web and commensalism) or negative (e.g. competition and inhibition). For simplicity and limited by the specific cases tested, only food web and inhibition were taken into consideration. The full range of microbial interactions involve links of different levels of nonlinearity and strength; the reliable inference of all these interactions remains a challenge for future explorations.

7.3. Future work

In addition to addressing the limitations discussed above, it is envisaged that future work may research into the following areas, to further improve our understanding and utilisation of microbial communities.

7.3.1 Flux balance analysis based models to predict phenotypes of microorganisms

In this work, a large amount of effort was spent to analyse the ecological interactions between microbial species in the community. However, the models to describe the dynamics of microbial species are still very simple, and hence may lead to inaccuracy by some assumptions, e.g. a constant yield factor. Therefore, a more comprehensive analysis on individual microbial species is needed. One possibility is to consider the flux balance analysis (FBA) model [1]. With a defined network of intra-cellular chemicals chemical reactions, a set of properly formulated constraints, cell phenotypes

such as growth rate and yield can be predicted solving an optimization problem for a chosen objective function, often by using a toolbox [2]. In this case, a more accurate model to describe the behaviour of microbial species is established, with which the ecological behaviours of a community could be modelled more accurately.

7.3.2 Linking data richness with statistical learning models for inferring microbial community networks

For network inference on microbial communities that involve a large number of species, a number of models have been developed to reconstruct networks from data. However, research is still weak on the data side, which becomes important especially when the network becomes very complex. There is theoretical work on conditions governing the possibility to reconstruct a network, however, there are few empirical studies on the link between inference models or approaches the richness of data. For example, in Chapter 6, it was observed that when different perturbation strategies were used to excite the system for data collection, different networks were inferred due to the difference in the information richness of the data collected. However, the exact connection between the type of networks to be inferred and the corresponding requirement for data or the perturbation strategy is still not clear. Future work to provide a more precise picture linking the statistical modelling tools with the data requirement would help modellers to determine the most appropriate data generation strategies for a network inference task at hand [3].

7.3.3 Applications of inferred microbial community networks

A network is a type of representation of the topological structure of a system, and it has been shown that social, technological and internet networks can be used for different purposes. In microbial community networks, real applications could gain from a

number of network measurements and analysis, such as centrality [4], community detections [6] and control [5]. It is obvious that accurate network inference from data becomes a key for further applications. However, this work and other research only focuses on measure a network inference from a statistical point of view. Future work could examine network inference accuracy in connection with certain intended applications, to make notion of accuracy more context-specific.

References

[1]. Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? Nature biotechnology, 28(3), 245.

[2]. Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., ... & Kang, J. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature protocols, 6(9), 1290.

[3]. Angulo, M. T., Moreno, J. A., Lippner, G., Barabási, A. L., & Liu, Y. Y. (2017). Fundamental limitations of network reconstruction from temporal data. Journal of the Royal Society Interface, 14(127), 20160966.

[4]. Shaw, G. T. W., Liu, A. C., Weng, C. Y., Chou, C. Y., & Wang, D. (2017). Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of hog waste. PloS one, 12(7), e0181395.

[5]. Angulo, M. T., Moog, C. H., & Liu, Y. Y. (2019). A theoretical framework for controlling complex microbial communities. Nature communications, 10.

[6]. Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582.