Modeling and optimization of the production of lipids by oleaginous yeasts

Carlos Eduardo Robles Rodriguez

Université Fédérale

Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue par :
Carlos Eduardo ROBLES RODRIGUEZ

le mercredi 19 octobre 2016

Titre :

Modélisation et optimisation de la production de bio-lipides par les levures oléagineuses

---

École doctorale et discipline ou spécialité :
ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP)

Directeur/trice(s) de Thèse :

César Arturo ACEVES LARA (Maître des Conférences, INSA)
Carole MOLINA-JOUVE (Professeure des Universités, INSA)

Jury :
Ramón VILANOVA-ARBOS (Professeur, Université Autonome de Barcelone) Rapporteur
Olivier BERNARD (Directeur de Recherche, INRA) Rapporteur
Gilles ROUX (Professeur des Universités, UPS) Examinateur
Fatiha NEJJARI (Professeure, Université Polytechnique de Catalogne) Examinatrice
Alain RAPAPORT (Directeur de Recherche, INRA) Examinateur
Elsa PARMENTIER (Scientifique d'Innovation, TEREOS) Invitée

Last name: ROBLES RODRIGUEZ **Name:** Carlos Eduardo

Title: Modeling and optimization of the production of lipids by oleaginous yeasts.

Year: 2016 **Place:** INSA Toulouse

Abstract

This PhD thesis aims at optimizing lipid accumulation by oleaginous yeast, and most particularly by the yeast *Yarrowia lipolytica* from glucose. This optimization is addressed from a mathematical point of view based on automatic control laws, where model-based control strategies are proposed. The bibliographic review compiles and evaluates previous works to identify the different existing methodologies to attain the optimization, which is divided in three main axes: modeling, control strategies, and monitoring. In this context, five different models are proposed to describe lipid accumulation. The first model is based on Monod and inhibition kinetics (unstructured), and the second on the Droop quota model (quota) previously used for microalgae. The last three are dynamic metabolic models that combine kinetics with metabolic models based on the stoichiometry of metabolism. These three models used a reduced metabolic network decomposed into elementary flux modes. The five models were successfully calibrated and validated with different experimental data. Nonetheless, the dynamic metabolic models presented highlighting features such as the description of metabolic shifts. Two approaches of multi-objective control strategies aiming at maximizing lipid productivity and lipid content fraction were proposed. In the first, the two objectives were weighted by static calculation of Pareto fronts, and integrated to the control strategy by dynamic optimization algorithms with a dynamic metabolic model. The second strategy used a constant weighed objective function solved by piecewise linear functions by integrating the quota model. The simulation results of the optimization attained lipid contents between 0.2 – 0.23 g*LIP*.gX^{-1} and productivities between 0.78 – 0.95 g.(L-h)$^{-1}$ shortening the culture time to 20 h. Soft-sensors were developed by correlating on-line measurements (*i.e.* pO_2 and the added base for pH) through support vector machines in order to overcome the lack of measurements. The perspective is to experimentally validate the control strategies.

Keywords:

Optimization, Modeling, *Yarrowia lipolytica*, State estimation, Optimal control, Lipid accumulation, Metabolic shifts

Nom: ROBLES RODRIGUEZ **Prénom:** Carlos Eduardo

Titre: Modélisation et optimisation de la production des bio-lipides par les levures oléagineuses

Année: 2016 **Lieu:** INSA Toulouse

Résumé

Le but de ce travail de doctorat est de contribuer à l'optimisation de l'accumulation de lipides par les levures oléagineuses, et plus particulièrement par la levure *Yarrowia lipolytica* à partir du glucose, avec une stratégie d'optimisation dynamique à partir d'une commande basée sur un modèle. L'étude bibliographique permet de faire le bilan des connaissances antérieures pour identifier les différentes méthodologies existantes pour l'optimisation des procédés en structurant trois parties principales: la modélisation, la commande, et le suivi des procédés. Dans ce contexte, cinq modèles ont été proposés pour décrire l'accumulation de lipides. Le premier est un modèle non structuré basé sur des cinétiques de Monod et d'inhibition. Le deuxième s'appuie sur le modèle de Droop (quota) précédemment utilisé pour décrire l'accumulation de lipides dans les microalgues. Les trois derniers sont des modèles métaboliques dynamiques qui combinent les cinétiques avec un réseau métabolique réduit, obtenu à partir des modes élémentaires. Les cinq modèles ont été calibrés et validés en utilisant plusieurs jeux des données expérimentales. Cependant, un des avantages des modèles métaboliques dynamiques présentés est la possibilité de décrire les basculements métaboliques. Deux stratégies de commande multi-objective visant à maximiser la productivité des lipides et la fraction en teneur lipidique ont été proposées. Dans la première, les deux objectifs ont été pondérés par le calcul statique des fronts de Pareto, et intégrés à la stratégie de commande avec un modèle dynamique métabolique. La deuxième stratégie est basée sur des fonctions linéaires par morceaux en intégrant le modèle quota. Les simulations de la commande montrent la possibilité d'atteindre des teneurs en lipides entre 0,2 – 0,23 gLIP.gX^{-1} et productivités entre 0,78 – 0,95 g.(L-h)$^{-1}$ en diminuant le temps de la culture à 20 h. Des capteurs logiciels ont été proposés afin de pallier le manque de capteurs en ligne en corrélant des mesures en ligne (*i.e.* pO2 et la base ajoutée pour le pH) par des algorithmes de types machines à vecteurs supports. La validation expérimentale des stratégies de commande est la principale perspective de ce travail.

Mots clés:

Optimisation des procédés, Modélisation, *Yarrowia lipolytica*, Estimation des Etats, Commande optimale, Accumulation des lipides, Basculements métaboliques.

"It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."

Carl Friedrich Gauss

REMERCIEMENTS

Je tiens à remercier à toutes les personnes qui ont contribué à la réalisation de mes travaux de thèse.

Tout d'abord, je remercie aux membres du jury: M. Ramón VILANOVA, M. Olivier Bernard, Mme. Fatiha NEJJARI, M. Alain RAPAPORT, Mme. Elsa PARMENTIER, et M. Gilles ROUX pour avoir accepté de juger cette thèse, et le court délai pour faire un retour sur ce long travail.

Je remercie à la direction du Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, précédemment dirigée M. Nic LINDLEY et actuellement par Mme. Carole MOLINA-JOUVE pour avoir autorisé le déroulement ma thèse au sein de son laboratoire.

Une reconnaissance spéciale va en direction de M. Claude MARANGES, directeur de l'école doctorale SEVAB.

Je tiens à remercier au projet ProBio3, et plus particulièrement à TEREOS pour le financement de cette thèse.

Je remercie également à ma directrice Mme Carole MOLINA-JOUVE pour ses remarques et suggestions dans mon travail, toujours très pertinents, qui m'ont vraiment aidé à mieux expliquer le côté bio de ma thèse.

Des très grands remerciements vont pour M. Gilles ROUX qui a été toujours disponible pour des discussions, relectures et améliorations de ce travail.

Les plus grands des remerciements pour mon directeur M. César Arturo ACEVES-LARA pour m'avoir donné l'opportunité de venir à Toulouse à réaliser ce travail. Merci pour toutes les discussions professionnelles et personnelles, ainsi comme les conseils pour continuer dans ce beau métier.

Je remercie à l'équipe FAME, encadrée par M. Stéphane GUILLOUET, pour sa réactivité et les connaissances partagées. Je tiens à remercier Mme. Nathalie GORRET et l'ensemble de l'équipe, dont plus particulièrement à Mme. Carine BIDEAUX pour ses conseils de modélisation et les corrections des publications.

Un grand merci à tous mes collègues et amis avec qui j'ai partagé mon séjour dans le laboratoire. Je pense à Xiaomin, Rana, Asma, Marine D, Tuan, Susana, Jian, Çeren, Claudio, Ana, Cléa, Agustina, Nicolas, Florence, Marine C, et ceux qui sont parti Henrily, Chainese, Cuong, Rafael, Lucie, Céline, Aurore C, Yannick, Ulysse. Merci pour tous les moments partagés surtout au repas et/ou au tour d'une bière.

Au-delà du labo, je remercie mes amis à Toulouse Lauren, Sergio, Lexuri, Christian, et à mes amis géographiquement éloignés qui sont toujours resté proche de moi. Toño, Pris, Luis Jorge, Abraham, Adriana, Maira, Brayan, Karla, Pepe (avec mes excuses pour les oubliés).

Pour finir, j'expresse ma profonde gratitude à ma famille, mes parents, mon frère, et ma sœur pour sa patience et son soutien pendant tous ces années. 'No tengo palabras para agradecerles todo lo que han hecho por mí y su apoyo incondicional, los amo'.

PUBLICATIONS AND COMMUNICATIONS

International Journals

Robles-Rodriguez, C.E., Bideaux, C., Guillouet S.E., Gorret, N., Cescut, J., Uribelarrea, J-L., Molina-Jouve, C., Roux, G., Aceves-Lara,C.A. 2017. 2017. "Dynamic metabolic modeling of lipid accumulation and citric acid production by *Yarrowia lipolytica". Computers & Chemical Engineering*, *100*, 139–152. https://doi.org/10.1016/j.compchemeng.2017.02.013.

Robles-Rodriguez, C.E., Munoz-Tamayo, R., Bideaux, C., Guillouet S.E., Gorret, N., Molina-Jouve, C., Roux, G., Aceves-Lara, C.A. "Modeling and Optimization of lipid accumulation during Fed-batch fermentation by *Yarrowia lipolytica* from glucose under nitrogen depletion conditions". (Submitted to *Biotechnology and Bioengineering*).

Robles-Rodriguez, C.E., Bideaux, C., Guillouet S.E., Molina-Jouve, C., Roux, G., Aceves-Lara, C.A. "An overview on Dynamic Metabolic Modeling of yeasts towards process control". (Under preparation).

Book Chapter

Robles-Rodriguez, C.E., Bideaux, C., Roux, G., Molina-Jouve, C., Aceves-Lara, C.A. "Soft-sensors for lipid fermentation based on PSO Support Vector Machine (PSO-SVM). *Distributed Computing and Artificial Intelligence, 13th International* Conference. Volume 474 of the series Advances in Intelligent Systems and Computing. Pp 175-183. ISSN: 2194-5357. ISBN: 978-3-319-40162-1. DOI: 10.1007/978-3-319-40162-1_19.

International Conferences with online proceedings

Robles-Rodriguez, C.E., Bideaux, C., Gaucel, S., Laroche, B., Gorret, N., Aceves-Lara, C.A. "Reduction of metabolic models by polygons optimization method applied to bioethanol production with co-substrates". *19th World Congress of the International Federation of Automatic Control (IFAC 2014).* August 24-29 2014, Cape Town, South Africa. (Oral Presentation). Pp 6198-6203. ISSN: 14746670. ISBN: 978-390282362-5. DOI: 10.3182/20140824-6-ZA-1003.01037.

Robles-Rodriguez, C.E., Guillouet, S.E., Bideaux, C., Roux, G., Gorret, N., Hulin, S, Molina-Jouve, C., Aceves-Lara, C.A. "Dynamical metabolic modelling for lipid accupulation by Yarrowia lipolytica growing on glucose". *23th Mediterranean Conference on Control and Automation (MED 2015).* June 16 – 19, 2015. Malaga, Spain. (Oral Presentation). Pp 815-820. ISBN: 978-147999936-1. DOI: 10.1109/MED.2015.7158846.

Robles-Rodriguez, C.E., Bideaux, C., Roux, G., Molina-Jouve, C., Aceves-Lara, C.A. "Soft-sensors for lipid fermentation based on PSO Support Vector Machine (PSO-SVM). *13th International Conference on Distributed Computing and Artificial Intelligence.(DCAI 2016).* June 1-3, 2016. Seville, Spain. (Oral Presentation). Pp 175-183 ISSN: 2194-5357. ISBN: 978-3-319-40162-1. DOI: 10.1007/978-3-319-40162-1_19.

Robles-Rodriguez, C.E., Guillouet, S.E., Bideaux, C., Roux, G., Gorret, N., Molina-Jouve, C., Aceves-Lara, C.A. "Multi-objective particle swarm optimization (MOPSO) of lipid accumulation in Fed-batch cultures". *24th Mediterranean Conference on Control and Automation. (MED 2016).* June 21 – 24, 2016. Athens, Greece. (Oral Presentation). Pp 979 – 984. ISBN: 978-1-4673-8345-5. DOI: 10.1109/MED.2016.7535934.

INDEX OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

GENERAL INTRODUCTION

In the last years, the declining of petroleum reserves, and the utilization of fossil resources have been discussed regarding environmental impacts due to CO_2 emissions (Fernando *et al*., 2006; Naik *et al*., 2010). In 2013, world statistics reported that 46% of CO_2 emissions were produced from fuel combustion, 33% from oil and 20% from gas (CO_2 Emissions, 2015). Therefore, in order to reduce greenhouse gases emissions, many countries have adopted governmental measures to increase the production and use of biofuels (*e.g.* Energy Independence and Security Act of 2007) (Vicente *et al*., 2009). This fact has encouraged research towards the development of new alternatives for producing biofuels with major interest in transportation fuels.

Biofuels were proposed to ensure environmentally friendly fuels. The first generation biofuels are from edible feedstocks including sugar, starch and oil crops. The most common first generation biofuels are biodiesel and bioethanol. Even if efficient, some concerns surround this type of biofuels about the overall carbon footprint, environmental damage, and the food versus fuel competition, which set its principal drawback (Havlík *et al*., 2011; Luque *et al*., 2008).

Second generation biofuels were introduced to overcome the major drawbacks of producing first generation biofuels. In this case, alternative feedstocks, generally non-edible feedstocks including waste vegetable oils and fats, non-food crops and biomass sources, and/or technologies started to be developed (Koutinas *et al*., 2014). In theory, these could solve food versus fuel debate and could supply a larger proportion of fuel in a more sustainable way with greater environmental benefits, such as the use of alternative lands so-called "wastelands" (Janaun and Ellis, 2010).

Production of second generation biofuels from non-edible feedstocks (*e.g.* forestry wastes, food industrial wastes, agro industrial by-products) can be obtained from hydrolysis (*i.e.* enzymatic, acid, *etc.*) of the lignocellulosic non-edible biomass to be further transformed into biofuels by microorganisms (Jin *et al*., 2015). The possibility to use lignocellulosic biomass suggests lower costs and a better environmental performance (Granda *et al*., 2007)

Recently, aviation industry investigates new manners to produce sustainable fuels. The European aviation stakeholders are promoting researches into bio jet fuel production in order to reach a target of 2 million tons of biofuel produced in Europe by 2020 (European Advanced

Biofuels Flight Path Initiative). This industry is committed to ambitious and challenging targets to improve its environmental performance, including a 50% reduction in the aviation net carbon emissions by 2050 (Babau *et al*., 2013). These bio jet fuels, also called 'drop-in' fuels, will be blended with fossil fuels in order to lower the changes in aircraft engines (25 years of average lifetime) and to achieve identical overall performances as with jet A1. Among the different alternative emerging biofuels certified by ASTM, the Hydro-processed Esters and Fatty Acids (HEFA) Jet Fuels are of major interest (Standard ASTM D1655). HEFA are produced from oils that are converted into bio-Synthetic Paraffinic Kerosene by hydro-treatment. These physicochemical reactions are well known and therefore the challenge lies in oil production and its supply (Llamas *et al*., 2012).

In this context, this PhD thesis was carried out within the ProBio3 project (ref. ANR-11-BTBT-0003) and supported by TEREOS, an industrial partner of this project. ProBio3 project aims at the bio-catalytic production of lipid bio-products from renewable raw materials and industrial by-products, with a special interest on Bio-jet fuel application. Therefore, this work was proposed for enabling advances in the knowledge of microbial lipid metabolism towards innovative bioprocesses providing high productivities and high lipid contents.

One promising alternative for the production of oils is the utilization of oleaginous microorganisms. Microbial lipids, also called single cell oils (SCO) present many advantages over vegetable oils because they do not require arable land, and this production does not compete with food since lignocellulosic biomass and industrial by-products can be used as carbon source (Vicente *et al*., 2009).

Oleaginous microorganisms can accumulate high content of lipids (more than 20% w/w) (Ratledge, 2004), among which yeasts are good candidates with high conversion yields, high growth rates, and high lipid contents with interesting lipid profiles for biofuel production (Meng *et al*., 2009).

One interesting yeast is *Yarrowia lipolytica*, which is able to accumulate lipids as triacylglycerol with a promising composition for the production of biofuels. The genome of this oleaginous yeast is fully sequenced (Barth and Gaillardin, 1997), and thus a genetically engineering yeast can be achieved with ease to modulate lipid profile and to optimize lipid

production. *Yarrowia lipolytica* has been studied in fermentation processes under various operating conditions in order to get higher quantity of lipids (Beopoulos et al., 2009). Lipid accumulation on *Yarrowia lipolytica* can be triggered, by nutrient depletion (*e.g.* nitrogen) and the excess of carbon. However, lipid conversion yield could be hampered by the production of by-products such as citric acid.

In this context, the principal aim of the project is to optimize lipid production, which could be seen as the maximization of three conflicting criteria: lipid productivity, lipid content, and the yield conversion of substrate into lipids. These three objectives could be handled by different strategies involving:

- Genetic engineering for the development of mutant strains where plenty of works have been performed (Friedlander et al., 2016; Ledesma-Amaro and Nicaud, 2016; Nicaud, 2012; Rakicka et al., 2015).
- Bioprocesses engineering that has tested different culture modes (Ochoa-Estopier and Guillouet, 2014; Papanikolaou et al., 2006; Papanikolaou and Aggelis, 2011), and
- Bioprocess control, which has not been yet exploited due to the lack of accurate models on the subject.

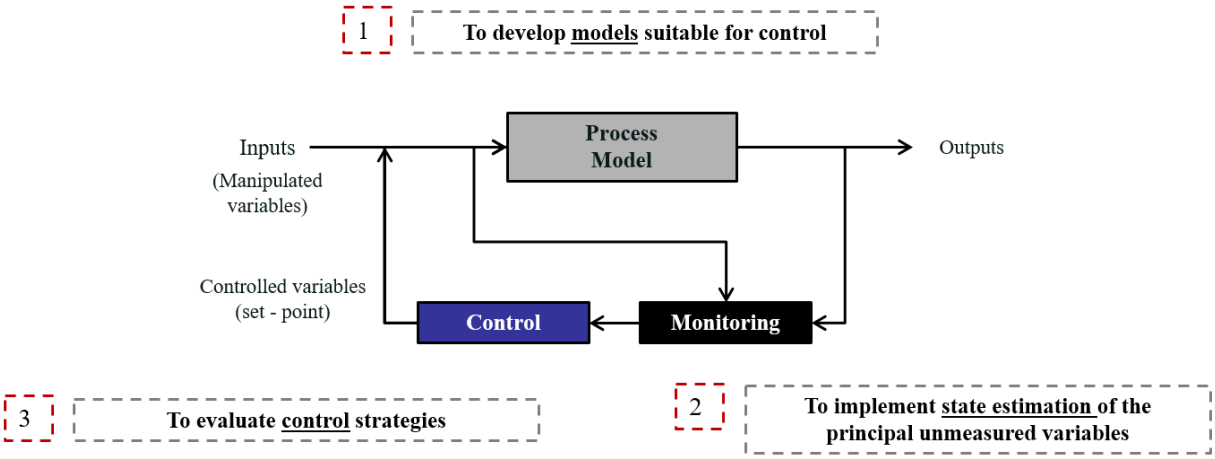


Figure 0.1. Approach of this PhD thesis.

In this regard, **the aim of this work is to maximize lipid accumulation in *Yarrowia lipolytica* from glucose by optimal control strategies**. In order to accomplish this objective, this work proposes the development of **models to describe the metabolic shifts of *Yarrowia lipolytica*, a control strategy based on a multi-objective predictive optimization, and the**

reconstruction of state variables by software sensors. This general objective could be thus seen and addressed by three complementary main axes: (i) modeling, (ii) control, and (iii) state estimation. The general approach of this PhD thesis is described in Figure 0.1.

This PhD thesis comprises four chapters. The first chapter is a state-of-the-art on lipid accumulation with three sections. The first section gives the insights of biofuels production, whose precursors could be lipids produced by oleaginous microorganisms, such as yeasts. In the second and third sections of this chapter, a general overview of lipid accumulation is presented regarding first the metabolism of *Yarrowia lipolytica* and then the factors influencing lipid accumulation process.

The second chapter, organized in four sections, reviews the three axes of this work. The first and second sections are dedicated to modeling, where the existing models of lipid accumulation are described first, followed by a focus on the Dynamic Metabolic Modeling (DMM) approach. The DMM approach is regarded in two steps: the static metabolic analysis, and the dynamic adaptation of metabolism. Afterwards, in a third section, different optimization algorithms for control strategies on fermentation are revised and compared. Finally, the monitoring tools called software-sensors are described analyzing their advantages and drawbacks.

The third chapter is composed by seven sections holding the main results of this work. The first is an introduction that summarizes the five following sections, which are presented as publications. The second section deals with a new method to reduce metabolic networks. Third section develops three dynamic metabolic models based on a reduced metabolic network. The methodology used for two of these models was found within previous works, whereas the third one is a new proposition. Fourth section presents the development of two macroscopic models, where one is further applied in a control strategy in order to optimize lipid productivity and lipid content. The fifth section exposes a control strategy based on the multi-objective model predictive optimization of lipid accumulation. The control law uses a dynamic metabolic model to predict the conflicting objectives. Section six develops software-sensors to estimate on-line the dynamics of some state variables. Closing the chapter, the seventh section provides the discussion and conclusions of the presented publications.

Finally, chapter four draws the conclusions and the perspectives of the obtained results and gives the insights of further works to be performed.

The material and methods for obtaining the experimental data used throughout this work are presented in an appendix which is divided in three sections. The strain and the culture medium are described in the first. The second section reports the description of the analytical methods to determine the major metabolites of these cultures. Finally, three experimental cultures are presented consisting on one fed-batch culture and two sequential batch cultures.

REFERENCES

CO2 Emissions from fuel combustion: HIGHLIGHTS. 2015. IEA Statistics. International Energy Agency. Retrived from (https://www.iea.org/publications/freepublications/publication/CO2EmissionsFromFuelCombustionHighlights2015.pdf)

Standard specification for aviation turbine fuels. Standard ASTM D1655-09a. West Conshohocken, Pennsylvania: ASTM International; 2009.

Biofuels for aviation. European Advanced Biofuels Flightpath Initiative. Retrieved from (https://ec.europa.eu/energy/en/topics/biofuels/biofuels-aviation).

Babau, M., Cescut, J., Allouche, Y., Lombaert-Valot, I., Fillaudeau, L., Uribelarrea, J.-L., Molina-Jouve, C., 2013. Towards a Microbial Production of Fatty Acids as Precursors of Biokerosene from Glucose and Xylose. Oil Gas Sci. Technol. – Rev. d'IFP Energies Nouv. 68, 899–911. doi:10.2516/ogst/2013148

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol. Rev. 19, 219–37.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Fernando, S., Adhikari, S., Chandrapal, C., Murali, N., 2006. Biorefineries: Current Status, Challenges, and Future Direction. Energy & Fuels 20, 1727–1737. doi:10.1021/ef060097w

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E.H., Consiglio, A.L., MacEwen, K., Crabtree, D. V., Afshar, J., Nugent, R.L., Hamilton, M.A., Joe Shaw, A., South, C.R.,

Stephanopoulos, G., Brevnova, E.E., 2016. Engineering of a high lipid producing Yarrowia lipolytica strain. Biotechnol. Biofuels 9, 77. doi:10.1186/s13068-016-0492-3

Granda, C.B., Zhu, L., Holtzapple, M.T., 2007. Sustainable liquid biofuels and their environmental impact. Environ. Prog. 26, 233–250. doi:10.1002/ep.10209

Havlík, P., Schneider, U.A., Schmid, E., Böttcher, H., Fritz, S., Skalský, R., Aoki, K., Cara, S. De, Kindermann, G., Kraxner, F., Leduc, S., McCallum, I., Mosnier, A., Sauer, T., Obersteiner, M., 2011. Global land-use implications of first and second generation biofuel targets. Energy Policy 39, 5690–5702. doi:10.1016/j.enpol.2010.03.030

Janaun, J., Ellis, N., 2010. Perspectives on biodiesel as a sustainable fuel. Renew. Sustain. Energy Rev. 14, 1312–1320. doi:10.1016/j.rser.2009.12.011

Jin, M., Slininger, P.J., Dien, B.S., Waghmode, S., Moser, B.R., Orjuela, A., Sousa, C., Balan, V., Sousa, L. da C., Balan, V., 2015. Microbial lipid-based lignocellulosic biorefinery : feasibility and challenges. Trends Biotechnol. 33, 43–54. doi:10.1016/j.tibtech.2014.11.005

Koutinas, A.A., Chatzifragkou, A., Kopsahelis, N., Papanikolaou, S., Kookos, I.K., 2014. Design and techno-economic evaluation of microbial oil production as a renewable resource for biodiesel and oleochemical production. Fuel 116, 566–577. doi:10.1016/j.fuel.2013.08.045

Ledesma-Amaro, R., Nicaud, J.-M., 2016. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. Prog. Lipid Res. 61, 40–50. doi:10.1016/j.plipres.2015.12.001

Llamas, A., García-Martínez, M., Al-Lal, A.-M., Canoira, L., Lapuerta, M., 2012. Biokerosene from coconut and palm kernel oils: Production and properties of their blends with fossil kerosene. Fuel 102, 483–490. doi:10.1016/j.fuel.2012.06.108

Luque, R., Herrero-Davila, L., Campelo, J.M., Clark, J.H., Hidalgo, J.M., Luna, D., Marinas, J.M., Romero, A.A., 2008. Biofuels: a technological perspective. Energy Environ. Sci. 1, 542. doi:10.1039/b807094f

Meng, X., Xian, M., Xu, X., Zhang, L., Nie, Q., Yang, J., Xu, X., Zhang, L., Nie, Q., Xian, M., 2009. Biodiesel production from oleaginous microorganisms. Renew. Energy 34, 1–5. doi:10.1016/j.renene.2008.04.014

Naik, S.N., Goud, V. V., Rout, P.K., Dalai, A.K., 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. 14, 578–597. doi:10.1016/j.rser.2009.10.003

Nicaud, J.M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Papanikolaou, S., Aggelis, G., 2011. Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. Eur. J. Lipid Sci. Technol. 113, 1031–1051. doi:10.1002/ejlt.201100014

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., Nicaud, J.M., 2015. Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. Biotechnol. Biofuels 8, 104. doi:10.1186/s13068-015-0286-z

Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. Biochimie 86, 807–15. doi:10.1016/j.biochi.2004.09.017

Vicente, G., Bautista, L.F., Rodríguez, R., Gutiérrez, F.J., Sádaba, I., Ruiz-Vázquez, R.M., Torres-Martínez, S., Garre, V., 2009. Biodiesel production from biomass of an oleaginous fungus. Biochem. Eng. J. 48, 22–27. doi:10.1016/j.bej.2009.07.014

CHAPTER I.

AN OVERVIEW TO LIPID ACCUMULATION

CHAPTER I. AN OVERVIEW TO LIPID ACCUMULATION

SUMMARY OF THE CHAPTER

This chapter presents a state-of-the-art on lipid accumulation. This is addressed in three sections revising first the oleaginous microorganisms, where oleaginous yeasts are presented as competitive microorganisms. Special interest is paid to *Yarrowia lipolytica*, which was the selected microorganism due to its ability to assimilate a great variety of substrates and its fully sequenced genome.

Afterwards, a brief review of the metabolism is reported containing the most significant pathways and enzymes involved in lipid accumulation of *Yarrowia lipolytica*. All these pathways were put together to construct a reduced metabolic network.

Finally, details on the culture modes and the different factors affecting the optimization of lipid accumulation this process are provided.

I. OLEAGINOUS MICROORGANISMS

The so called "oleaginous microorganisms" accumulate lipids or single cell oils (SCO) up to 20% of their dry cell weight. These microbial lipids are mainly stored as triacylglycerols (TAG) at different quantities within algae, bacteria, fungi, and yeasts (Table I.1) (Babau *et al*., 2013; Meng *et al*., 2009).

Regarding the performance of microorganisms, only few bacteria can be used for lipid accumulation from simple carbon sources (*e.g.* glucose) under growth-restricted conditions (Meng *et al*., 2009). However, the accumulated lipids in bacteria are generally not produced as triacylglycerols, they are stored as polymers, such as poly-b-hydroxy-butyrates and alkanaoates (Ratledge, 2004).

Table I.1. Microorganisms for production of SCO (Babau *et al*., 2013; Meng *et al*., 2009)**.**

Microorganism	Oil content (% dry cell)	Microorganism	Oil content (% dry cell)
Algae		Bacteria	
Botryococcus braunii	25 – 75	*Acinetobacter calcoaceticus*	27 – 38
Cylindrotheca sp.	16 – 37	*Arthrobacter AK 19*	78
Nannochloris sp	31 – 68	*Arthrobacter* sp.	>40
Nitzchia sp.	45 – 47	*Bacillus alcalophilus*	18 – 24
Schizochytrium sp.	50 – 77	*Rhodococcus opacus*	24 – 25
Chlorella sp.	28 – 32		
Isochrysis galbana	22 – 38		
Haematococcus pluvialis	30 – 40		
Fungi		Yeast	
Aspergillus oryzae	57	*Candida curvata*	58
Aspergillus terreus	57	*Cryptococcus albidus*	65
Entomopthora coronate	43	*Cunninghamella japonica*	60
Humicola lanuginose	75	*Lipomyces starkeyi*	64
Morterella vinacea	66	*Rhizopus arrhizus*	57
Mortierella isabellina	86	*Rhodosporidium toruloides*	72
Mortierella elongate	24	*Trichosporon fermentans*	35
Trichosporon cutaneum	39.2	*Trichosporon pullulans*	65
		Yarrowia lipolytica	36
		Debaryomyces etchellsii	61 – 66

Microalgae use CO_2 as carbon source and light as energy source. Nevertheless, the harvest of microalgae and the manner in which light is provided are great challenges to overcome (Huang *et al*., 2010). Moreover, they require aquatic environments that may vary from fresh-water to sea-water and the supply of nitrogen (ammonia or nitrate) as well as essential elements (Singh

et al., 2011). Those nitrates and phosphates are normally from chemical fertilizers (Lam and Lee, 2012).

Oil production by yeast is promising for the production of SCO because yeast exhibit high growth rate with low nutrient requirements, and certain species accumulate fatty acids with a composition comparable to plant oils (Meng *et al*., 2009). Although fungi reported high lipid contents, most of them are explored for the production of some special lipids such as γ-linoleic, arachidonic acid, and eicosapentoenoic acid (Li *et al*., 2008). Yeasts are better candidates than fungi for the production of SCO because of their faster growth and their easier cultivation at large scale (Ratledge and Cohen, 2008). The best known oleaginous yeasts are *Lipomyces starkeyi*, *Rhorosporidium toruloides,* and *Yarrowia lipolytica*, which are able to accumulate lipids up to 64, 66, 72 and 36 percent of dry cell weight respectively (Table I.1).

The main lipid compounds from yeast and fungi are triacylglycerols (more than 80% TAGs) with a fatty acid profile rich in compounds with a carbon chain length of C16 and C18, showing desaturations delta9, delta12 and delta15 (Beopoulos *et al*., 2009a; Cescut, 2009; Ratledge and Wynn, 2002). Table I.2 displays the fatty acid composition of yeasts, where oleic acid (C18:1) is the main compound (Ahmad *et al*., 2015; Babau *et al*., 2013; Li *et al*., 2008; Meng *et al*., 2009).

Table I.2. Fatty acid composition of several oleaginous microorganisms (Ahmad *et al*., 2015; Babau *et al*., 2013; Li *et al*., 2008; Meng *et al*., 2009).

Microorganism	Lipid (% dry cell)	**Fatty acid composition**					
		C16:0	C16:1	C18:0	C18:1	C18:2	C18:3
Candida curvata	58	25	1	10	57	7	0
Cryptococcus albidus	65	12	1	3	73	12	0
Cunninghamella japonica	60	16	0	14	48	4	8
Lipomyces starkeyi	63	34	6	5	51	3	0
Rhodotorula glutinis	72	37	7	3	47	8	0
Trichosporon pullulans	65	15	0	2	57	24	1
Yarrowia lipolytica	36	11	6	1	28	51	1

Several substrates have been studied for oleaginous yeasts and fungi (Table I.3). On glucose (Cescut, 2009; Ochoa-Estopier, 2012) the theoretical stoichiometric yield for oleaginous yeasts is 0.32 g of TAG per gram of glucose (Ratledge and Wynn, 2002). Regarding other osidic substrates, the ratio xylose/glucose was tested for the growth of *Rhodotorula glutinis* under fed-batch conditions with a constant total carbon flus (Babau *et al*., 2013). The use of both substrates achieved 72% of lipids with a 95% theoretical conversion yield of glucose into lipids.

Table I.3. Substrates used for lipid accumulation by oleaginous yeasts and fungi.

Reference	Substrate	Microorganism
(Arzumanov, *et al*, 2000),	Ethanol	*Yarrowia lipolytica*
(Ykema, *et al*, 1988),	Whey permeate	*Aprotichum curvatus*
(Nicaud *et al*., 2002)	Fatty acids	*Yarrowia lipolytica*
(P Fickers *et al*., 2005)	Fatty acids	*Yarrowia lipolytica*
(Papanikolaou and Aggelis, 2002, 2003b, 2011)	Glycerol co-substrate	*Yarrowia lipolytica*
(Zhu *et al*., 2008)	Molasses	*Trichosporon fermentans*
(Angerbauer *et al*., 2008)	Pectin	*Lipomyces starkeyi*
(Angerbauer *et al*., 2008)	Sewage sledge	*Lipomyces starkeyi*
(Papanikolaou *et al*., 2007)	Starch	*Mortierella isabelina*
(Papanikolaou *et al*., 2006).	Glycerol/Stearin mixtures	*Yarrowia lipolytica*
(Babau *et al*., 2013)	Glucose/Stearin Glucose/Xylose	*Rhodotorola glutinis*

Ethanol (Arzumanov *et al*., 2000), fatty acids (Fickers *et al*., 2005; Nicaud *et al*., 2002) and glycerol (Papanikolaou and Aggelis, 2003a, 2002) were studied for the cultivation of *Yarrowia lipolytica* aiming at producing citric acid and lipids. Considering co-substrates, glucose and glycerol were used in the cultivation of *Yarrowia lipolytica*, which improved lipid content in batch and continuous cultures (Papanikolaou and Aggelis, 2011b, 2003a, 2002). Other mixtures like glucose/stearin and glycerol/stearin were studied in the same yeast, where the enrichment of the stored lipids concentration with unsaturated fatty acids was observed (Papanikolaou *et al*., 2006).

From the reported works, it is straightforward to see that the oleaginous yeast *Yarrowia lipolytica* is an interesting microorganism that can use a great variety of substrates to produce a moderate lipid content with a similar profile to biodiesel (Beopoulos et al., 2009a, 2009b; F. A. G. Gonçalves et al., 2014). More of the characteristics that made this yeast a promising option for lipid accumulation are reviewed in the next subsection.

I.1. *Yarrowia lipolytica*

The oleaginous yeast *Yarrowia lipolytica*; formerly known as *Candida*, *Endomucopsis,* or *Sacchamomycopsis lipolytica* (Barth and Gaillardin, 1997) belongs to the family *Hemiascomycetes*. This yeast was isolated from dairy products such as cheese, yogurt, and sausages, (Barth and Gaillardin, 1996) and oil polluted media or marine and hypersaline environments (Beopoulos *et al*., 2009a). This yeast is considered as oleaginous due to the existence of the enzyme ATP-citrate lyase (Ratledge, 2004; Ratledge and Wynn, 2002).

Yarrowia lipolytica is strictly aerobic and unable to grow above 32°C (Patrick Fickers *et al*., 2005). Depending on growth conditions this yeast is polymorphic, since it can develop the forms of single cells, pseudo-mycelia, or even true mycelia, with the microbial morphology being substantially influenced by environmental and nutritional factors (Papanikolaou and Aggelis, 2010).

The yeast *Yarrowia lipolytica* is of great interest due to its high proportion of accumulated linoleic acid, and its ability to assimilate different substrates such as glucose, galactose and mannitol, alkanes, alkenes, ethanol, acetate, glycerol (Barth and Gaillardin, 1997; Beopoulos *et al*., 2009a; Finogenova *et al*., 2005; Jimenez-Bremont *et al*., 2001; Papanikolaou *et al*., 2009), sucrose (Lazar *et al*., 2011), and industrial by-products such as molasses (Rakicka *et al*., 2015).

With regards on the biotechnological applications of *Yarrowia lipolytica* (Coelho *et al*., 2011; Liu *et al*., 2015), this yeast can produce a great quantity of citric acid (Anastassiadis et al., 2002a), isocitric acid, ketoglutaric acid and pyruvic acids triggered by the exhaustion of some nutrients (*i.e.* nitrogen, phosphorus, magnesium) (Finogenova *et al*., 2005). Furthermore, *Yarrowia lipolytica* allows the production of lipases, which are applied to manufacture detergents, optically active compounds for the pharmaceutical industry (L-alanine) (Akın *et al*., 2003). Moreover, it has been used to produce aromatic compounds (Coelho *et al*., 2011) from the production of γ-decalactone. *Yarrowia lipolytica* has been also used for soil bioremediation and polluted marine environments as it can degrade hydrophobic substrates, (Zinjarde and Pant, 2002). Recently, special interest has been paid to the production of single cell oils for different application such as cacao-butter substitute (Papanikolaou *et al*., 2003) or precursors of biofuels (Cescut, 2009; Ochoa-Estopier, 2012).

II. A GLANCE TO LIPID METABOLISM

II.1. Lipids

Lipids are macromolecules constituted of carbon, hydrogen, oxygen, and in some cases also phosphor. These macromolecules have a polar head and a non-polar hydrophobic tail (Figure I.1). Due to this characteristic, lipids are materials sparingly soluble or insoluble in water, but soluble in selected organic solvents such as benzene, chloroform, diethyl ether, hexane, and methanol. Together with carbohydrates and proteins, lipids constitute the principal structural components of cells (Belitz *et al*., 2009).

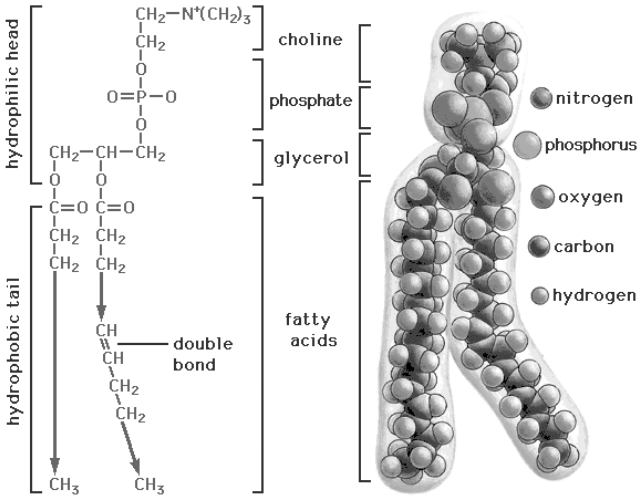


Figure I.1. Lipid general structure (Biochembayern, 2013)

Lipids play important roles and can be separated according to their cellular functions into:

- Structural: they constitute several structures in the cell and participate like barriers between the inner and outer parts of the cells to modulate the transport of amino acids.
- Accumulated: they are stored as carbon sources for the cell mainly in the form of Triacylglycerol (TAGs) and Free Fatty Acids (FFA) (Sorger and Daum, 2003).

There exists several classifications of lipids, but two which are of major interest are described hereafter.

II.1.1. Fatty Acids

Fatty acids (FAs) are groups of aliphatic chains with hydrophobicity. Fatty acids are normally activated in the form of acyl-CoA (Akoh and Min, 2008). In living organisms, FAs are found in paired number of carbons which can be separated into two main groups:

- Saturated: carbon bonds are single for long and short chains.
- Unsaturated: fatty acids with double carbon bonds. They are classified according to the number of bonds into monounsaturated (one bond) and polyunsaturated.

The unsaturated fatty acids allow the transport of electrons for oxidative phosphorylation. The nomenclature of the most common fatty acids produced by yeasts is presented in Table I.4.

Table I.4. Fatty acids nomenclature (Gunstone *et al*., 2007)**.**

Common name	Nomenclature	Chemical name
Palmitic acid	16:0	hexadecanoic acid
Stearic acid	18:0	octadecanoic acid
Oleic acid	18:1 (n-6)	9-octadecenoic acid
Linoleic acid	18:2 (n-6)	9,12-octadecadienoic acid
γ-linolenic acid	18:3 (n-6)	6,9,12-octadecatrienoic acid

II.1.2. Triacylglycerol

Triacylglycerols are molecules of glycerides where the three groups of hydroxyl glycerol are esterified by fatty acids (Figure I.2). They can constitute more than 15% of total structural lipids, which are found in the form of glycerophospholipids, sphingolipids, and glycosphingolipids (Coleman, 2004).

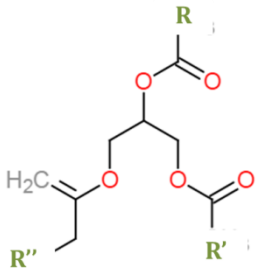


Figure I.2. General form of a molecule of triacylglycerol.

II.1.3. Phospholipids

Phospholipids are composed by two fatty acids, and one phosphate group esterified by a glycerol molecule. The phosphate group is hydrophilic (head), whereas the fatty acids are hydrophobic (tails) (Belitz *et al*., 2009).

Phospholipids are major components of cell membranes. Moreover, they form a lipid bilayer when hydrophobic tails line up against one another forming a membrane of hydrophilic heads on both sides. The lipid bilayer is semi-permeable and allows the diffusion of molecules (*e.g.* proteins) through the membrane.

II.2. Fatty acid synthesis

In yeasts, fatty acid synthesis (FAS) is triggered by the depletion of a nutrient in the medium (*e.g.* nitrogen, phosphorus, sulfate) (Cescut, 2009; Ochoa-Estopier, 2012; Papanikolaou and Aggelis, 2011b; Wu et al., 2010). Fatty acid synthesis is performed by two different pathways: a) "de novo" and b) "ex novo" synthesis. The "de novo" synthesis is in charge of the formation of precursors to lead lipid accumulation. Reaction (I.1) and (I.2) are catalyzed by the enzymes ATP Citrate Lyase (ACL), and Acetyl-CoA Carboxylase (ACC) respectively.

$$\text{Citrate} + \text{HS-CoA} + \text{ATP} \Rightarrow \text{Acetyl-CoA} + \text{oxaloacetate} + \text{ADP} + \text{Pi} \quad \text{(I.1)}$$

$$\text{Acetyl-CoA} + HCO_3^- + \text{ATP} \Rightarrow \text{Malonyl-CoA} + \text{ADP} + \text{Pi} \quad \text{(I.2)}$$

Before entering the FAS cycle, the acyl carrier CoA is converted into the acyl carrier protein (ACP), and thus the FAS cycle begins from Acetyl-ACP and Malonyl-ACP. The Malonyl-ACP

serves as a two carbon donor in a cyclic series of four sequential reactions catalyzed by fatty acid synthases and elongases. In yeasts, fatty acid synthase is composed of two subunits Fas1 (beta) and Fas2 (alpha) (Ratledge, 2004). Fas1 harbors the action of those sequential reactions catalyzed by the enzymes: β-ketoacyl synthase, β-ketoacyl reductase, β-hydroxyacyl-ACP or Enoyl dehydratase, and Enoyl-ACP reductase. The cyclic reactions involve 2-carbon addition in the termination of the overall reaction until the formation of C16 (palmate) as shown in Figure I.3. Summarizing, the synthesis of 1 mole of C16 fatty acid requires 14 mole of NADPH: 2 moles of NADPH are needed to reduce each 3-keto-fattyacyl group arising after every condensation reaction of Acetyl-ACP with Malonyl-ACP.

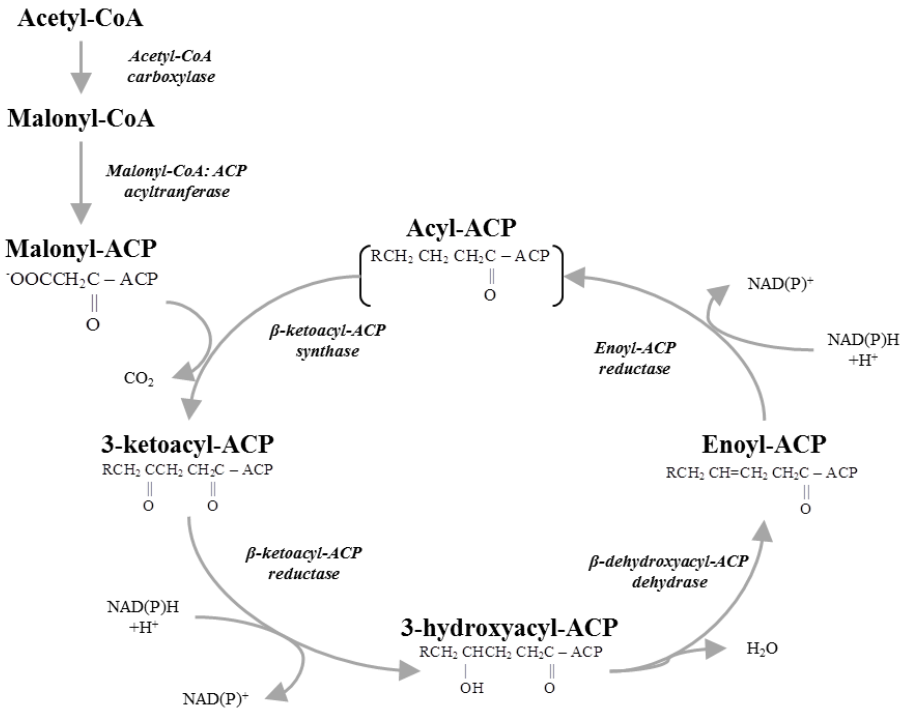


Figure I.3. Fatty acid synthesis (From Harwook 2010).

The "ex novo" synthesis adopts the incorporation of fatty acids from the culture medium to the cell through their degradation to be incorporated as Acetyl-CoA and to begin the FAS cycle again (Ochoa-Estopier, 2012).

II.3. Principal enzymes of fatty acid synthesis

During the lipid accumulation phase precursors (Acetyl-CoA, Malonyl CoA, and glycerol) and energy (ATP, NADPH) are required. Those are proportionated by some enzymes that play a key role in lipid accumulation.

II.3.1. AMP Deaminase

After nitrogen depletion (limitation or deficiency), oleaginous microorganisms show an increase of the enzyme AMP deaminase (Ratledge and Wynn, 2002), which cleaves the reaction:

$$AMP \Rightarrow IMP + NH_4^+ \quad \text{(I.3)}$$

As the ammonium concentration increases, the decrease of AMP inhibits isocitrate dehydrogenase, blocking the citric acid cycle on isocitrate.

II.3.2. ATP Citrate Lyase

Isocitrate cannot be metabolized due to the inhibition of isocitrate dehydrogenase and citrate is therefore accumulated. When the concentration of citric acid becomes higher than a critical value, it is secreted into the cytosol in exchange of incoming malate into the mitochondrion (Figure I.4). Citrate is used by the enzyme ATP Citrate Lyase (ACL) that catalyzes reaction (I.1) allowing the generation of Acetyl-CoA units, which are the precursors of fatty acid synthesis. This enzyme is characteristic of oleaginous microorganism. If the enzyme is absent the cells will be unable to accumulate lipids and will be therefore non-oleaginous (Ratledge and Wynn, 2002). Citric acid transport to cytosol is cleaved by ACL into Acetyl-CoA and Oxaloacetate. This Acetyl-CoA is used for FAS, whilst Oxaloacetate is converted into malate which enters the mitochondria to be included in the citric acid cycle.

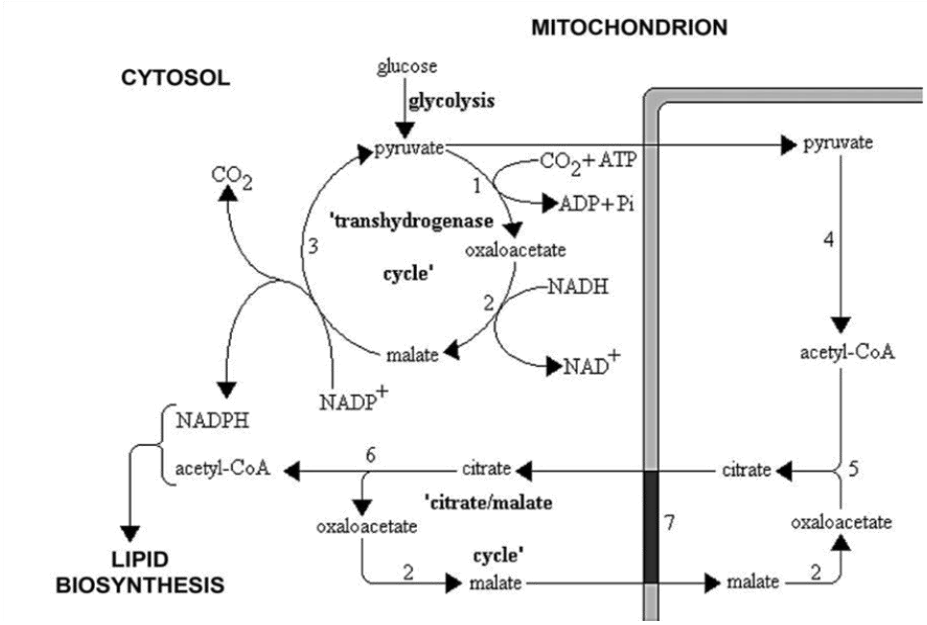


Figure I.4. Citrate/Malate cycle (Ratledge, 2004)**.**

II.3.3. Malic Enzyme

The major supplier of NADPH is the malic enzyme (Reaction 3, Figure I.5).

$$\text{Malate} + NADP^+ \Rightarrow \text{Pyruvate} + CO_2 + NADPH \quad \text{(I.4)}$$

Only this enzyme can provide the NADPH that is needed for fatty acid biosynthesis, and it is thus vital in the metabolism of lipid accumulation (Ratledge, 2004).

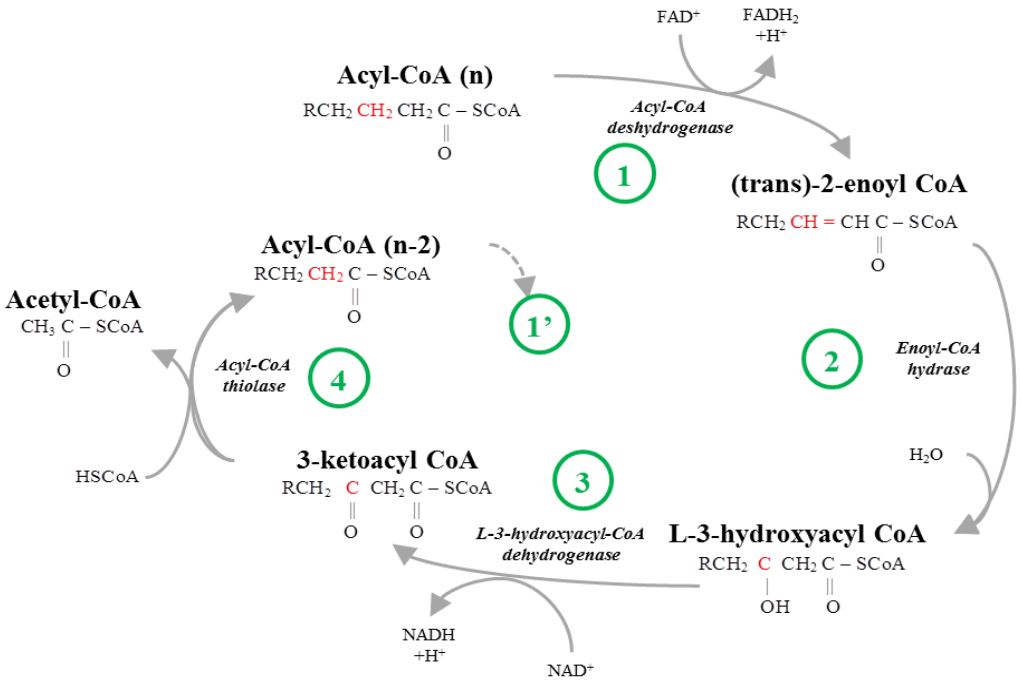


Figure I.5. β-oxidation of fatty acids (Paselk, 2006).

II.4. Elongation and desaturation of Fatty acids

The fatty acids cycle, which adds 2 carbons per cycle, normally ends by the formation of palmitic or stearic acids (C16:0 and C18:0, respectively). The elongation of these fatty acids occurs in the endoplasmic reticulum, where the same FAS cycle reactions are executed (Tehlivets *et al*., 2007). However, the fatty acids from FAS cycle can also be modified by the action of desaturases, which add a double bond in carbons Δ9, Δ12, and so on as displayed in Figure I.6.

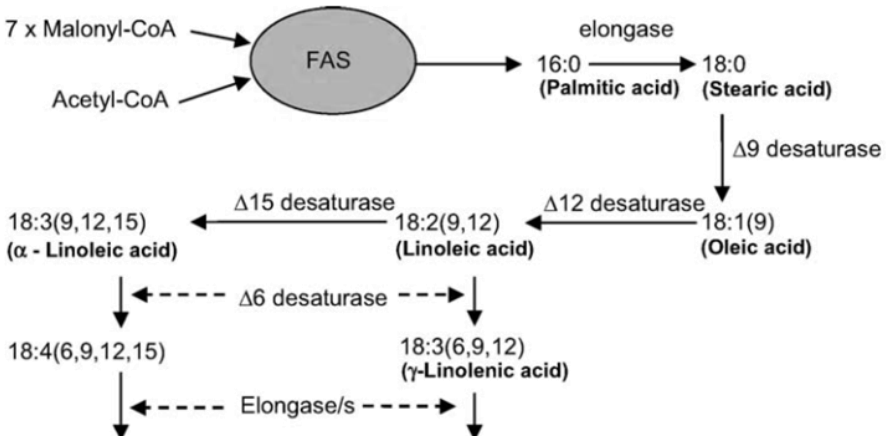


Figure I.6. Desaturation of fatty acids (Ratledge, 2004)**.**

The length of unsaturated fatty acids in oleaginous yeasts is often limited to palimitoleic acid (C16:2) and linoleic acid (C18:2) (see Table I.2).

II.5. Synthesis of TAG

Fatty acids, resulted from FAS and elongation, are then esterified with a molecule of glycerol-3-phosphate into TAG by the action of several acyl transferases. In order to form TAG some precursors are required such as phosphatidic acid (PA) and diacylglycerols (DAG) (Sorger and Daum, 2003).

In yeasts, phosphatidic acid (PA) is mostly formed from the acylation of glycerol-3-phosphate or by the reduction of dihydroxyacetone phosphate. Once PA is formed, it is dephosphatized by the enzyme phosphatidic acid phosphatase (PAP) to release one molecule of DAG and one organic phosphor (Figure I.7) (Coleman, 2004). These DAG are converted into TAG in another acylation step utilizing acyl donors, which can be acyl-CoA, phosphatidylcholine, or fatty acids (Sorger and Daum, 2003). Nonetheless, when metabolism is directed towards lipid accumulation, fatty acids are the main acyl donors which are esterified to lysophosphatidic acid, phosphatidic acid, and DAG successively (Cescut, 2009).

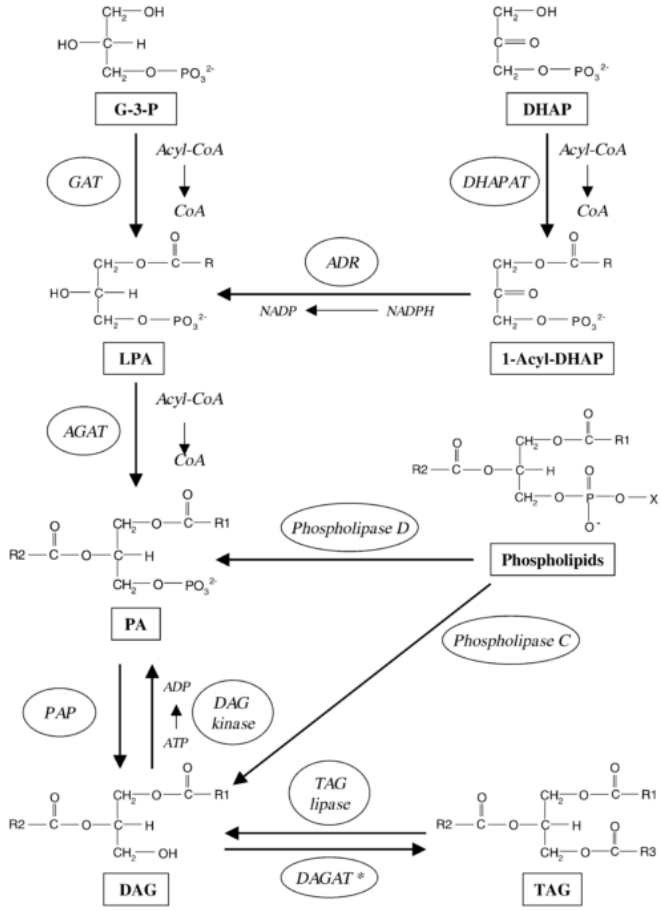


Figure I.7. Formation of Triacylglycerol (Sorger and Daum, 2003)**.**

II.6. Fatty acid degradation

Degradation of fatty acids is performed though the β-oxidation pathway (Schulz and Kunau, 1987), which allows the formation of acetyl-CoA units from pair chains of fatty acids by four successive reactions (Figure I.5).

1. Dehydrogenation by FAD by the enzyme acyl-CoA dehydrogenase. This enzyme catalyzes the formation of a double bond between carbon C2 and C3 generating a trans-Δ^2-enoyl-CoA.
2. Hydration. The double carbon bond is hydrated to form the isomer L-β-hydroxyacyl CoA by the action of the enoyl-CoA hydratase.
3. Oxidation by NAD^+. In this step, the hydroxyl group is converted into a β-ketoacyl CoA.

4. Thiolysis. The β-ketoacyl CoA is cleaved by the thiol group of another molecule of CoA. The thiol is inserted between C2 and C3 to generate an acetyl-CoA molecule, and an acyl-CoA two carbons shorter.

One molecule of NADH and one molecule of $FADH_2$ are formed for every produced molecule of acyl-CoA. In the case of oleaginous yeasts, β-oxidation is performed in the peroxisome.

II.7. Reduced metabolic network or reduced metabolic network

The metabolism of *Yarrowia lipolytica* has been studied using *in silico* models (1143 reactions) (Pan and Hua, 2012) and genomic models (2002 reactions) (Loira *et al*., 2012). Figure I.8 presents the reduced metabolic network to be used in this work, which takes into account: several data bases (KEEG, MetaCyc), previous experimental results (Cescut, 2009; Ochoa-Estopier, 2012), and the *in silico* models (Loira *et al*., 2012; Pan and Hua, 2012). In this schema only two compartments are considered (cytosol and mitochondria) to eliminate transport reactions. Furthermore, some linear pathways were considered into one general reaction that includes the actions of the four enzymes for fatty acid synthesis described along this section (Ratledge, 2004) and reduces those pathways into the global reaction:

$$\text{Acetyl-ACP} + 14\ \text{NADPH} + 7\ \text{Malonyl-ACP} \Rightarrow \text{palmitate-ACP} + 14\ \text{NADP} + 7\ \text{CO2} \quad \text{(I.5)}$$

The reduced metabolic network (Figure I.8) consists of 59 reactions with 51 metabolites. The principal interest of this reduced metabolic network is its use on structured models, whose description is presented in CHAPTER III, section III.

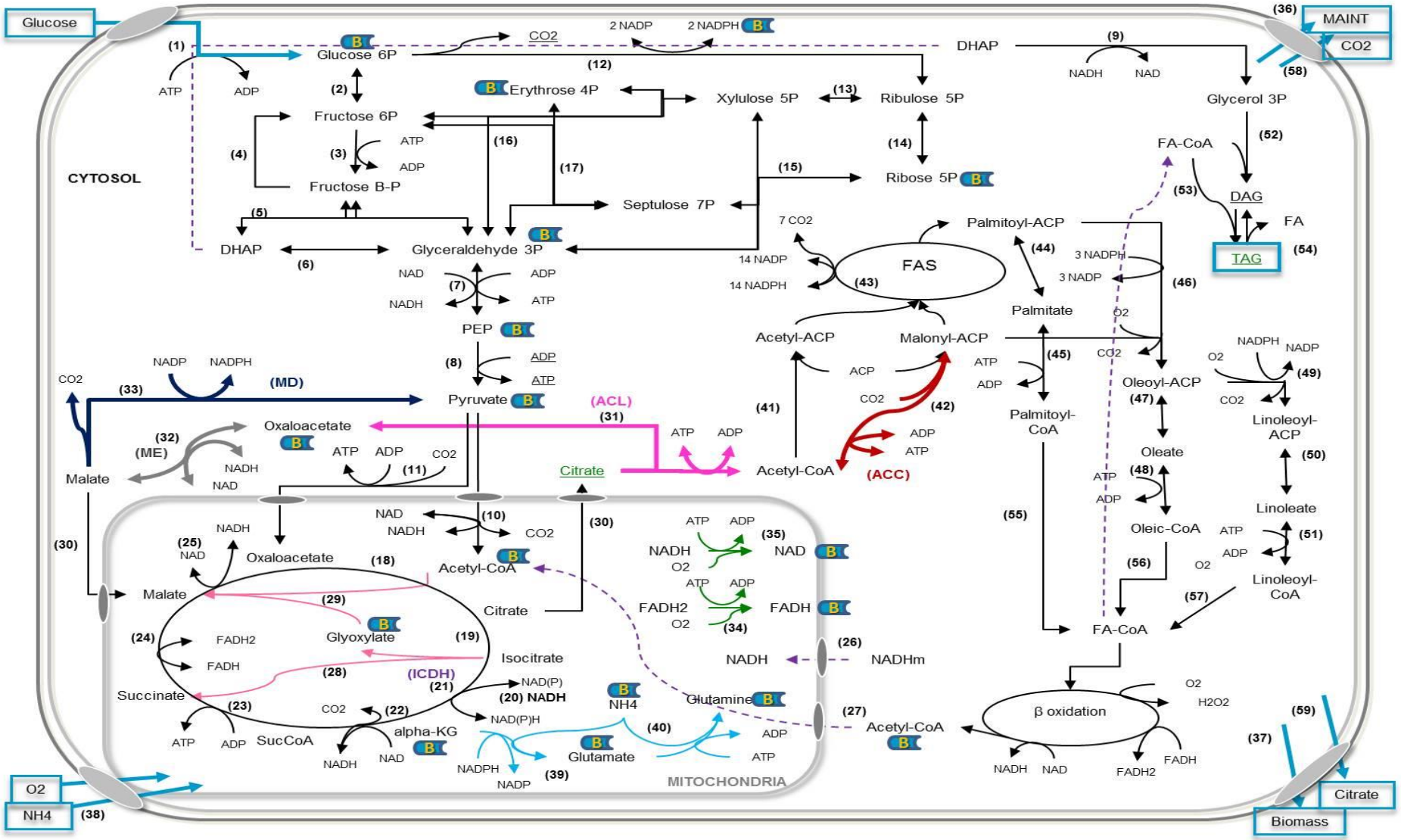


Figure I.8. Reduced Metabolism of *Yarrowia lipolytica*

III. GENERAL ASPECTS OF LIPID ACCUMULATION BIOPROCESSES

As mention above, lipid accumulation is triggered by the depletion of a nutrient in the medium (*e.g.* nitrogen, phosphorus, sulfate) (Anastassiadis et al., 2002a; Cescut, 2009; Ochoa-Estopier, 2012; Walther et al., 2005; Wu et al., 2010). The most studied nutrient depletion is nitrogen (Beopoulos et al., 2009a; Papanikolaou and Aggelis, 2011a; Ratledge and Cohen, 2008). When nitrogen is depleted, the activity of the enzyme phosphofructokinase is inhibited leading to the accumulation of cytosolic citrate, which is the precursor of the Acetyl-CoA in reaction (I.1) to produce storage lipids (Song *et al*., 2001). Additionally to the deprivation of nitrogen, lipid accumulation is affected by other factors (Beopoulos *et al*., 2009a; Gonçalves *et al*., 2014; Ratledge and Wynn, 2002; Sitepu *et al*., 2014), which will be described in the following subsections.

III.1. Factors influencing lipid accumulation

III.1.1. pH

Several studies have shown that oleaginous yeast can accumulate lipids in a range of pH between 3.5 to 7.5 with small changes in the quantity of accumulated lipids (Ratledge, 2004). The mould *Trichosporon fermentans* reported higher biomass concentration at pH 6.5, where the lipid content was also substantially higher (Zhu *et al*., 2008) when growing on cane molasses. For *Yarrowia lipolytica*, Moeller *et al*., (2007) demonstrated that this yeast grows faster as the pH is low (pH 4). At pH 7, the growth rate of *Yarrowia* is slower, and its growth is inhibited at pH 8.

According to (Papanikolaou et al., 2009), keeping pH in a range between 5.2 and 6.2 could enhance higher production of biomass, lipids, and citric acid in *Yarrowia lipolytica*. Based on the given range, the experiments were set within these values at a pH around 5.5.

III.1.2. Temperature

A decrease on the temperature favored the presence of unsaturated fatty acids. This was observed when growth temperature was reduced from 25 to 10 °C resulted in an increase of total lipid as well as an increase in unsaturated fatty acids for *Yarrowia lipolytica* (Kates and Baxter, 1962). However, low temperatures are not desirable when working on production scale because keeping low or high temperatures will increase the production cost. In the case of lipase production by *Yarrowia lipolytica*, the optimal range was set between 28 and 30 °C (Fontanille et al., 2012; Papanikolaou et al., 2001; Papanikolaou and Aggelis, 2002). Regarding growth of *Yarrowia lipolytica* from glucose, the optimal temperature was found in the rage of 30 – 34°C (Moeller *et al*., 2007) when the maximum growth rate was 0.132 h^{-1}. At 37 °C and 38 °C, no growth was observed. A temperature between 28 – 32 °C was optimal for citric acid production from glucose (Moeller *et al*., 2007).

For the growth of *Thricosporon fermentans* on cane molasses (Zhu *et al*., 2008), biomass concentration and lipid content reached their maxima at 27.5 $g.L^{-1}$ and at 60.6% , respectively, at 25 °C. Moreover, in the range from 20 to 35 °C, the total amount of unsaturated fatty acids decreased from 71.8% to 52.0%, indicating that a low temperature was favorable for the formation of unsaturated fatty acids. This was also reported for the mould *Entomophthora exitalis* (Kendrick and Ratledge, 1992), where the abatement of temperature in a range of 10°C favored the formation of long chain PUFAs, while the saturated fatty acids slightly decreased.

Regarding lipid profile, this was found not altered with a change in growth temperature, which limits the opportunities to manage the fatty acid by changes in the temperature. Nonetheless, when temperature was changed in the lipid accumulation phase, the maximum lipid content (19.1%) was achieved at 21 °C for the fungus *Pythium irregular*. This value decreased to 12.9% and 17% when temperatures were of 14 and 28 °C, respectively (Cantrell and Walker, 2009).

III.1.3. Oxygen

Fatty acids are less saturated when oxygen is diminished (Coelho *et al*., 2011). This can be explained from the metabolism, where oxygen is needed to convert palmitic acid to stearic acid (C16:0 – C18:0) and the desaturation of stearic acid to oleic acid and further to linoleic acid.

This was observed by (Davies *et al*., 1990), where the oleic acid increased from 42% to 55% as the oxygen uptake rated increased from 2.3 to 10 $mmol.L^{-1}.h^{-1}$.

Other authors showed that at a constant temperature and growth rate, oxygen depletion has no significant effect on lipid composition of *Entomophthora exitalis* (Kendrick and Ratledge, 1992).

III.1.4. Carbon source

The carbon source can alter the composition and the content of accumulated lipids (F. a. G. Gonçalves et al., 2014). Strains able to convert a higher diversity of sugars are preferable (Sitepu *et al*., 2014). As mentioned above, one of those microorganisms is *Yarrowia lipolytica*.

The content of triacylglycerol diminished, but the profile on oleic acid in the cell increased (Athenstaedt *et al*., 2006) when *Yarrowia lipolytica* was cultivated on oleic acid. *Yarrowia lipolytica* ACA-DC 50109 growing on raw glycerol achieved 43% of lipid content (Papanikolaou and Aggelis, 2009), whilst on glucose it attained 23% (Cescut, 2009). The opposite was observed for the production of citric acid, when this was produced on higher quantities when growing from glucose (Papanikolaou et al., 2009; Papanikolaou and Aggelis, 2003a).

The mixture of different carbon sources, such as glucose/stearin, have also been reported (Papanikolaou *et al*., 2006). In this work, glucose was fed first, where the production of citric acid was favored. When stearin was added at low concentrations, a decrease on the reserve lipid in *Yarrowia lipolytica* occurred. On the other hand, biomass yield presented notably high values on fat-substrate, while the effects on citric acid yield were neglected. Moreover, in mixtures *Yarrowia lipolytica* had a preference to assimilate glycerol on mixtures of glycerol, glucose, oleic acid and stearin.

A particularly marked difference was seen when comparing Crabtree negative and Crabtree positive yeasts. Crabtree positive (respiratory deficient) yeasts such as *Saccharomyces cerevisiae* and *Yarrowia lipolytica* are subject to catabolic repression, but Crabtree negative (respiratory sufficient) yeasts such as *Rhodotorula glutinis, Candida utilis* and *Pichia stipitis*

are not. It has been known for quite some time that while Crabtree positive yeasts produce less lipid under high glucose concentrations, Crabtree negative species accumulate higher total lipids (Sitepu *et al*., 2014).

III.1.5. Nitrogen

Several studies showed that nitrogen depletion induced lipid accumulation (Pomraning *et al*., 2016). Sitepu *et al*., (2014) regarded three cases in continuous cultures: two with constant nitrogen additions with a different final culture time, and other were a shift from low nitrogen to nitrogen-free medium was done after three days. In their results, lipid content was higher for the last culture where cultures where nitrogen free. Cescut, (2009) reported a critical concentration of nitrogen of 10^{-3} mol.L^{-1} for lipid induction in *Yarrowia lipolytica*. When lower values than the critical were achieved, the yeast tended to produce by-products.

The type of nitrogen source can impact the accumulation of lipids in *Yarrowia lipolytica* (Fickers *et al*., 2005). This was observed by Pereira-Meirelles *et al*. (1997), where peptone favored lipase activity over urea and ammonium sulfate when growing on olive oil as carbon source. Moreover, a pH decrease was always observed when urea and ammonium sulfate were used. Nitrogen source impact was also studied on the fungus *Thricosporon fermentans* (Zhu *et al*., 2008), where peptone reported the best lipid contents (54.9%). Disappointingly, $(NH_4)_2SO_4$, NH_4Cl and NH_4NO_3 gave low biomass concentration and lipid content.

III.1.6. C/N ratio

The most studied factor to modulate the conversion yield and lipid productivity is the C/N feeding ratio (Beopoulos *et al*., 2009; Cescut, 2009; Ochoa-Estopier and Guillouet, 2014; Sitepu *et al*., 2014; Ykema *et al*., 1986), which depends on the carbon and nitrogen flow rates. Oleaginous yeasts accumulate more lipids as the C/N ratio rises to a certain point. For example in *Rhodotorula glutinis*, glucose to lipid conversion yield increased from 0.25 to 0.4 Cmol.$Cmol^{-1}$ as C/N increases from 150 to 350. However, at 350, the severe nitrogen exhaustion led to cell death (Beopoulos *et al*., 2009a).

At a constant carbon concentration, nitrogen containing compounds tended to be depleted in cells grown under nitrogen limited medium (C/N=10) for *Yarrowia lipolytica*, whilst non-nitrogen containing carbohydrates, alcohols and acids were presented under higher nitrogen limited levels (C/N=150) (Pomraning *et al*., 2016). Moreover, Ochoa-Estopier and Guillouet, (2014) identified an optimal inverse range N/C from 0.021 to 0.084 mol*N*.Cmol^{-1} for the accumulation of lipids without the concomitantly production of citric acid. Therefore, the design of cultures must take into account these values.

Furthermore, (Beopoulos *et al*., 2009a) reported a manner to calculate the maximum suitable C/N ratio for lipid accumulation estimated from the theoretical biomass to substrate yield ($Y_{X/S}$) and the nitrogen concentration content in the biomass (N_X) (mol*N*.Cmol^{-1}) as $Y_{X/S} / N_X$, which could be obtained from the biomass formula.

III.2. Culture modes

The design of bioprocesses for lipid accumulation should be conceived on the basis of regulating the factors reported in last section.

Different culture modes of *Yarrowia lipolytica* have been carried out under nitrogen depletion conditions (limitation and deficiency) for lipid accumulation by monitoring and/or controlling the C/N ratio. Table I.5 summarizes the performance of different strains of *Yarrowia lipolytica* in different culture modes. The highest results for productivity and lipid content were obtained in continuous cultures on glycerol (Rakicka *et al*., 2015). However, the same substrate gave the worst performance when operated in batch mode (Papanikolaou and Aggelis, 2002). These contrasting differences are related to the culture mode, but also to the strain. In order to have some insights about it, the used culture modes are described in the following subsections.

Table I.5. Performance of different culture conditions on lipid accumulation by *Yarrowia lipolytica.*

Culture mode	Strain	Substrate	X ($g.L^{-1}$)	Volumetric Productivity ($g.(L\text{-}h)^{-1}$)	Lipid content ($g.g^{-1}$)	Reference
Batch	LGAM S(7)1	Glycerol	7.1	0.001	0.05 - 0.1	(Papanikolaou and Aggelis, 2002)
	LGAM S(7)1	Stearin/oleic acid	7.2	0.06	0.4	(Papanikolaou and Aggelis, 2003b)
	LGAM S(7)1	Glycerol/stearin	11.4	0.034	0.30	(Papanikolaou *et al*., 2003)
	LGAM S(7)1	Glucose/stearin	11.3	0.02	0.22	(Papanikolaou *et al*., 2003)
	LGAM S(7)1	Glucose/stearin	8.32	0.022	0.19	(Papanikolaou *et al*., 2006)
Continuous	LGAM S(7)1	Industrial Glycerol	8.1	0.11	0.4	(Papanikolaou and Aggelis, 2002)
	LGAM S(7)1	Glucose	9.3	0.08	0.25	(Aggelis and Komaitis, 1999)
	JMY4086	Glycerol	60	0.31	0.46	(Rakicka *et al*., 2015)
D-Stat	W29	Glucose	9.8	0.0008	0.16	(Ochoa-Estopier and Guillouet, 2014)
Repeated batch	LGAM S(7)1	Glycerol	5	0.021	0.223	(Makri *et al*., 2010)
Fed-Batch	W29	Glucose	73.1	0.27	0.25	(Cescut, 2009)
	W29	Glucose	50.26	0.10	0.17	(Ochoa-Estopier, 2012)
	JMY4086	Glycerol	50	0.18	0.31	(Rakicka *et al*., 2015)

III.2.1. Batch mode cultures

This type of culture is large used in the industry to guarantee a pure culture (Dochain, 2003). Batch cultures are mainly used to evaluate the growth of microorganisms on different substrates. They have the advantage of being easy to perform and could produce large volumes of experimental data in short periods of time. However, the optimization of the production of specific metabolites is difficult to achieve due to the lack of control on the substrate feeding (Stephanopoulos *et al*., 1998). In the case of lipid accumulation by *Yarrowia lipolytica*, these

cultures aimed at studying the microbial behavior on the given substrates (Papanikolaou and Aggelis, 2003b, 2002). Those cultures started with a known value of C/N ratio that tends to infinite as nitrogen is being depleted enhancing lipid accumulation. Lipid to biomass yield of 0.3 – 0.4 $g.g^{-1}$ were obtained after 40 hours of culture, where lipids were further consumed as the carbon source was been depleted (Papanikolaou and Aggelis, 2002). The lack of C/N regulation makes difficult to achieve high productivity of lipid accumulation in batch cultures. In such conditions, lipid accumulation was followed by citric acid production (Papanikolaou *et al*., 2006). These findings highlight the need to manage both carbon and nitrogen flows to optimize lipid accumulation and reduce citric acid production, which cannot be achieved on batch mode.

III.2.2. Continuous cultures

Continuous cultures allow the characterization of steady states for the determination of specific rates under well-defined environmental conditions (Adamberg *et al*., 2015). Here, the dilution rate is kept constant in order to the study the response to disturbances of a stable state. The drawback of using this type of culture is the laborious operation as large amounts of fresh, sterile medium have to be prepared and it requires long periods of time for achieving the steady state (Kasemets *et al*., 2003). Moreover, continuous cultures are rarely used in industrial processes because they are sensitive to contaminations. With a particular interest on lipid accumulation, the optimal dilution rate (*D*) is intrinsically linked to the optimal C/N ratio. Ykema *et al*., (1986) studied different C/N ratios by the regulation of substrate concentration. This led to a fine tuning of growth rate proving that lower growth rates promote more extensive lipid accumulation.

Papanikolaou and Aggelis, (2002), regarded the impact of dilution rates when *Yarrowia lipolytica* was grown on glycerol. The results showed that storage lipids were favored at low dilution rates, where the highest value (0.4 $g.g^{-1}$) was attained at D = 0.04 h^{-1}. On the other hand, when dilution rate was fixed to 0.01 h^{-1}, Rakicka *et al*., (2015) studied the effect of glycerol concentration, where lipid to biomass yield of 0.46 $g.g^{-1}$ was obtained with 250 $g.L^{-1}$ of glycerol.

Even if good results were obtained in continuous cultures, their scale up is more complicated due to the diverse conditions to take into account such as sterile conditions and product recovery, among others (Beopoulos et al., 2009a; Dochain, 2008). For this reason, Fed-batch cultures should be preferable aiming industrialization of the process.

III.2.3. D-stat cultures

The D-stat approach is very useful to study the impact of an environmental parameter on cell physiology at constant dilution rate *D*. The concept of the D-stat is simple: one environmental parameter (*e.g.* temperature, pH, or cultivation medium composition) is smoothly changed whilst *D* and other parameters are kept constant (Adamberg *et al*., 2015). This type of culture has been used for the study of: (i) the transition of metabolism from carbon to nitrogen limitation by smoothly decreasing the N/C ratio in *Saccharomyces cerevisiae* and *Saccharomyces urvarum* (Kasemets *et al*., 2003); (ii) the dependence on galactose/arginine ratio of lactic acid bacteria (Adamberg *et al*., 2006); and (iii) the overflow metabolism by acetate co-utilization with glucose on *Escherichia coli* (Valgepea *et al*., 2010).

Regarding lipid accumulation, the impact of the ratio C/N was evaluated to identify the metabolic shifts of *Yarrowia lipolytica* at constant dilution rate, pH = 5.5, and T = 28 °C (Ochoa-Estopier and Guillouet, 2014). In this case, the best N/C ratio in the cultivation medium for producing lipids without carbon loss into citric acid was identified between 0.021 and 0.085 mol*N*.Cmol^{-1}. Although the insights obtained from this culture mode, it is not suitable for lipid accumulation due to the low dilution rates.

III.2.4. Fed-batch cultures

This culture is the most common operation in industrial practice because it allows controlling substrate additions as well as the environmental conditions, and it facilitates physiological studies for a limited period of time (Beopoulos *et al*., 2009a; Stephanopoulos *et al*., 1998). In addition, this operating mode is more recommended when the recovery of the products is carried out at intervals (*e.g.* intracellular accumulation) (Dochain, 2008).

For lipid accumulation, it is desirable to have a precise control and monitoring of carbon and nitrogen flow rates since both induce activation-deactivation of the metabolic pathways (Papanikolaou and Aggelis, 2011b; Ykema et al., 1986). In addition, this type of culture allowed a well management of the N/C ratio to ensure a production of lipids without citric acid, which was shown by Ochoa-Estopier, (2012) for *Yarrowia lipolytica* on glucose. In fed-batch culture, if the nitrogen feeding is considered constant, and the carbon flow rate is varying, lipid accumulation can be generally seen as two metabolic phases (Beopoulos et al., 2009a; Papanikolaou and Aggelis, 2011b): (i) Growth phase, and (ii) lipid accumulation phase.

Phase (i) occurs when all nutrients are available in the medium, and carbon source (*i.e.* sugars) is consumed. Structural lipids are produced around 5-10% w/w (Papanikolaou and Aggelis, 2010). The biomass produced in this phase is called catalytic biomass (biomass which no contains accumulated lipids).

Phase (ii), is activated by a nutrient depletion, mostly nitrogen, and an excess on the carbon source. Under these excess conditions, the exceeding carbon is stored in lipid bodies mainly in the form of TAGs accumulated inside the cell (Beopoulos *et al*. 2011; Meng *et al*. 2009). Catalytic biomass can also increase due to the accumulation of non-lipid materials such as polysaccharides.

IV. CONCLUSIONS

This chapter presented a brief summary of oleaginous microorganisms, where the yeast *Yarrowia lipolytica* was highlighted for its ability to assimilate a great variety of substrates, its fully sequenced genome, and its products that could be of industrial interest. The principals of metabolism towards lipid accumulation were also addressed. These metabolic reactions allowed the construction of a reduced metabolic network for *Yarrowia lipolytica*. This was based on two compartments: cytosol and mitochondria. However, metabolic networks are expected to gain in accuracy to make some improvements particularly on the lipids, carbohydrates and degradation pathways.

The studies on fermentations by *Yarrowia lipolytica* reported that lipid accumulation is triggered by nutritional limitation. This could lead to the production of other by-products such

as citric acid. The production of the last was increased when this yeast was on cultures from glucose. The use of a co-substrate (*e.g.* glycerol, stearin) was efficient to redirect metabolism towards lipid accumulation. Nevertheless, it would be of great interest to study the dynamics of the most important metabolites to determine the metabolic shifts inducing citric acid production in order to optimize the carbon flux towards the production of lipids.

The factors influencing lipid accumulation were presented. Among these factors, the carbon source for this work was fixed to be glucose, pH and temperature were regulated, and thus the main focus was the N/C ratio. Additionally, different culture conditions were reviewed; where continuous cultures were the most effective in terms of lipid content ($gLIP.gX^{-1}$). Fed-batch cultures from glucose, for instance, were better to provide both good lipid productivities and content, which are the two important criteria to take into account with the aim of industrializing the production of lipids. Furthermore, fed-batch culture will allow the recovery of the cells accumulating lipids, the industrial control of contamination, and the controlling of the N/C ratio, reported as the most influencing factor of lipid accumulation.

V. REFERENCES

Paselk, R.A. 2006. Fat Metabolism. *Introductory Biochemistry*. Retrieved from http://users.humboldt.edu/rpaselk/C438.S07/C438Notes/C438nLec35.htm.

Biochembayern 2013. Structure of Lipids. Retrieved from https://biochembayern.wordpress.com/2013/04/14/structure-of-lipids/

Adamberg, K., Adamberg, S., Laht, T.-M., Ardö, Y., Paalme, T., 2006. Study of Cheese Associated Lactic Acid Bacteria Under Carbohydrate-Limited Conditions Using D-Stat Cultivation. Food Biotechnol. 20, 143–160. doi:10.1080/08905430600709412

Adamberg, K., Valgepea, K., Vilu, R., 2015. Advanced continuous cultivation methods for systems microbiology. Microbiol. (United Kingdom) 161, 1707–1719. doi:10.1099/mic.0.000146

Aggelis, G., Komaitis, M., 1999. Enhancement of single cell oil production by Yarrowia lipolytica growing in the presence of Teucrium polium L. aqueous extract. Biotechnol. Lett. 21, 747–749. doi:10.1023/A:1005591127592

Ahmad, F.B., Zhang, Z., Doherty, W.O.S., O'Hara, I.M., 2015. A multi-criteria analysis approach for ranking and selection of microorganisms for the production of oils for

biodiesel production. Bioresour. Technol. 190, 264–273. doi:10.1016/j.biortech.2015.04.083

Akın, N., Aydemir, S., Koçak, C., Yıldız, M.A., 2003. Changes of free fatty acid contents and sensory properties of white pickled cheese during ripening. Food Chem. 80, 77–83. doi:10.1016/S0308-8146(02)00242-X

Akoh, C.C., Min, D.B., 2008. Food Lipids: Chemistry, Nutrition, and Biotechnology, Third Edition. CRC Press.

Anastassiadis, S., Aivasidis, A., Wandrey, C., 2002. Citric acid production by Candida strains under intracellular nitrogen limitation. Appl. Microbiol. Biotechnol. 60, 81–7. doi:10.1007/s00253-002-1098-1

Angerbauer, C., Siebenhofer, M., Mittelbach, M., Guebitz, G.M., 2008. Conversion of sewage sludge into lipids by Lipomyces starkeyi for biodiesel production. Bioresour. Technol. 99, 3051–6. doi:10.1016/j.biortech.2007.06.045

Arzumanov, T.E., Sidorov, I.A., Shishkanova, N. V., Finogenova, T. V., 2000. Mathematical modeling of citric acid production by repeated batch culture. Enzyme Microb. Technol. 26, 826–833. doi:http://dx.doi.org/10.1016/S0141-0229(00)00178-2

Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J.-M., Chardot, T., 2006. Lipid particle composition of the yeast Yarrowia lipolytica depends on the carbon source. Proteomics 6, 1450–9. doi:10.1002/pmic.200500339

Babau, M., Cescut, J., Allouche, Y., Lombaert-Valot, I., Fillaudeau, L., Uribelarrea, J.-L., Molina-Jouve, C., 2013. Towards a Microbial Production of Fatty Acids as Precursors of Biokerosene from Glucose and Xylose. Oil Gas Sci. Technol. – Rev. d'IFP Energies Nouv. 68, 899–911. doi:10.2516/ogst/2013148

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. Fems Microbiol. Rev. 19, 219–237.

Barth, G., Gaillardin, C., 1996. Nonconventional yeasts in Biotechnology 313–388.

Belitz, H.-D., Grosch, W., Schieberle, P., 2009. Food Chemistry, 4th revise. ed. Springer.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009a. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Beopoulos, A., Chardot, T., Nicaud, J.-M., 2009b. Yarrowia lipolytica: A model and a tool to understand the mechanisms implicated in lipid accumulation. Biochimie 91, 692–696. doi:10.1016/j.biochi.2009.02.004

Beopoulos, A., Nicaud, J.-M., Gaillardin, C., 2011. An overview of lipid metabolism in yeasts and its impact on biotechnological processes. Appl. Microbiol. Biotechnol. 90, 1193–206. doi:10.1007/s00253-011-3212-8

biochembayern, n.d. Structure of Lipids | biochembayern on WordPress.com [WWW Document]. URL http://biochembayern.wordpress.com/2013/04/14/structure-of-lipids/ (accessed 10.21.14).

Cantrell, K.B., Walker, T.H., 2009. Influence of Temperature on Growth and Peak Oil Biosynthesis in a Carbon-Limited Medium by Pythium irregulare. J. Am. Oil Chem. Soc. 86, 791–797. doi:10.1007/s11746-009-1409-0

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Coelho, M.A.Z., Amaral, P.F.F., Belo, I., 2011. Yarrowia lipolytica : an industrial workhorse, Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology. Formatex.

Coleman, R., 2004. Enzymes of triacylglycerol synthesis and their regulation. Prog. Lipid Res. 43, 134–176. doi:10.1016/S0163-7827(03)00051-1

Davies, R.J., Holdsworth, J., Reader, S., 1990. The effect of low oxygen uptake rate on the fatty acid profile of the oleaginous yeast Apiotrichum curvatum. Appl. Microbiol. Biotechnol. 33. doi:10.1007/BF00172553

Dochain, D., 2008. Bioprocess Control, First edit. ed. John Wiley & Sons, Inc., London.

Dochain, D., 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. J. Process Control 13, 801–818. doi:10.1016/S0959-1524(03)00026-X

Fickers, P., Benetti, P.-H., Waché, Y., Marty, A., Mauersberger, S., Smit, M.S., Nicaud, J.-M., 2005. Hydrophobic substrate utilisation by the yeast Yarrowia lipolytica, and its potential applications. FEMS Yeast Res. 5, 527–43. doi:10.1016/j.femsyr.2004.09.004

Fickers, P., Fudalej, F., Nicaud, J.-M., Destain, J., Thonart, P., 2005. Selection of new over-producing derivatives for the improvement of extracellular lipase production by the non-conventional yeast Yarrowia lipolytica. J. Biotechnol. 115, 379–386. doi:10.1016/j.jbiotec.2004.09.014

Finogenova, T. V., Morgunov, I.G., Kamzolova, S. V., Chernyavskaya, O.G., 2005. Organic Acid Production by the Yeast Yarrowia lipolytica: A Review of Prospects. Appl. Biochem. Microbiol. 41, 418–425. doi:10.1007/s10438-005-0076-7

Fontanille, P., Kumar, V., Christophe, G., Nouaille, R., Larroche, C., 2012. Bioconversion of volatile fatty acids into lipids by the oleaginous yeast Yarrowia lipolytica. Bioresour. Technol. 114, 443–449. doi:10.1016/j.biortech.2012.02.091

Gonçalves, F. a. G., Colen, G., Takahashi, J. a, 2014. Yarrowia lipolytica and its multip le applications in the biotechnological industry. Sci. World J. 2014, 1–14. doi:in press

Gonçalves, F.A.G., Colen, G., Takahashi, J.A., 2014. Yarrowia lipolytica and Its Multiple Applications in the Biotechnological Industry. Sci. World J. 2014, 1–14. doi:10.1155/2014/476207

Gunstone, F.D., Harwood, J.L., Dijkstra, A.J., 2007. The Lipid Handbook, Third Edit. ed.

Harwook, J.L., 2010. The AOCS lipid Library [WWW Document]. Plant Fat. acid Synth. URL http://lipidlibrary.aocs.org/plantbio/fa_biosynth/index.htm

Huang, G., Chen, F., Wei, D., Zhang, X., Chen, G., 2010. Biodiesel production by microalgal biotechnology. Appl. Energy 87, 38–46. doi:10.1016/j.apenergy.2009.06.016

Jimenez-Bremont, J.F., Ruiz-Herrera, J., Dominguez, A., 2001. Disruption of gene YlODC reveals absolute requirement of polyamines for mycelial development in Yarrowia lipolytica. FEMS Yeast Res. 1, 195–204.

Kasemets, K., Drews, M., Nisamedtinov, I., Adamberg, K., Paalme, T., 2003. Modification of A-stat for the characterization of microorganisms. J. Microbiol. Methods 187–200. doi:10.1016/S0167-7012(03)00143-X

Kates, M., Baxter, R.M., 1962. Lipid composition of mesophilic and psychrophilic yeasts (Candida species) as influenced by environmental temperature. Can. J. Biochem. Physiol. 40, 1213–27.

Kendrick, A., Ratledge, C., 1992. Lipid formation in the oleaginous mould Entomophthora exitalis grown in continuous culture: effects of growth rate, temperature and dissolved oxygen tension on polyunsaturated fatty acids. Appl. Microbiol. Biotechnol. 37, 18–22. doi:10.1007/BF00174196

Lam, M.K., Lee, K.T., 2012. Microalgae biofuels: A critical review of issues, problems and the way forward. Biotechnol. Adv. 30, 673–90. doi:10.1016/j.biotechadv.2011.11.008

Lazar, Z., Walczak, E., Robak, M., 2011. Simultaneous production of citric acid and invertase by Yarrowia lipolytica SUC+ transformants. Bioresour. Technol. 102, 6982–6989. doi:10.1016/j.biortech.2011.04.032

Li, Q., Du, W., Liu, D., 2008. Perspectives of microbial oils for biodiesel production. Appl. Microbiol. Biotechnol. 80, 749–756. doi:10.1007/s00253-008-1625-9

Liu, H.-H., Ji, X.-J., Huang, H., 2015. Biotechnological applications of Yarrowia lipolytica: Past, present and future. Biotechnol. Adv. 33, 1522–1546. doi:10.1016/j.biotechadv.2015.07.010

Loira, N., Dulermo, T., Nicaud, J.-M., Sherman, D.J., 2012. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Syst. Biol. 6, 35. doi:10.1186/1752-0509-6-35

Makri, A., Fakas, S., Aggelis, G., 2010. Metabolic activities of biotechnological interest in Yarrowia lipolytica grown on glycerol in repeated batch cultures. Bioresour. Technol. 101, 2351–2358. doi:http://dx.doi.org/10.1016/j.biortech.2009.11.024

Meng, X., Xian, M., Xu, X., Zhang, L., Nie, Q., Yang, J., 2009. Biodiesel production from oleaginous microorganisms. Renew. Energy 34, 1–5. doi:10.1016/j.renene.2008.04.014

Moeller, L., Strehlitz, B., Aurich, A., Zehnsdorf, A., Bley, T., 2007. Optimization of Citric Acid Production from Glucose byYarrowia lipolytica. Eng. Life Sci. 7, 504–511. doi:10.1002/elsc.200620207

Moeller, L., Strehlitz, B., Aurich, A., Zehnsdort, A., Bley, T., 2007. Optimization of citric acid production from glucose by Yarrowia lipolytica. Eng. Life Sci. 7, 504–511. doi:10.1002/elsc.200620207

Nicaud, J.-M., Madzak, C., Broek, P., Gysler, C., Duboc, P., Niederberger, P., Gaillardin, C., 2002. Protein expression and secretion in the yeast Yarrowia lipolytica. FEMS Yeast Res. 2, 371–379. doi:10.1111/j.1567-1364.2002.tb00106.x

Ochoa-Estopier, A., 2012. Analyses systématique des bascules métaboliques chez les levures d'intérêt industriel: application aux bascules du métabolisme lipidique chez Yarrowia lipolytica. Institut National des Sciences Appliquées de Toulouse.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Pan, P., Hua, Q., 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS One 7, e51535. doi:10.1371/journal.pone.0051535

Papanikolaou, S., Aggelis, G., 2011a. Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. Eur. J. Lipid Sci. Technol. 113, 1031–1051. doi:10.1002/ejlt.201100014

Papanikolaou, S., Aggelis, G., 2011b. Lipids of oleaginous yeasts. Part II: Technology and

potential applications. Eur. J. Lipid Sci. Technol. 113, 1052–1073. doi:10.1002/ejlt.201100015

Papanikolaou, S., Aggelis, G., 2010. Yarrowia lipolytica: A model microorganism used for the production of tailor-made lipids. Eur. J. Lipid Sci. Technol. 112, 639–654. doi:10.1002/ejlt.200900197

Papanikolaou, S., Aggelis, G., 2009. Biotechnological valorization of biodiesel derived glycerol waste through production of single cell oil and citric acid by Yarrowia lipolytica. Lipid Technol. 21, 83–87. doi:10.1002/lite.200900017

Papanikolaou, S., Aggelis, G., 2003a. Modelling aspects of the biotechnological valorization of raw glycerol: production of citric acid by Yarrowia lipolytica and 1,3-propanediol by Clostridium butyricum. J. Chem. Technol. Biotechnol. 78, 542–547. doi:10.1002/jctb.831

Papanikolaou, S., Aggelis, G., 2003b. Modeling lipid accumulation and degradation in Yarrowia lipolytica cultivated on industrial fats. Curr. Microbiol. 46, 398–402. doi:10.1007/s00284-002-3907-2

Papanikolaou, S., Aggelis, G., 2002. Lipid production by Yarrowia lipolytica growing on industrial glycerol in a single-stage continuous culture. Bioresour. Technol. 82, 43–49.

Papanikolaou, S., Chatzifragkou, A., Fakas, S., Galiotou-Panayotou, M., Komaitis, M., Nicaud, J.-M.M., Aggelis, G., 2009. Biosynthesis of lipids and organic acids by Yarrowia lipolytica strains cultivated on glucose. Eur. J. Lipid Sci. Technol. 111, 1221–1232. doi:10.1002/ejlt.200900055

Papanikolaou, S., Chevalot, I., Komaitis, M., Aggelis, G., Marc, I., 2001. Kinetic profile of the cellular lipid composition in an oleaginous Yarrowia lipolytica capable of producing a cocoa-butter substitute from industrial fats. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 80, 215–224. doi:10.1023/A:1013083211405

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Papanikolaou, S., Galiotou-Panayotou, M., Fakas, S., Komaitis, M., Aggelis, G., 2007. Lipid production by oleaginous Mucorales cultivated on renewable carbon sources. Eur. J. Lipid Sci. Technol. 109, 1060–1070. doi:10.1002/ejlt.200700169

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., Marc, I., 2003. Accumulation of a cocoa-butter-like lipid by Yarrowia lipolytica cultivated on agro-industrial residues. Curr.

Microbiol. 46, 124–130. doi:10.1007/s00284-002-3833-3

Pereira-Meirelles, F.V., Rocha-Leão, M.H.M., Sant Anna Jr, G.L., 1997. A Stable Lipase from Candida lipolytica. Appl. Biochem. Biotechnol. 63-65, 73–85.

Pomraning, K.R., Kim, Y.-M., Nicora, C.D., Chu, R.K., Bredeweg, E.L., Purvine, S.O., Hu, D., Metz, T.O., Baker, S.E., 2016. Multi-omics analysis reveals regulators of the response to nitrogen limitation in Yarrowia lipolytica. BMC Genomics 17, 138. doi:10.1186/s12864-016-2471-2

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., Nicaud, J.M., 2015. Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. Biotechnol. Biofuels 8, 104. doi:10.1186/s13068-015-0286-z

Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. Biochimie 86, 807–15. doi:10.1016/j.biochi.2004.09.017

Ratledge, C., Cohen, Z., 2008. Microbial and algal oils: Do they have a future for biodiesel or as commodity oils? Lipid Technol. 20, 155–160. doi:10.1002/lite.200800044

Ratledge, C., Wynn, J.P., 2002. The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. Adv. Appl. Microbiol. 51, 1–51.

Schulz, H., Kunau, W.-H., 1987. Beta-oxidation of unsaturated fatty acids: a revised pathway. Trends Biochem. Sci. 12, 403–406. doi:10.1016/0968-0004(87)90196-4

Singh, A., Nigam, P.S., Murphy, J.D., 2011. Renewable fuels from algae: an answer to debatable land based fuels. Bioresour. Technol. 102, 10–6. doi:10.1016/j.biortech.2010.06.032

Sitepu, I., Selby, T., Lin, T., Zhu, S., Boundy-Mills, K., 2014. Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. J. Ind. Microbiol. Biotechnol. 41, 1061–70. doi:10.1007/s10295-014-1447-y

Sitepu, I.R., Garay, L.A., Sestric, R., Levin, D., Block, D.E., German, J.B., Boundy-Mills, K.L., 2014. Oleaginous yeasts for biodiesel: current and future trends in biology and production. Biotechnol. Adv. 32, 1336–60. doi:10.1016/j.biotechadv.2014.08.003

Song, Y., Wynn, J.P., Li, Y., Grantham, D., Ratledge, C., 2001. A pre-genetic study of the isoforms of malic enzyme associated with lipid accumulation in Mucor circinelloides. Microbiology 147, 1507–1515.

Sorger, D., Daum, G., 2003. Triacylglycerol biosynthesis in yeast. Appl. Microbiol. Biotechnol. 61, 289–99. doi:10.1007/s00253-002-1212-4

Stephanopoulos, G.G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles

and Methodologies, Metabolic Engineering. doi:10.1016/B978-0-12-666260-3.50019-4

Tehlivets, O., Scheuringer, K., Kohlwein, S.D., 2007. Fatty acid synthesis and elongation in yeast. Biochim. Biophys. Acta 1771, 255–70. doi:10.1016/j.bbalip.2006.07.004

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L., Vilu, R., 2010. Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Syst. Biol. 4, 166. doi:10.1186/1752-0509-4-166

Walther, T., Reinsch, H., Ostermann, K., Deutsch, A., Bley, T., 2005. Coordinated development of yeast colonies: An experimental analysis of the adaptation to different nutrient concentrations - Part 1. Eng. Life Sci. 5, 115–124.

Wu, S., Hu, C., Jin, G., Zhao, X., Zhao, Z.K., 2010. Phosphate-limitation mediated lipid production by Rhodosporidium toruloides. Bioresour. Technol. 101, 6124–9. doi:10.1016/j.biortech.2010.02.111

Ykema, A., Verbree, E.C., Van Verseveld, H., Smit, H., 1986. Mathematical modelling of lipid production by oleaginous yeasts in continuous cultures. Antonie Van Leeuwenhoek 52, 491–506.

Zhu, L.Y., Zong, M.H., Wu, H., 2008. Efficient lipid production with Trichosporon fermentans and its use for biodiesel preparation. Bioresour. Technol. 99, 7881–5. doi:10.1016/j.biortech.2008.02.033

Zinjarde, S.S., Pant, A., 2002. Emulsifier from a tropical marine yeast, yarrowia lipolytica NCIM 3589. J. Basic Microbiol. 42, 67–73. doi:10.1002/1521-4028(200203)42:1<67::AID-JOBM67>3.0.CO;2-M

CHAPTER II.

MODELING AND PROCESS OPTIMIZATION

CHAPTER II. MODELING AND PROCESS OPTIMIZATION

SUMMARY OF THE CHAPTER

This chapter presents a bibliographic review involving the three main axes of this thesis: modeling, control, and state estimation. In the first section the existing models for lipid accumulation by yeasts and fungi are revised. The uptake and production rate kinetics of these models are analyzed for the proposition of new models. The second section assess the dynamic metabolic modeling frameworks, which are based on the balance growth condition (also called quasi-steady state assumption) that allows the implementation of the hereafter presented modeling approaches (CHAPTER III, section III). These models are regarded at two levels: static modeling of metabolism, and dynamic modeling frameworks. The static metabolic models are based on the stoichiometric matrix and are interested on analyzing and providing insights about metabolic fluxes, where the most important yeast models are also presented. The dynamic models link the analysis of the metabolic fluxes to the mass balances of extracellular variables. Several approaches are here addressed to illustrate and to highlight the manner in which this metabolic fluxes are dynamically adapted.

Later, the optimization of bioprocess attained by model based control strategies is regarded. The general principles, advantages, and drawbacks of those strategies are reported with some of their applications to fermentations. In this section, some optimization algorithms are also presented to solve single objective and multi-objective optimization problems. However, these control strategies are almost not feasible without the development of monitoring tools such as software sensors (*i.e.* model-based and data-based soft-sensors) whose description and application to fermentation close this chapter.

I. UNSTRUCTURED MODELING OF LIPID ACCUMULATION

Mathematical modeling is an understanding tool that captures the relevant features of systems at different levels of complexity depending on the knowledge objectives. Models are representations and approximations of the systems designed to attain a great variety of goals, for instance, bioprocess models allow the estimation of non-measured variables, detection of failures and optimization of process performances by control laws (Dochain, 2008). Therefore, different model frameworks can be proposed for the same system to achieve their specific objectives.

The most common way to develop models characterizing the process dynamics is to consider the material balances of major components of the process. For the sake of illustration, let us express the mass balances of a perfectly mixed bioreactor, in a general from as,

$$\frac{d(\xi.V)}{dt} = K.r(\xi,\theta,t) + F_{IN}.\xi_{IN} - F_{OUT}.\xi \qquad \text{(II.1)}$$

where ξ is the state variable vector to be analyzed, V holds for the volume of the reactor. F_{IN} and F_{OUT} are the respective inflow and outflow of the reactor related to their respective concentrations ξ_{IN} and ξ. The term $r(\xi,\theta,t)$ is the uptake rate vector that may be function, in the more general case, of the state vector ξ, the vector of kinetic parameters θ and the time t. K represents the matrix of yields.

Equation (II.1) can be used for the different culture modes:

- Batch $F_{IN} = F_{OUT} = 0$
- Fed-batch $F_{OUT} = 0$
- Continuous culture $F_{IN} = F_{OUT} = F$. In this kind of culture it is possible to define the dilution rate as $D = F/V$.

In a general overview, models can be categorized into unstructured and structured models (Bailey, 1998). The unstructured models are essentially kinetic equations that describe the variation of substrate or product concentrations. According to this concept, the bioprocess dynamics depends, directly and only, on the macroscopic variables representing the working

conditions in the bioreactor (Almquist *et al*., 2014). However, unstructured models have proven enormous interest in bioengineering for the design of the on-line algorithms for bioreactor monitoring, control, and optimization (Bastin and Dochain, 1990). Most process comprises nonlinearities as a result of their complex dynamics. For a long time the macroscopic perspective of modeling biotechnological processes has proven to be efficient for solving many bioengineering problems (Bastin and Dochain, 1990; Provost, 2006).

Structured models designate a formulation in which the cell is composed of multiple chemical components (Fredrickson *et al*., 1970). These models provide information about the physiological state of the cells. The structured models can be: multi compartment models, genetically structured models, and biochemical structured models (Koutinas *et al*., 2012).

Lipid accumulation on oleaginous yeast and fungi has been scarcely modeled by the unstructured modeling approach (Papanikolaou et al., 2006; Papanikolaou and Aggelis, 2003b; Ykema et al., 1986). As presented in CHAPTER I, section I.1, special interest is paid to the yeast *Yarrowia lipolytica*, in which lipid accumulation is accompanied by citric acid production. Therefore, in the next subsections, the unstructured models are addressed by their aimed product: lipids, citric acid, or the concomitant production of both. Special interest is brought to their various kinetic equations, as displayed in Table II.1. Finally, the limitations of these models are regarded towards their extension.

I.1. Modeling lipid accumulation

The first models proposed for lipid accumulation in yeast and fungi were based on the idea that the C/N ratio could be used to predict growth (biomass concentration) and lipids concentration. That was the case of the model of Ykema *et al*. (1986) who studied lipid accumulation in the fungus *Aprotrichum curvatum*, growing on glucose under strictly nitrogen-limited conditions. In this model, two phases were distinguished: (i) nitrogen excess, and (ii) nitrogen limitation. This model showed that the lipid content of the given fungus could be predicted at any C/N ratio when yields of intermediates, lipids, and biomass were known. However, the definition of those yields is not straightforward which limits the application of this model. The model of Ykema *et al*. (1986) was extended to be applied in other substrate (*i.e*. whey permeate) (Ykema

et al., 1988), where the model could predict the lipid yield, but it neither predicted the lipid content, nor the concentrations of biomass and intermediate metabolites.

Continuing the trend of defining different phases, other authors (Aggelis et al., 1995a; Aggelis and Sourdis, 1997) proposed a three-phased model, where lipids were divided into: structural lipids (constituents of the membrane) and storage lipids. Hence, biomass was also separated into: catalytic biomass (without storage lipids) and total biomass (with lipids). Those authors considered in a first time a growth phase (catalytic biomass and structural lipids formation). Afterwards, substrates were converted into storage lipids and catalytic biomass. In a third time, carbon source was exhausted, and storage lipids were consumed to produce catalytic biomass. This phenomenon is known as turnover. Even if all these models could predict biomass and lipid concentrations, the specific growth rates (μ) were considered as constants limiting the information related to the kinetics of each variable.

Different kinetics equations were presented to describe the three phases of lipid accumulation (Table II.1). Models considered Williams-type kinetics to define growth based on the concentrations of: the carbon substrate (Papanikolaou and Aggelis 2003a), nitrogen (Papanikolaou *et al*., 2006), or based on double kinetics involving both carbon and nitrogen concentrations (Economou *et al*., 2011; Fakas *et al*., 2009). However, it is worth noting that Economou *et al*. (2011) included a term to take into account that lipid accumulation is favored in low concentrations of nitrogen.

In addition to nitrogen depletion conditions, Meeuwse *et al*. (2011b) studied three limitation cases on continuous cultures with the fungus *Umbelopsis isabellina:* (a) nitrogen limitation, (b) carbon limitation and (c) double limitation. Differently from the other works, they assumed that the specific production rate of structural lipids was proportional to the specific growth rate. The inclusion of other fermentation indicators (*i.e.* respiratory quotient) have been also proposed to easily relate monitoring variables with the dynamics of lipid production (Meeuwse *et al*., 2012). The calculation of the respiratory quotient, as well as the state variables, was described differently for each of the three phases: (1) growth phase, (2) lipid accumulation phase, and (3) turnover phase. The drawback of this model relies on the determination of switching times (t_{12} and t_{23}) for phases (1) – (2), and (2) – (3), respectively.

Table II.1. Summary of kinetics for citric acid production and lipid accumulation models by yeast and fungus.

Reference	Microorganism	Culture	Conditions	Substrate	Product	Kinetics
(Ykema *et al*., 1986)	*Aprotrichum curvatum*	Chemostat	T=30°C pH=5.5	Glucose	Lipids	$D_{\max} = \mu^{\max} \frac{(C/N)_{critical}}{(C/N)}$ $\pi_L = (L)(D)(S)$
(Ykema *et al*., 1988)	*Aprotrichum curvatum*	Chemostat	T=30°C pH=5.5	Whey permeate	Lipids	$N_A = N_B + [N_{A0} - N_B]e^{-Dt}$
(Aggelis *et al*., 1995a, 1995b)	*Mucor circinelloides*	Chemostat	Not reported	Sunflower oil	Lipids	$L = L_0 + S_0(1 - e^{k_2 \cdot t}) - \frac{\ln X - \ln X_0}{k_1}$
(Aggelis and Sourdis, 1997)	*Mucor circinelloides*	Batch	Simulation	Sunflower oil	Lipids	$\mu = \mu(S).Y_{X/S}S + \mu(L_S)Y_{X/Ls}Ls$ $X = X_f + L_X$ $L = L_X + L_S$
(Arzumanov *et al*., 2000)	*Yarrowia lipolytica*	Fed-Batch	T=28°C pH=4.5	Ethanol	Citrate	$\mu = \mu^{\max} \frac{N}{N + K_N}$ $\pi_{CIT} = a + b\mu$
(Papanikolaou and Aggelis, 2002)	*Yarrowia lipolytica*	Chemostat	T=28°C pH=6	Glycerol	Lipids	$\mu = \frac{D.X_f}{X}$ $\pi_L = \frac{D.L}{X}$

Reference	Microorganism	Culture	Conditions	Substrate	Product	Kinetics
(Papanikolaou and Aggelis, 2003a)	*Yarrowia lipolytica*	Batch	T=28°C pH=6.5	Glycerol	Citrate	$\mu = \begin{bmatrix} \mu^{\max}\frac{N}{N+K_N} \\ \mu^{\max}\frac{N}{N_0} \\ \mu^{\max}\left(1-\frac{x}{x_{\max}}\right) \end{bmatrix}$ $\pi_{CIT} = \alpha$
(Papanikolaou and Aggelis, 2003b)	*Yarrowia lipolytica*	Batch	T=30°C pH=6 C/N=35	Fatty acids	Lipids	$\pi_L = \pi_L^{\max}\frac{S}{S_0}$ $\mu(L) = \mu^{\max}\frac{L}{L_0}$ $\mu(S) = \mu^{\max}\frac{S}{S_0}$
(Papanikolaou *et al*., 2006)	*Yarrowia lipolytica*	Batch	T=28°C pH=5.1-6 C/N=110,138,172,500	Glucose Glucose/Stearin	Citrate Citrate/Lipids	$\mu = \mu^{\max}\frac{N}{N_0}$ $\pi_{CIT} = \alpha$ $\pi_L = \beta$

Reference	Microorganism	Culture	Conditions	Substrate	Product	Kinetics
(Fakas *et al*., 2009)	*Thamnidium elegans*	Batch	T=28°C	glycerol	Lipids	$\pi_L = \pi_L^{\max} \frac{N_0 - N}{N_0} \frac{S}{K_S + S}$
(C. N. Economou et al., 2011)	*Mortierella isabellina*	Batch	T=28°C pH=6 C/N=52,44,53	Sweet sorghum	Lipids	$\mu_{SN} = \mu_{SN}^{\max} \frac{S}{K_S + S + \frac{S^2}{K_I}} \frac{N}{K_N + N}$ $\pi_L = \pi_L^{\max} \frac{S}{K_S + S + \frac{S^2}{K_I}} \frac{k_2}{k_2 + N}$ (Lipid turnover) $\mu_L = \mu_L^{\max} \frac{L}{K_L + L} \frac{k_1}{k_1 + S}$
(Meeuwse *et al*., 2011)	*Umbelopsis isabellina*	Chemostat	T=28°C pH=6	Glucose	Lipids	$X = N_0 Y_{X/N}$ $L = \frac{\pi_L X}{D}$ $\pi_L = \pi_L^{\max}$
(Meeuwse *et al*., 2012)	*Umbelopsis isabellina*	Submerged Batch	T=28°C pH=6	Glucose	Lipids	$L = \left(\frac{f_{L_0}}{1 - f_{L_0}} + \frac{\pi_L^{\max}}{k_d} \left(1 - e^{-k_d \cdot (t - t_{12})}\right) \right) X^{\max}$ $L = L(t_{23}) - m_S / 1.4.X^{\max}$

I.2. Modeling Citric acid

Concerning *Yarrowia lipolytica*, lipid accumulation is accompanied by the formation of citric acid (Anastassiadis et al., 2002b), which can be produced in even greater quantities than lipids. The production of this organic acid has been modeled neglecting lipid accumulation. The kinetic equations describing citric production were assumed as linear equations when produced from ethanol (Arzumanov *et al*., 2000), glycerol (Papanikolaou and Aggelis, 2003a), and glucose (Papanikolaou *et al*., 2006).

I.3. Modeling citric acid and lipids

In literature, only one model has been proposed to describe the concomitantly production of citric acid and lipids (Papanikolaou *et al*., 2006). The authors studied the production of citric acid from glucose by *Yarrowia lipolytica*. The aim of this model was to calculate and to quantify the dynamics of biomass, storage lipids, and citric acid concentrations. In a first time, glucose was used as unique substrate where citric acid was produced with negligible quantities of lipids. Afterwards, experiments with mixtures of glucose/stearin were carried out to promote lipid accumulation. Differently to the propositions of Aggelis *et al* (Aggelis et al., 1995a, 1995b; Aggelis and Sourdis, 1997), they identified five pathways. (i) incorporation of the exo-cellular aliphatic chains into the stored lipid structures; (ii) elongation and desaturation reactions; (iii) partial or total degradation of acyl-CoA units by the β-oxidation; (iv) catabolism of the acetyl-CoA through Krebs cycle and the glyoxylate acid; and (v) de novo biosynthesis of fatty acids, which completes the cycle for lipid accumulation. In this model, the production of citric acid was assumed to be the consequence of a change of substrate to glucose, which caused metabolic carbon overflow (carbon excess) (Papanikolaou *et al*., 2006). Even if the model construction is interesting for application, the results only showed adequate performance for structural lipids, since the reported experiments did not reported accumulated lipids when glucose was used as substrate.

I.4. Limitations of unstructured models

The models presented in Table II.1 are regarded as unstructured models, which are used for phenomenological description of processes and provide correct predictions for batch and

continuous cultures. However, those models fall short to predict the switch towards lipid accumulation and citric acid production. As this is mainly due to environmental and internal disturbances, these unstructured models will not predict accurately the real process (*e.g.* intracellular mechanisms).

Even if those models were defined on several phases, they neither considered internal regulation of metabolism, nor metabolic shifts, which has to be defined by new approaches integrating kinetics with metabolic models (Ramkrishna, 2003).

The models, here presented, relate extracellular substrates with the final products by a single macroscopic reaction. Therefore, this approach is unable to represent the interaction of internal metabolites and the complexity of cell metabolism. Following this ideas, it is necessary to use several reactions that are linked in a so-called metabolic network to have a better and more complete representation. Dynamic Metabolic Models (DMM) (H.-S. S. Song et al., 2013) deal with the link between metabolic and macroscopic description, which will be described in the next section.

II. MODELING METABOLISM

II.1. Metabolic modeling framework

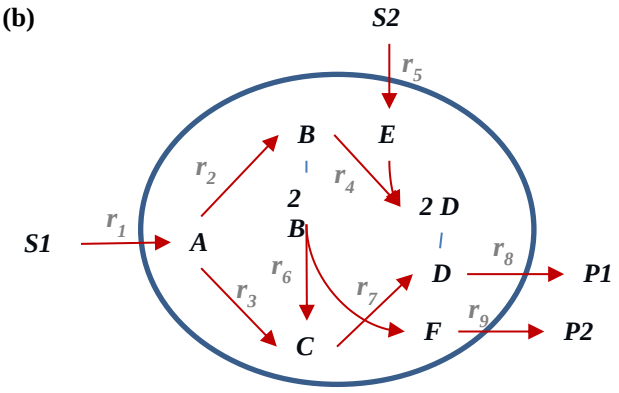
Usually, dynamic bioprocess models used for control purposes describe the conversion of initial substrates into final products by a small set of macroscopic reactions connecting the substrates to the targeted products. These models are known as unstructured models because they relate extracellular substrates with the final products by macroscopic reactions. Unfortunately, this approach is unable to represent the interaction of internal metabolites and the complexity of cell metabolism. Following this ideas, it is thus necessary to use several reactions that are linked in a so-called metabolic network to have a better and more complete representation of the microbial behavior. However, the number of reactions should be moderate in order to find a trade-off between simplicity and good fitting between data and simulations.

Table II.2. Example of representation of a stoichiometric matrix (Trinh *et al*., 2009)**.**

(a)

v1: S1 => A
v2: A => B
v3: A => C
v4: B + E => 2 D
v5: S2 => E
v6: 2 B => C + F
v7: C => D
v8: D => P1
v9: F => P2

(b)



(c)

Reactions

Metabolites	v1	v2	v3	v4	v5	v6	v7	v8	v9	
S1	−1	0	0	0	0	0	0	0	0	r_1
A	1	−1	−1	0	0	0	0	0	0	r_2
B	0	1	0	−1	0	−2	0	0	0	r_3
C	0	0	1	0	0	1	−1	0	0	r_4
D	0	0	0	2	0	0	1	−1	0	r_5
E	0	0	0	−1	1	0	0	0	0	r_6
F	0	0	0	0	0	1	0	0	−1	r_7
S2	0	0	0	0	−1	0	0	0	0	r_8
P1	0	0	0	0	0	0	0	1	0	r_9
P2	0	0	0	0	0	0	0	0	1	

S r

The metabolic modeling framework relies on the knowledge of the metabolic network and its stoichiometry. The characterization of cell metabolism has led to the connection of the set of metabolic pathways into a metabolic network (Llaneras and Picó, 2008). A metabolic network is the representation of a set of extra and intracellular metabolites (nodes) connected through reactions (links or arrows). To clarify the concept, let us take the set of reactions of Table II.2 (a), and their graphical representation in Table II.2 (b). The cell wall is illustrated as the blue line containing all the intracellular metabolites *A* – *F*, and portraying the extracellular metabolites *S1*-*S2* and *P1*-*P2*. This can be mathematically expressed thanks to the stoichiometric matrix *S* of the metabolic network. This matrix encloses the stoichiometric information, where the rows denote the l_i metabolites, and the columns designate the v_j reactions. The coefficients $S_{i,j}$ of the matrix *S* are the coefficients of the reactions where,

- $S_{i,j} = 0$ if the metabolite l_i does not participate in the reaction v_j
- $S_{i,j} < 0$ if the metabolite l_i is a substrate in the reaction v_j

- $S_{i,j} > 0$ if the metabolite l_i is a product in the reaction v_j

In this case, irreversible reactions are described by '=>', whereas reversible reactions are denoted by '='. However, reversible reaction can be split into two (backward and forward reactions) (Maertens and Vanrolleghem, 2010). Table II.2 (c) exhibits the stoichiometric matrix of the example previously defined.

As seen in the example, the stoichiometric matrix portrays cell metabolism in a mathematical form to describe the behavior of cells. However, it is relevant to know how the evolution of those metabolites could affect the production of a targeted metabolite. A commonly used framework is based on dynamic models.

II.1.1. Quasi-Steady State Assumption

Dynamic models are usually defined by mass-balances express as a set of Ordinary Differential Equations (ODE) that change according to reactor configuration (Bastin and Dochain, 1990). For the sake of simplicity, let us consider a microorganism growing in a perfectly mixed batch bioreactor with a constant volume. It consumes extracellular substrates, to synthesize biomass X ($n_X \times 1$), and excrete products. The corresponding ODE system relating substrates and products, which are included in the vector M ($n_M \times 1$) of external or extracellular metabolites, and biomass obtained is:

$$\frac{d}{dt} M = S_M . r . X \tag{II.2}$$

$$\frac{d}{dt} X = \mu . X \tag{II.3}$$

where S_M ($n_M \times n_r$) is the stoichiometric matrix containing the n_M external metabolites, and the n_r reactions, as defined in Table II.2 (c).The vector r ($n_r \times 1$) represents the reaction rates (per unit of biomass) of the reaction in the metabolic network, whilst μ is the term of specific growth rate.

Following equations (II.2) and (II.3), the kinetics of each metabolic reaction, is thus needed to simulate the transient dynamics of intracellular metabolites m ($n_m \times 1$).

$$\frac{d}{dt} m = S_m.r - \mu.m \tag{II.4}$$

where S_m ($n_m \times n_r$) is the stoichiometric matrix of intracellular metabolites, and the term $\mu.m$ represents their dilution with respect to the biomass concentration. However, the experimental measurement of intracellular metabolites along cultures is complicated, reason why the so-called quasi-steady state assumption (QSSA) should be considered (Schuster *et al*., 2002; Song and Ramkrishna, 2009a). This assumption implies that the intracellular metabolites consumption and production fluxes are balanced. In other words, QSSA states that the dynamics of the internal or intracellular metabolites *m* are faster than the dynamics of the external metabolites *M*, therefore, internal metabolites can be considered at steady state. The mathematical representation of this assumption is,

$$\frac{d}{dt} m = 0 \tag{II.5}$$

and if the term $\mu.m$ is considered negligible, it is possible to define

$$S_m.r = 0\,,\; r \geq 0 \tag{II.6}$$

Consequently, the n_r reactions only depend on the stoichiometry of the metabolic network, and the uptake of substrates. Nevertheless, the solution space of equation (II.6) implies all the feasible fluxes described by the metabolic network. Usually, the n_r reactions are larger than the n_m internal metabolites. Hence, the solution of (II.6) is undetermined.

In order to handle this difficulty, it is necessary to define some constrains that features differences in cell behavior. Depending on the structure and knowledge of the stoichiometric matrix, the constraints could be non-adjustable and adjustable (Trinh *et al*., 2009). The non-adjustable are time-invariant, for example irreversibility of fluxes ($r \geq 0$), and enzyme capacities ($r < r_{\max}$). On the other hand, adjustable constraints considered measured fluxes, regulatory constraints, and kinetics, which could change due to environmental conditions or from one cell to another (Llaneras and Picó, 2008).

The analysis of fluxes through a metabolic network provides advantageous information to study metabolism and to enhance the production of targeted metabolites for biotechnological interests. Concerning flux distribution, irreversible fluxes assumed to follow only one direction and are naturally constrained to be non-negative (Maertens and Vanrolleghem, 2010). Convex analysis shows that the space of flux is a convex polyhedral cone (Schuster *et al*., 2002). Under

this principle the number of possible solutions is bounded. This mathematical description becomes the core of this framework.

Several methods were proposed in order to study the solution space of equation (II.6) such as: Flux Balance Analysis (FBA) (Raman and Chandra, 2009), Elementary Flux Modes (EFM) (Schuster *et al*., 2002), Metabolic Flux Analysis (MFA)(Zamboni, 2011). These static frameworks were modified in order to simulate dynamic systems, for example: Macroscopic Bioreaction Models (MBM) (Provost *et al*., 2006), Hybrid Cybernetic Models (HCM)(Kim *et al*., 2008; Song *et al*., 2009), Dynamic Flux Balance Analysis (DFBA) (Mahadevan *et al*., 2002), and Dynamic Reduction of Unbalanced Metabolism (DRUM) (Baroukh *et al*., 2014).

Then, two types of frameworks will be described and commented in the two next sections. The static metabolic models are interested on analyzing and providing insights about metabolic fluxes whose results are achieved by convex analysis, linear programming and the imposition of constraints. The dynamic metabolic models link the analysis of the metabolic fluxes to the mass balances of extracellular variables. This is accomplished by the definition of time-variant uptake and excretion rate kinetics. Several approaches are here addressed to illustrate and to highlight the manner in which this metabolic fluxes are dynamically adapted.

Figure II.1 depicts a schematic summary of those frameworks.

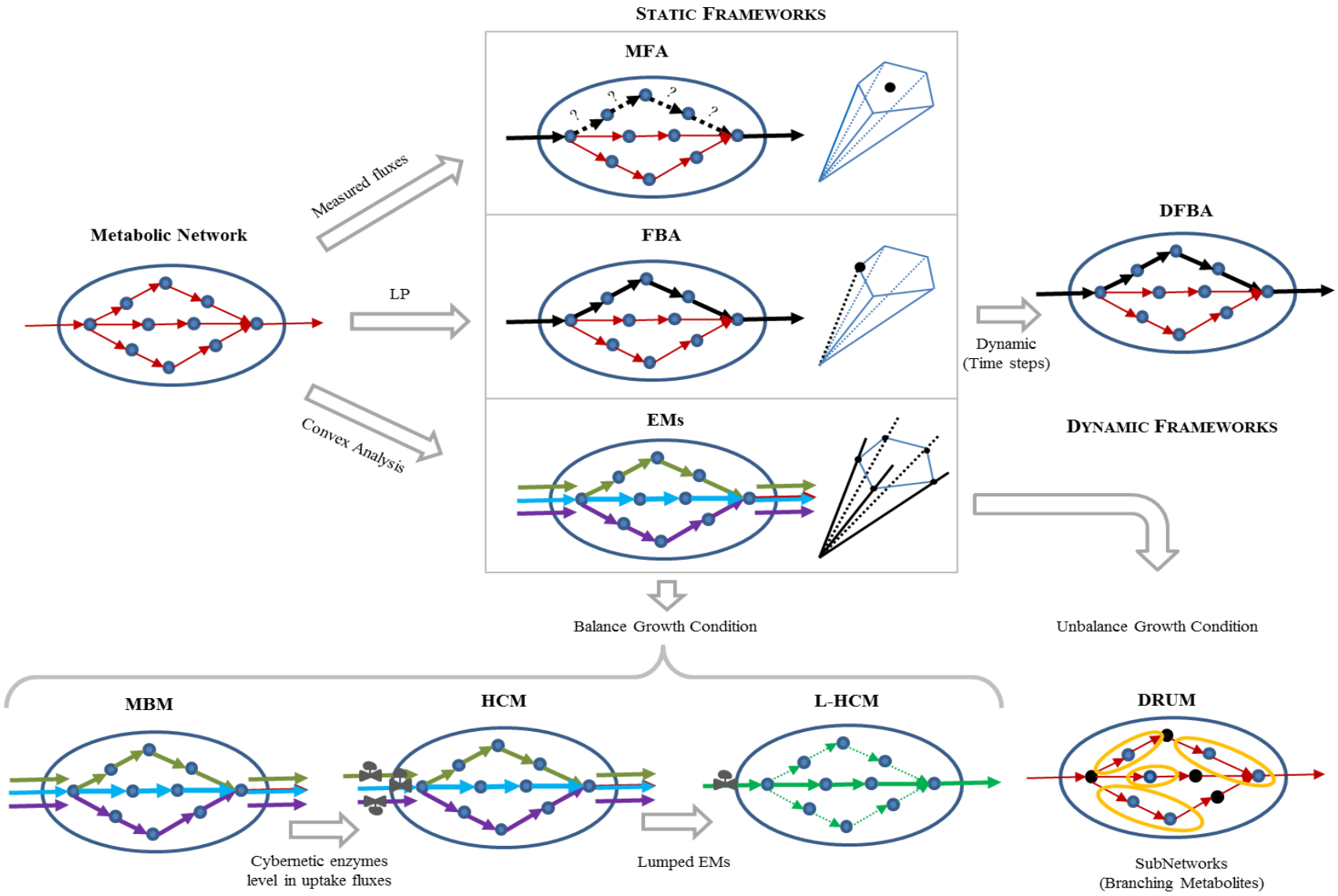


Figure II.1. Summary of some existing Metabolic Models.

II.2. Static Metabolic modeling

II.2.1. Metabolic Flux Analysis (MFA)

Metabolic Flux analysis was proposed to quantify the metabolic fluxes relying on a known metabolic network and the QSSA (Varma and Palsson, 1994; Wiechert, 2001). MFA analyze the topology of the metabolic network at steady state at which each reaction occurs (Zamboni, 2011). In this context, a set of measured fluxes is used in combination with the knowledge of cell metabolism to determine the non-measured fluxes. Furthermore, it is possible to use constraints imposed by the measured fluxes to offset the indeterminacy of the system (Llaneras and Picó, 2008). The reaction rates *r* is thus partitioned into two subgroups (vectors) regarding the r_{mes} ($n_{r_{mes}} \times 1$) measured fluxes and the unmeasured fluxes r_u ($n_{r_u} \times 1$), which are linked to their respective stoichiometric matrices S_{mes} ($n_{mes} \times n_{r_{mes}}$) and S_u ($n_r \times n_{r_u}$). The solution of equation (II.6) is then reduced to compute the unmeasured fluxes r_u. as,

$$S_u.r_u - S_{mes}.r_{mes} = 0 \tag{II.7}$$

When the system is determined, a unique set of values for the unmeasured fluxes that satisfy (II.7) can be calculated as,

$$r_u = -S_u^*.S_{mes}.r_{mes} \tag{II.8}$$

where S_u^* is the Penrose pseudo-inverse of the stoichiometric matrix of unmeasured fluxes (Llaneras and Picó, 2008).

Some examples of this method comprise the study of: transient processes in animal cell cultures (Herwig and von Stockar, 2002); lactate production in mammalian cells (Crown *et al*., 2012); and the optimization of ethanol from xylose by *Candida shehatae* (Bideaux *et al*., 2016). Nevertheless, its limitations deal with the number of fluxes that should be provided to MFA for the determination of the unmeasured fluxes (Maertens and Vanrolleghem, 2010; Trinh *et al*., 2009).

II.2.2. Flux Balance Analysis (FBA)

FBA is based on convex analysis under a constraint optimization problem. This approach was developed by (Varma and Palsson, 1994) based on linear optimization. FBA is one of the most widely used algorithms for analyzing metabolic systems, to dynamic modeling (Papin *et al*., 2004; Ruppin *et al*., 2010). The main assumption of FBA is that cell evolves to achieve an optimal metabolic objective. Some commonly used objective functions include: production of ATP, production of a desired by-product, or growth rate (Edwards *et al*., 2002; Maertens and Vanrolleghem, 2010).

FBA employs linear programming (LP) to solve the equation (II.6) (Orth *et al*., 2010). The uptake fluxes are required to be given as an input to a LP problem where the other of fluxes are calculated to maximize a prescribed objective function (*e.g.* biomass yield) (Provost 2006; Song *et al*. 2014). This flux distribution is then used to understand the metabolic capabilities of the system. In the case of maximizing biomass concentration, the LP problem can be represented as,

$$\begin{aligned} &\max r_{Biomass} / r_{Substrate} \\ &s.t \\ &b^L < \frac{r_{Biomass}}{r_{Substrate}} < b^U \\ &r > 0 \\ &S_m.r = 0 \end{aligned} \tag{II.9}$$

where the solution is constrained between a lower and an upper bound, b^L and b^U, respectively. Thermodynamic constraints (Beard *et al*., 2002) can be also applied to define limits on the range of values for individual fluxes in the network.

There are several important issues that arise when analyzing the optimal metabolic flux distribution. First, FBA relies on *a priori* knowledge (measured fluxes), but the knowledge on the regulatory mechanisms (*e.g.* regulation of enzymatic reactions) are still lacking (Maertens and Vanrolleghem, 2010). Second, the optimal solution may not always correspond to the real flux distribution because the solution is bounded to attain an specific objective (Llaneras and Picó, 2008). To overcome these hurdles, assumptions regarding cellular behavior based on the optimal solution must be made. Some of them assume that cells have evolved towards an optimal behavior,

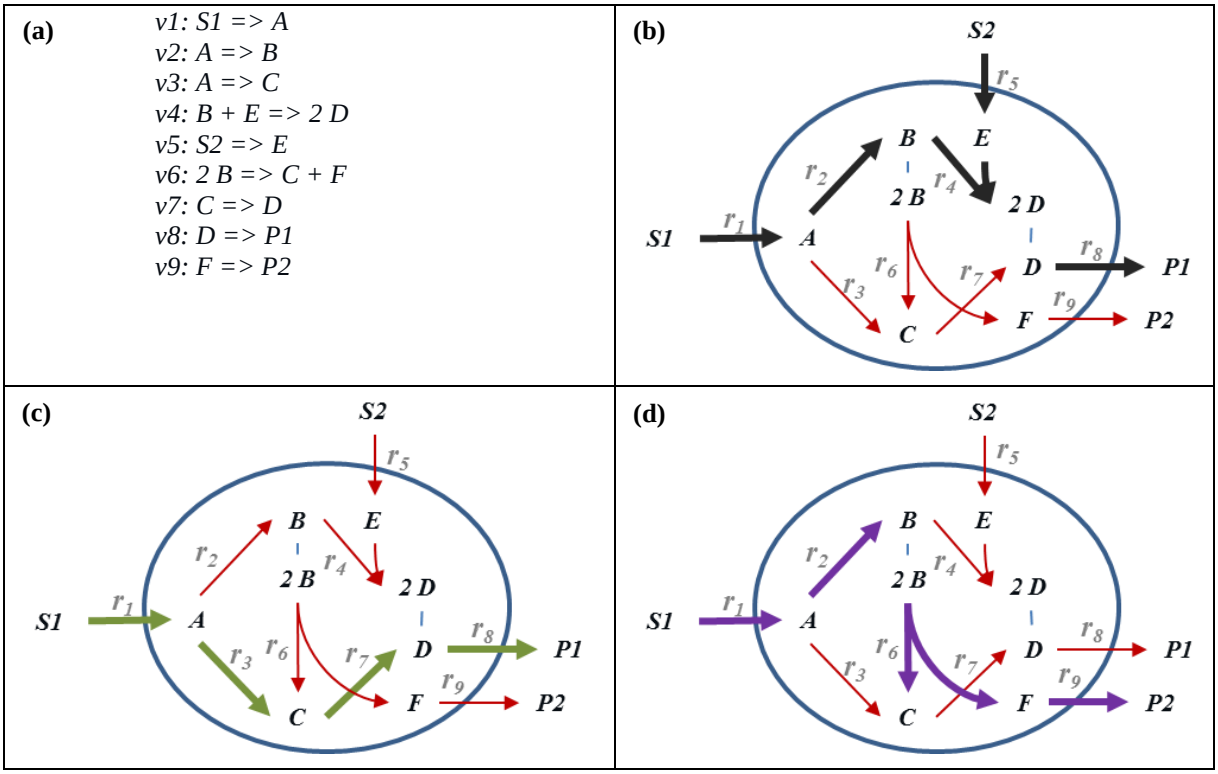
and that the objective function that has been mathematically imposed is consistent with the evolutionary objective (Feist and Palsson, 2010). However, the efficiency of the method highly depends on proper selection of the objective function, which is intended to represent the whole system (complete metabolism of a microorganism). This is not an easy task since each objective leads to different solutions. This was proven by Schuetz *et al*., (2007) studied the effect of changing objective functions by evaluating flux prediction with 11 linear and nonlinear objective functions in *Escherichia coli* with eight adjustable constraints. One of the most popular tool used to compute FBA is the COBRA toolbox implemented in MATLAB (Becker *et al*., 2007; Schellenberger *et al*., 2011). Nevertheless, the advantage of FBA is that it needs less data than MFA (Song et al., 2014a).

FBA has been used for the genome-scale reconstruction of the metabolic networks. For example, *Escherichia coli* K-12 (Feist *et al*., 2007); *Pseudomonas putida* (887 reactions) (Puchałka *et al*., 2008), which was also used for predicting the growth yield (*i.e.* biomass/substrate); and *Yarrowia lipolytica* metabolism (1143 reactions) (Pan and Hua, 2012), where the model was analyzed *in silico* for 24 different substrates. FBA has also been applied for strain improvement by metabolic engineering, for example Koffas and Stephanopoulos, (2005) achieved higher product yield and productivities of lysine production with strains of *Corynebacterium glutamicum*. In general, FBA applications are focused on metabolic engineering, construction and analysis of genome-scale metabolic models, and the refinement of the existing bio-chemical/metabolic networks. A detailed review on applications is revised by Raman and Chandra, (2009).

II.2.3. Elementary Modes analysis

Elementary Modes (EMs) are defined as non-decomposable pathways consisting on the minimum set of reactions under quasi steady-state assumption (Schuster *et al*., 2002). They can be seen as the set of successive reactions to transform substrates into products. Table II.3 could be considered as an example to clarify this concept. Here, the sets of reactions in Table II.3 (a) are schematized on Table II.3 (b-d), which take into account three different paths to obtain P1 from S1 and S2 (Table II.3 b), P1 from S1 (Table II.3 c), and/or P2 from S2 (Table II.3 d).

Table II.3. Examples of Elementary modes.



EMs analysis is oriented to system analysis without seeking for a particular flux solutions (Song et al., 2014a). Here, the vector *r* is considered non-negative and all the fluxes irreversible. Therefore, the solution corresponds to a cone that contains the set of EMs. In this context, flux vector *r* can be represented by a combination of Elementary Modes (EMs) (Schuster *et al.*, 2002) as,

$$r = Z.r_M \tag{II.10}$$

where Z ($n_r \times n_Z$) is the EMs matrix, and r_M ($n_Z \times 1$) is the kinetic vector that describes the n_Z elementary flux modes, so that (II.2) can be rewritten,

$$\frac{d}{dt}M = S_M Z.r_M.X \tag{II.11}$$

$$\frac{d}{dt}X = \mu.X \tag{II.12}$$

and $\mu = S_X Z$ where S_X is the stoichiometric matrix for the biomass. The set of EMs can be computed by different software implemented in MATLAB®. The most popular are METATOOL (von Kamp and Schuster 2006), *efmtool* (Terzer and Stelling, 2008), or CellNetAnalyzer (Klamt *et al.*, 2007) that compiles both software.

Applications of EMs regard the analysis of the distribution of metabolic processes to assess the effects of kinetic changes on microbial production; for example in the metabolism of *Escherichia coli* (Pfeiffer *et al*., 1999) and *Saccharomyces cerevisiae* (Schwartz and Kanehisa, 2006).

Unfortunately, in a real scheme the number of reaction is larger, and thus the number of EMs. This is known as combinatorial explosion. For example, a system of 112 reactions and 89 internal metabolites describing *Escherichia coli* computed 2 456 787 EMs (Kamp et al., 2006), reason why a reduction of the solution space is needed. Several efforts for reducing the number of EMs have been presented (Wagner and Urbanczik, 2005) using Flux Balance Analysis (FBA), or using the analysis of thermodynamic feasibility (Jol *et al*., 2012) of EMs, where a reduction to 54% was achieved.

Nevertheless, other straightforward strategies are based on convex analysis, under the principle of a convex cone (Donoho and Stodden, 2004), where it is assumed that all the points inside this convex cone could be represented by the border points, also known as Generating Modes (GMs). These methods represented an important step towards making full quantitative use of stoichiometry for maximizing biomass/substrate yield with information about the metabolic state of the organism.

In addition to FBA, Song and Ramkrishna, (2009b) proposed a reduction method based on the transformation of the EMs into yields (*i.e.* method of Yield Analysis) by normalizing EMs to one compound (*e.g.* substrate) of the metabolic pathways. Yield analysis (YA) allowed the simplification of the convex cone into a convex hull, which can be visualized in 2-D. The method of YA implies the reduction of convex hull by (i) selecting the Generating Modes (GMs) that form the greatest triangle and (ii) increasing the area of the convex hull by adding one vertex (one GM) at a time until reaching 99% of the total area of the convex hull. Once the generating modes are selected, they are called active modes (AMs) which are supposed to best represent the analyzed system.

Another method of reduction was proposed by (Robles-Rodriguez *et al*., 2014). This method follows the idea of formerly presented Yield Analysis (Song and Ramkrishna, 2009b), but varies

in the way of selecting the active modes. In our method, the selecting of the active modes is performed by an optimization attempting to minimize the error between a centroid and the experimental data maximizing, at the same time, the area of the convex hull. The utilization of this method will be developed in CHAPTER III, section II, and further applied to the study of lipid accumulation.

II.2.4. Metabolic Modeling yeasts

The presented frameworks allow for the analysis of metabolic networks in order to model metabolism. However, since this work focuses on lipid accumulation by yeast a brief summary of the previous yeasts models is here presented. Most of these models were called genomic scale models because they tried to cope gene actions with their metabolic activity. Taking advantage of these genes similarities, new models were proposed growing up in complexity and insights about the metabolism of yeasts. The most important models are presented on Table II.4.

The first model was proposed by (van Gulik and Heijnen, 1995), where they used 129 reactions and 98 metabolites to explain the metabolism of *Saccharomyces cerevisiae*. An important remark from their work was the introduction of two important factors: maintenance and the P/O ratio. These ideas were also taken by (Vanrolleghem et al., 1996), who estimated both parameters for biomass yield of S. cerevisiae. Their model called iFF708 was used to predict the transition of feeding from glucose to ethanol by adding carbon balances. Compartmentation was first included on the work of (Nissen et al., 1997) by treating compounds unable to cross the inner mitochondrial membrane adding two different compounds (*i.e.* cytosol and mitochondria), where cofactors could be compartmented.

The first model to perform a comprehensive reconstruction of the yeast metabolism was the model by (Forster et al., 2003). Furthermore, this model was validated by FBA, where the results showed to have good agreements with many experimental data bases. This model included 3 compartments: cytosol, mitochondria, and the extracellular space. Another model was presented by (Duarte et al., 2004) including 8 compartments (nucleus, endoplasmic reticulum, golgi, vacuole and peroxisome). In addition, all the reactions were charge balanced. The ability of this model to predict growth

phenotype for single mutants was lower than the one obtained by the model from (Forster et al., 2003). This model named iND750 was used for phenotype phase plane analysis to consider transcription factors related with carbon and nitrogen metabolism. However, it failed when using mutant strains. To overcome this hurdle, (Kuepfer et al., 2005) presented a 13C validation of the central carbon metabolism evaluating possible mutations.

From Table II.4, it is worth noting the work of (Nookaew et al., 2008), who included more reactions to detail lipid metabolism. In 2008, the different approaches were standardized with a common nomenclature, including the international chemical identifier and the KEGG identifiers. This new model was called Yeast 1 (Herrgård et al., 2008), which contained 1857 reactions relating the actions of 832 genes and 1168 metabolites from 15 different compartments. Nevertheless, a new version of this model (Yeast 4) (Dobson et al., 2010) was proposed to provide a mathematical formulation for constrained based simulations.

Additionally to the models on Saccharomyces cerevisiae, other authors focused on enlarging the metabolic models to non-conventional yeasts such as *Yarrowia lipolytica*. A state of the art was presented by (Loira et al., 2012) to predict physiological and metabolic functions. Otherwise, (Pan and Hua, 2012) tested their model for 24 substrates, indicating that the model could be used to qualitatively predict the production of biomass under different substrates.

The two main goals of genome scale modeling are to test the ability of the model to predict the phenotype as close as the experimental one, and to describe as large as possible the metabolism including biological information to serve as a comprehensive knowledge of yeast metabolism. Simulations with the genome scale models are a tool for quality control of the experimental data, and for hypothesis testing. However, these models are static representations and further approaches are needed to understand how the metabolism evolve dynamically, which will be described in the next subsection.

Table II.4. Yeast metabolic models.

Model	Microorganism	#React	#Met	#Gen	Comp.	Valid	Objective/Application	Reference
iGH99	*Saccharomyces cerevisiae*	129	98	-	1	MFA	Description of growth on glucose and ethanol	(van Gulik and Heijnen, 1995)
-	*Saccharomyces*	88	84	-	1	MFA	Estimation of PO ratio and k on acetate and ethanol consumption cultures	(Vanrolleghem et al., 1996)
-	*Saccharomyces cerevisiae*	37	43	-	2	MFA	Calculate production rates of malate and fumarate. Estimation of carbon channeling.	(Nissen et al., 1997)
iFF708	*Saccharomyces cerevisiae*	1175	733	708	3	FBA	Predicting genes essentiality when comparting to in vivo knock out phenotypes	(Forster et al., 2003)
iND750	*Saccharomyces cerevisiae*	1489	972	750	8	-	Addition of compartments to describe the biological knowledge.	(Duarte et al., 2004)
iLL672	*Saccharomyces cerevisiae*	1038	636	672	2	FBA MoMA	Removal of dead ends from iFF708	(Kuepfer et al., 2005)
iMH805/775	*Saccharomyces cerevisiae*			805	-		Transcription factor activation or repression of metabolic genes	(Herrgård et al., 2006)
iIN800	*Saccharomyces cerevisiae*	1446	1013	800	3	FBA 13C	Inclusion of lipid metabolism to iFF708	(Nookaew et al., 2008)
iMM904	*Saccharomyces cerevisiae*	1402	713	904	8	-	Correlate concentrations of metabolites measured by metabolomics approaches	(Mo et al., 2009)
Yeast 1.0	*Saccharomyces cerevisiae*	1857	1168	832	15	-	Merge and unification of approaches iLL672 + iMM04	(Herrgård et al., 2008)

Model	Microorganism	#React	#Met	#Gen	Comp.	Valid	Objective/Application	Reference
-	*Saccharomyces cerevisiae*	-	-	900	-	-	Revised model of the iMM904 based on synthetic lethality/gene essentiality analysis. Correction of multi compartment models	(Zomorrodi and Maranas, 2010)
Yeast 4.0	*Saccharomyces cerevisiae*	1102	924	932	16	-	Adaptation of Yeast 1 for computational analysis	(Dobson et al., 2010)
iJH732	*Saccharomyces cerevisiae*	1152	935	732	2	dFBA	Dynamic FBA simulation. Reduction from iND750	(Hjersted and Henson, 2009)
Non-conventional yeasts								
-	*Pichia Pastoris*	141	102	-	1	MFA	Investigate the metabolic effects of methanol/sorbitol	(Çelik et al., 2010)
iBB814	*Pichia stipites*	1371	971	814	3	FVA	In silico computed growth rates to evaluate anaerobic growth on xylose	(Balagurunathan et al., 2012)
iNL895	*Yarrowia lipolytica*	2002	1847	895	16	FBA	Prediction of growth under different media conditions and gene knock outs. Provide a state of the art in *Y lipolytica* metabolism.	(Loira et al., 2012)
iYL619_PCP	*Yarrowia lipolytica*	1142	843	619	3	FBA	Understanding of metabolic characteristics towards lipid production	(Pan and Hua, 2012)
iMK755	*Yarrowia lipolytica*	1336	1111	735	8	dFBA	Simulate growth and lipid production phases of this yeast to optimize lipid productivity	(Kavšček et al., 2015)

II.3. Dynamic Metabolic Models

II.3.1. Dynamic Flux Balance Analysis (DFBA)

Dynamic Flux Balance Analysis (DFBA) was introduced by Mahadevan *et al*., (2002) to incorporate the rate of change of flux constraints. They incorporated some kinetic expressions when the kinetics was well characterized in order to predict the concentrations of intracellular and extracellular metabolites.

Dynamic Flux balance Analysis (DFBA) offers the possibility of formulating substrate uptake kinetics to account for known regulatory processes (Hjersted and Henson, 2009). DFBA is based on a dynamic model attaining a specific objective (*e.g.* maximization of biomass concentration) where the fluxes from FBA are calculated at every time step.

At each time step, FBA is used to solve (II.13) by linear programming to compute the flux rates,

$$\begin{aligned} &\max r_{Biomass} \\ &s.t \\ &r^L < r_{Biomass} < r^U \\ &r > 0 \\ &S_m.r = 0 \end{aligned} \qquad \text{(II.13)}$$

where the maximization is bounded to r^U and r^L. These rates are further used to calculate metabolites and biomass concentrations at the end of each time step.

$$\begin{aligned} &M(t+dt) = S_M.r(t).dt \quad \forall t \in \lfloor t_0, t_f \rfloor \\ &X(t+dt) = \mu(t).dt \end{aligned} \qquad \text{(II.14)}$$

DFBA has been applied to simulate the production of ethanol from glucose and xylose (Hanly and Henson, 2011) by *Saccharomyces cerevisiae;* and a mutated strain of *Escherichia coli* able of consuming only xylose. The same authors, also studied a model were *Saccharomyces cerevisiae* and *Pichia stipites* had competition for the consumption of glucose and xylose (Hanly and Henson, 2013). DFBA has also been used for characterization of metabolic shifts on the overflow of *Saccharomyces cerevisiae* (Jouhten *et al*., 2012), and to simulate the optimization of lipid accumulation on *Yarrowia lipolytica* (Kavšček et al., 2015). A drawback

of this modeling approach is its inability to account for unbalance growth conditions encountered in batch and fed-batch cultures (Jouhten *et al*., 2012).

II.3.2. Macroscopic Bioreaction Model (MBM)

Macroscopic Bioreaction Model (MBM) is another approach that tries to represent the metabolism dynamically. It was proposed by (Provost *et al*., 2006), where metabolism is described by a set of key macroscopic bioreactions that directly connect the substrates to products. This approach was also based under the QSSA, contemplating a balanced consumption and production of intracellular metabolites.

Elementary Modes (EMs) are used to express the metabolism by macroscopic reactions. As mention is section II.2.3, the EMs are computed by convex analysis, in which a convex cone includes the complete set of non-negative solutions (Figure II.2). However, the reduction of the set of EMs is achieved by imposing specific uptake and production rates. This addition lowers the number of EMs telling which EMs or which macroscopic bioreactions are sufficient to be combined in the spite of building a model that explains the extracellular rates.

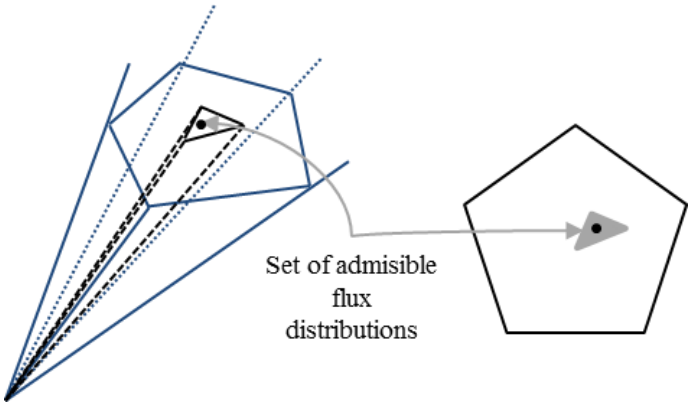


Figure II.2. Cone of flux distribution of the solution space. (The black point represents the experimental measurements) (Provost *et al*., 2006).

From a graphical point of view, EMs are chosen by according to experimental uptake and production rates taking into account the EMs that surround the experimental point (Figure II.2) (Provost *et al*., 2007). Hence, the necessity to have a fixed measured rate limits the application of this method. Even if different, MBM can also be expressed with equations (II.11) and (II.12), making the matrix $S_M Z$ be the minimal area around the experimental point.

In MBM, the dynamic metabolic behavior can be decomposed into several time phases during which some extracellular metabolites depict a tendency of consumption or production rate by biomass unit. For each phase, one set of EMs is selected and applied to describe the cell dynamics through the inclusion of smooth switching functions (*e.g.* sigmoidal functions) (Figure II.3). Nevertheless, the sigmoidal functions are time dependent and characteristic of the culture. This approach is thus neither straightforward for extrapolation to other culture, nor for application on control strategies (lack of dynamic adaptation).

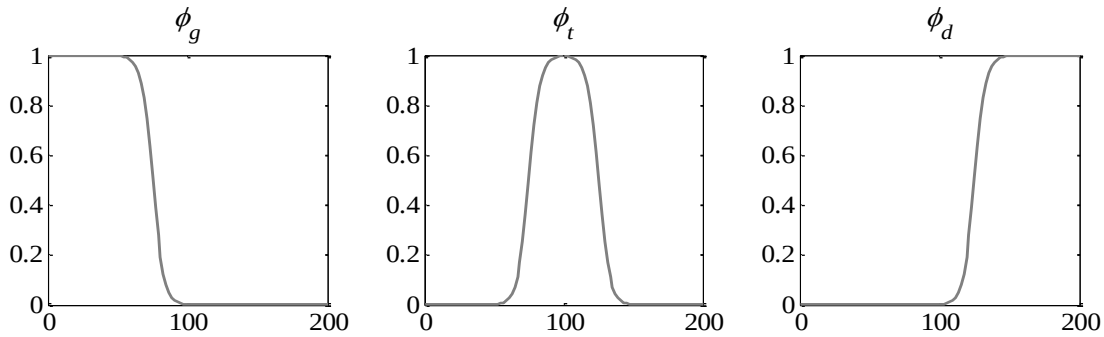


Figure II.3. Coupling of the three models corresponding to (ϕ_g) growth, (ϕ_t) transition, and (ϕ_d) death phases. (Provost *et al.*, 2006).

This method was applied on Chinese hamster ovary (CHO) cells (Provost *et al.*, 2006), where the model was defined separating cell metabolism into three phases: growth, transition and death phase. The complete model can be represented as an extension of equation (II.4) as,

$$\frac{d}{dt}M = \phi_g.(S_M Z)_g.(r_M)_g.X + \phi_t.(S_M Z)_t.(r_M)_t.X + \phi_d.(S_M Z)_d.(r_M)_d.X \qquad \text{(II.15)}$$

The functions ϕ_g, ϕ_t, and ϕ_d are the sigmoidal functions displayed in Figure II.3. Each matrix $(S_M Z)_{g,t,d}$ depicts the chosen EMs for each phase (growth phase, transition phase and death phase), and their respective kinetics vector $(r_M)_{g,t,d}$.

II.3.3. Hybrid Cybernetic Model (HCM)

Hybrid cybernetic model (HCM) (Kim *et al.*, 2008; Song *et al.*, 2009) allows to include regulatory effects at the level of "enzymes" (one cybernetic "enzyme" per EM) focused on the

description of cellular growth on multiple substrates with kinetic and stoichiometric models. This approach was constructed under the quasi-steady state assumption, in which the metabolism is taken into account by the elementary flux modes (EMs). Therefore, the description of the external metabolites (*M*) is taken as in (II.11) and (II.12), where the vector of fluxes $r_{M,i}$ is calculated as,

$$r_{M,i} = v_i.e_i^{rel}.k_i^{\max}.r_{M,i}^{kin} \tag{II.16}$$

The kinetic terms $r_{M,i}^{kin}$ is the vector of non-regulated rates which are represented, in most of the cases, by Monod-type kinetics,

$$r_{M,i}^{kin} = \prod_i \frac{M_i}{K_{M,i} + M_i} (i = 1,\ldots,n_Z) \tag{II.17}$$

$K_{M,i}$ is the saturation constant of the i^{th} mode for the extracellular metabolite *M*. The term $v_i.e_i^{rel}.k_i^{\max}$ holds the regulated rates. Hereof, each i^{th} mode is regulated by the vector of cybernetic variable v_i controlling enzyme activity and catalyzed by a vector e_i of cybernetic "enzymes", in which one "enzyme" represents one elementary mode. The term e_i^{rel} is the relative "enzyme" with respect to its maximum value $e_i^{\max}$ calculated at steady state.

$e_i^{rel} = \frac{e_i}{e_i^{\max}}$, and $e_i^{\max} = \frac{\alpha_i + k_{E,i}^{\max}}{\beta_i + k_i^{\max}}$

The dynamics of the cybernetic "enzymes" are

$$\frac{de_i}{dt} = \alpha_i + r_{EM,i} - (\beta_i + \mu).e_i \tag{II.18}$$

where the parameters α_i and β_i represent the constitutive synthesis and degradation rate respectively. The vector product $\mu.e_i$ is the dilution rate of enzymes due to growth. The term $r_{EM,i}$ is the inducible synthesis rate of "enzymes" regulated by the cybernetic variable u_i as,

$$r_{EM,i} = u_i.k_{E,i}^{\max}.b.r_{M,i}^{kin} \tag{II.19}$$

where parameter *b* represents the fraction of biomass that is not used for other intracellular metabolites (biomass for enzyme synthesis), which is only true if constitutive enzyme synthesis rate is much slower than inducible synthesis rate $\alpha_i > r_{EM,i}$, and is defined by

$$b = 1 - \sum_i m_{S\,i} \tag{II.20}$$

where the term $m_{S\,i}$ represent the accumulated metabolites. Finally, the parameters $k_i^{\max}$ and $k_{E,i}^{\max}$ from equations (II.16) and (II.19) are the maximum reaction rate of metabolites and "enzyme" synthesis of the i^{th} mode.

The formulation of cybernetic variables is stated by the proportional law (II.21) and matching law (II.22) based on the return on investment (ROI) p_i (II.23) of the metabolic objective function. This ROI considers the distribution of uptake fluxes among EMs derived from the synthesis of "enzyme" e_i, which subsequently catalyzes all the reactions involved in each i^{th} elementary mode (Song and Ramkrishna, 2010; Young and Ramkrishna, 2007). The most common choice of (ROI) in cybernetic models were the maximization of the carbon uptake rate (Song *et al*., 2009) for each i^{th} mode

$$v_i = \frac{p_i}{\|p_i\|_\infty},\ v_i \le 1 \tag{II.21}$$

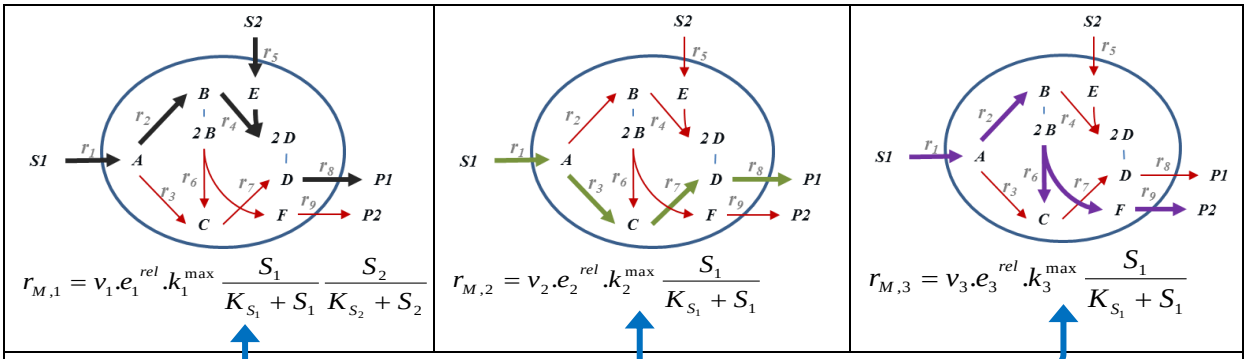
$$u_i = \frac{p_i}{\|p_i\|_1},\quad \sum_i u_i = 1 \tag{II.22}$$

$$p_i = f_{C,i} e_i^{rel} r_{M,i}^{kin} \tag{II.23}$$

where $f_{C,i}$ is the carbon number per unit mole of substrate consumed through each mode.

One example of how HCM relates the cybernetic "enzymes" is illustrated in Table II.5. Let us consider the stoichiometric matrix from Table II.3 and the three elementary modes that were identified. These EMs are represented in Table II.5 as the black, green, and purple pathways. Following (II.16) and (II.17), the flux vector $r_{M,i}$ could be defined by the uptake of substrates expressed by Monod-type equation, and a regulatory part regarding the cybernetic variable v_i and the enzyme level whose dynamics were presented on (II.18). The solution for the cybernetic enzymes gives then a dynamic weight to each EM. Finally, the flux vector is included in the mass balances of the corresponding extracellular metabolite (*i.e.* substrates, products, and biomass).

Table II.5. Example of application of HCM.



$\frac{d}{dt}S_1 = S_{S_1}Z.(r_{M,1}+r_{M,2}+r_{M,3}).X$	Three EMs
$\frac{d}{dt}S_2 = S_{S_2}Z.r_{M,1}.X$	One EM
$\frac{d}{dt}P_1 = S_{P_1}Z.(r_{M,1}+r_{M,2}).X$	Two EMs
$\frac{d}{dt}P_2 = S_{P_2}Z.r_{M,3}.X$	One EM
$\frac{dX}{dt} = S_X Z(r_{M,1}+r_{M,2}+r_{M,3}).X$	

HCM has been applied for the aerobic growth of *Escherichia coli* on glucose (Kim *et al*., 2008). However, it has also been implemented in the modeling of cultures where microorganisms co-consume several substrates, such as glucose and xylose by recombinant yeasts (Song *et al*., 2009); glucose and glutamate by mammalian cells (Geng *et al*., 2013); and mannose, galactose, glucose, and xylose by three yeasts *Saccharomyces cerevisiae*, *Pichia stipites*, and *Kluyveromyces marxianus* (Geng *et al*., 2012).

The HCM methodology, has been extended for the production of accumulated metabolites (*i.e.* PHA) by *Ralstonia eutropha* (Franz *et al*., 2011). This was accomplished by the addition of an assumption implying that the internal metabolites could be divided into two subgroups according to their dynamics: m_f and m_s fast and slow dynamics respectively. Thus, equation (II.4) can be redefined as,

$$\frac{d}{dt}\begin{bmatrix} m_f \\ m_s \end{bmatrix} = \begin{bmatrix} 0 \\ S_{m_s}.r - \mu.m_s \end{bmatrix} \quad \text{(II.24)}$$

For m_f, the quasi-steady state assumption holds and the reduction is done as in (II.5). Otherwise, m_s that represents intracellular accumulated metabolites (with slow dynamics) maintaining the expression of (II.4) with S_{m_s} the stoichiometric matrix of intracellular metabolites associated to the slow dynamics.

HCM has also been coupled with model predictive control (MPC) for the optimization of ethanol from lignocellulose (Wong *et al*., 2010). Simulations demonstrated that the use of HCM enabled to conduct accurate *in silico* experiments.

II.3.4. Lumped Hybrid Cybernetic Model (L-HCM)

Lumped Hybrid Cybernetic Model was introduced by Song and Ramkrishna, (2010). This modeling framework is an extension, of HCM in which all the equations presented in section II.3.3 hold. The difference with HCM and the essence of L-HCM relies in getting the L-EMs. To do so, the EMs are classified into different families according to their commonalities (*e.g.* substrates shared among EMs). In each family, EMs are subdivided into biomass producing group (including both biomass and ATP-producing modes); and ATP-only producing group (referred to as B- and A- groups, hereafter).

Flux distribution at EM family level can be described using the HCM framework by representing the flux vector *r* as follows,

$$r = Z_F r_F \quad \text{(II.25)}$$

where r_F is the vector of n_F fluxes through the lumped elementary modes and Z_F is $(n_r \times n_F)$ L-EM matrix. In an equivalent form r_F can be expressed as r_M. The lumped HCM considers that the J^{th} column vector of the L-EM matrix Z_F is given by the following lumping rule

$$Z_{F,J} = w_J.z_{B,J} + (1 - w_J).z_{A,J} \quad \text{(II.26)}$$

where the parameter w_J is determined such that the Growth rate dependent on ATP Requirement (GAR) is satisfied. $z_{B,J}$ and $z_{A,J}$ denote the weighted average of the EMs

producing biomass (and both biomass and ATP) (Group B) and ATP (Group A), respectively in the J^{th} family, and are calculated as,

$$z_{B,J}\left(\text{or } z_{A,J}\right)=\frac{\sum_{j\in L(J_B)} z_{B,J}\eta_J^{n_{v+2}}}{\sum_{j\in L(J_B)}\eta_J^{n_{v+2}}} \tag{II.27}$$

The lumping of EMs is also reflected in the computation of the return on investment (ROI) p (II.23), which is redefined within the J^{th} family as $p_J=\eta_J$. Under this regard, L-HCM assumes that the non-regulated rates $r_{M,J}^{kin}$ for each J^{th} family are proportional to a sub ROI (sROI) ($r_{M,J}^{kin}\propto\eta_J$), which is in turn given as the yields of vital products.

$$\eta_J=\begin{cases}Y_{B,J}\\ Y_{A,J}\end{cases} \tag{II.28}$$

And therefore, $r_{M,J}^{kin}$ from equation (II.17) becomes,

$$r_{Mi,J}^{kin}=k_{i,J}^{\max}\eta_J\prod_i\frac{M_i}{K_{M,i}+M_i}\left(i=1,\ldots,n_Z\right) \tag{II.29}$$

where the calculation is performed per i^{th} EMs in the J^{th} family. sROI is included to distinguish the two subgroups.

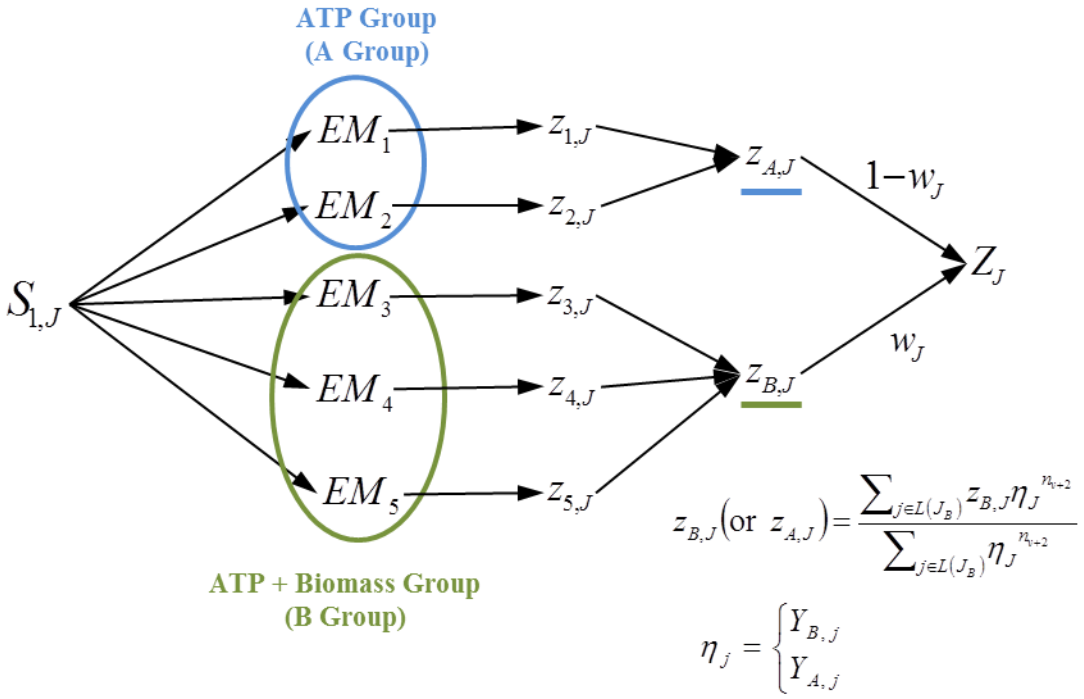


Figure II.4. Lumping of Elementary modes in the Lumped Hybrid Cybernetic Model (L-HCM).

The methodology for lumping inside a J^{th} family is displayed on Figure II.4, where the Elementary Modes (EMs) are classified into two subgroups to be further weighted to have a unique elementary mode Z_J.

In summary, the methodology can be summarized in several steps (Song and Ramkrishna, 2011).

(i) Distribution of uptake flux is regulated at two different levels, that is, first among lumped pathways (EM families) and next among individual EMs.

(ii) The regulation is split at both levels and described by the cybernetic laws.

(iii) Uptake fluxes through the lumped pathways are treated as independent reactions, while uptake fluxes through EMs within the same family are assumed to be proportional to η_J.

The number of EM families become identical to the number of EMs if they are classified according to the stoichiometry of all fluxes (intracellular and exchange) associated therein. In this case, L-HCM is equivalent to the original formulation of HCM (Song and Ramkrishna, 2011).

Applications of L-HCM concern the production of ethanol by *Saccharomyces cerevisiae* from glucose (Song and Ramkrishna, 2010); the anaerobic growth of different strains of *Escherichia coli* (Song and Ramkrishna, 2011) with the extension of the model to the prediction of mutant strains behavior (Song and Ramkrishna, 2012); and the triauxic growth of the microalgae *Shewanella oneidensis* (H.-S. Song et al., 2013).

The main advantage of this framework is the reduction of the number of EMs for model application (and parameter reduction), in which only one EMs is considered by family. However, L-HCM requires the estimation of two other parameters with respect to HCM, w_J and n_v.

II.3.5. Dynamic Reduction of Unbalanced Metabolism (DRUM)

This methodology was developed for unbalanced metabolisms, as microalgae (Baroukh *et al*., 2014). Here the equations of balances consider a metabolic network to determine the stoichiometric coefficients. Differently than the other approaches, DRUM assumes that QSSA is applicable only to groups of metabolic reactions called sub-networks (*SNs*). Therefore, there are some metabolites (*A*) interconnecting the *SNs*, which are not under the quasi-steady state constraint, and can, thus, accumulate.

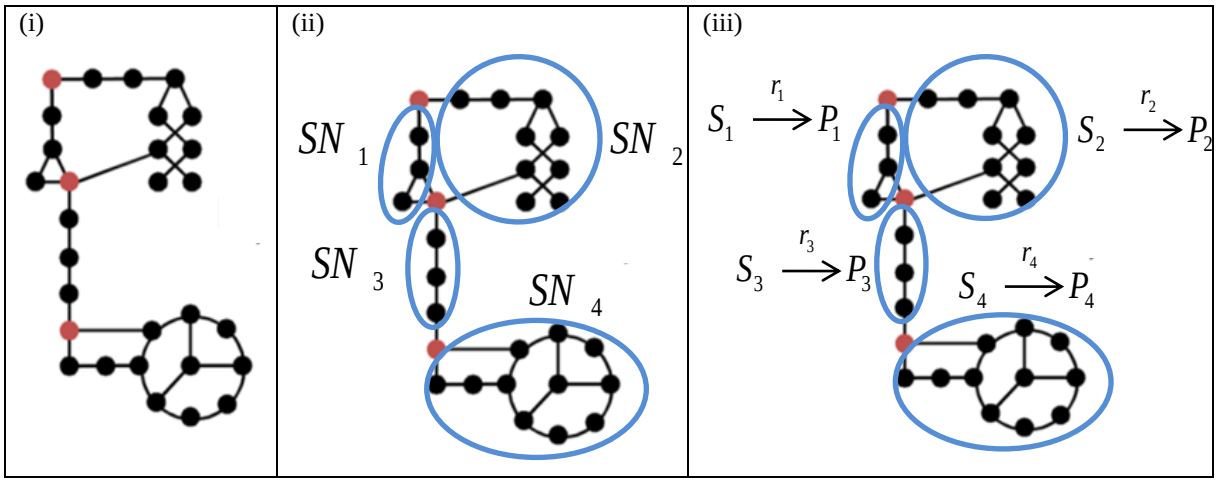
The methodology follows three steps (Table II.6):

(i) Identification of the complete network.

(ii) Decomposition of the network into to subnetworks (*SNs*) assumed at quasi-steady state, where the red points are the accumulated metabolites (*A*).

(iii) Reduction of the subnetworks into macroscopic reactions. Kinetics are defined for each macroscopic reaction.

Table II.6. DRUM approach (Taken from (Baroukh *et al*., 2014)**).**



In DRUM approach, equations (II.11) and (II.12) can be extended for each subnetwork as follow,

$$\frac{dM}{dt} = (S_M Z)_{SN} . r_{SN} . X \qquad \text{(II.30)}$$

$$\frac{dX}{dt} = (S_X Z)_{SN} . r_{SN} . X \qquad \text{(II.31)}$$

$$\frac{dA}{dt} = (S_A Z)_{SN} . r_{SN} . X \qquad \text{(II.32)}$$

where the $S_M Z$ $(n_M \times n_Z)$, $S_X Z$ $(n_X \times n_Z)$, and $S_A Z$ $(n_A \times n_Z)$ are the matrices of the macroscopic reactions comprising the set of subnetworks the k^{th} subnetworks *SN* as,

$$(S_M Z)_{SN} = \left[(S_M Z)_{SN,1} \quad \dots \quad (S_M Z)_{SN,k}\right] \tag{II.33}$$

The kinetics r_{SN} are defined by subnetworks instead of being describing per elementary mode. Special attention must be payed to biomass because in this approach due to the accumulation of certain metabolites (*A*). Therefore, not all the elements end up in biomass or excreted products. Total biomass X_T must be determined by the sum of the accumulated metabolites to the biomass X ($X_T = A + X$). The authors of DRUM approach suggest doing this sum per chemical element (*i.e.* C; N; O; H; P; S….) to have a more precise biomass composition.

This modeling approach was applied to the microalgae *Tisochrysis lutea* in day/night cycles (Baroukh *et al*., 2014). It demonstrated to be effective on describing the metabolism microalgae predicting accurately the experimental concentrations of intracellular metabolites (lipids, carbohydrates and chlorophyll), as well as the concentration of the extracellular metabolites (substrates, biomass). Nevertheless, DRUM highly depends on the selection of the branching points, which could affect the biomass composition and the flux distribution. This decomposition into subnetworks was performed by trial and error, which was time consuming and limits the further extension of the model. Clustering methods or additional optimization methods should be proposed to overcome this problem.

III. CONTROL OF BIOPROCESSES

Fermentations involve the use of microorganisms (bacteria, yeast, fungi, *etc.*) to transform substrates and nutrients into added-value goods (Dochain, 2008). The growth of those microorganisms depends on the preservation of the proper environmental conditions. However, fermentations are highly interactive systems where small fluctuations can influence the overall performance of process (Yamuna Rani and Ramachandra Rao, 1999).

Over the last years, several control approaches have been proposed to maintain the fermentations under optimal operating conditions in spite of disturbances that can affect its proper operation (Alford, 2006). The control approaches used from chemical processes, such as the Proportional Integral Derivative (PID) controller have been extrapolated for fermentations. PID controllers were designed for processes with linear dynamics; therefore,

their utilization on fermentations (nonlinear processes) require the adaptation of parameters to the changing process dynamics (Simutis and Lübbert, 2015). Moreover, control of bioprocess is challenging because the accurate models are rarely available or limited, they are usually over parametrized and there is lack of online sensors.

With the aim of defining a proper control strategy for lipid accumulation by *Yarrowia lipolytica*, two important aspects should be considered: (i) fed-batch cultures (see CHAPTER I), (ii) concomitant production of citric acid (overflow) (I. R. Sitepu et al., 2014). Therefore, special interest is paid to fed-batch fermentations due to its industrial applicability and its capability of managing the substrate additions. Control strategies on these cultures with overflow have been mainly focused on the definition of those substrate feeding profiles (Hjersted and Henson, 2006). These profiles are quite sensitive because they have important impacts on the performance of the microorganism, and thus in the bioprocess itself, where a bad estimation can lead to starvation, formation of undesirable products (overflow), and inhibition. This can be seen as a dynamic optimization problem in which a final time objective is maximized while it is subject to constraints imposed by dynamic model equations. Therefore, model based optimization is an essential tool for determining fed-batch operating strategies.

However, when no model is available, the operating conditions are often optimized by the use of the "knowledge" acquired from previous experiments and highly skilled operators. This "knowledge" must be translated to a quantitative form to be used. Some approaches, as fuzzy sets theory (Zadeh, 1965) has dealt with the adaptation of linguistic actions to mathematical equations through if-then rules. This type of control has been applied to the baker's yeast fed-batch fermentation to optimize biomass concentration by regulating the glucose feeding rate (Hisbullah *et al*., 2003; Kasperski and Miśkiewicz, 2002) with the respiratory quotient (carbon dioxide evolution rate by oxygen uptake rate) as an indicator of the metabolic behavior of the yeast *Saccharomyces cerevisiae*. Other works focused on the alcoholic fermentation controlled by the substrate feeding rate. The type-2 fuzzy controller proved to be an effective solution to control systems in the presence of parameter uncertainties (Galluzzo and Cosenza, 2010). Fuzzy strategies have also been coupled to optimal control and predictive control strategies (Jin *et al*., 2014).

A summary of applications on fermentation process with overflow are presented in Table II.7. In order to provide some insights about the applied model-based control algorithms, their description is addressed in the following subsection.

Table II.7. Control applications to fermentations.

Type of Control	**Application**	**Manipulated Variables**	**Objective(s)**	**Actions / Method**	**Reference**
Adaptive fuzzy logic	Continuous Alcoholic fermentation by *Saccharomyces cerevisiae*	Dilution rate	Improve ethanol concentration	Tracking glucose concentration	(Galluzzo and Cosenza, 2010)
Adaptive neuro fuzzy inference system	Chlortetracycline production by *Streptomyces aurefaciens*	Substrate feeding	Improve final titer of Chlortetracycline	Tracking Substrate concentration (%)	(Jin *et al*., 2014).
Adaptive fuzzy logic	Baker's yeast fermentation from glucose	Glucose feed rate	Maximize biomass concentration	Tracking difference of carbon dioxide evolution rate and oxygen evolution rate	(Hisbullah *et al*., 2003)
Adaptive	Alcoholic fermentation	Glucose feed rate	Maximize ethanol production	Control of DOT CO_2, alcohol or RQ	(Williams *et al*., 1986)
Adaptive	Continuous alcoholic fermentation by *Saccharomyces cerevisiae*	Dilution rate	Improve ethanol concentration	Regulate an track substrate concentration	(Maher *et al*., 1995)
Adaptive	Overflow metabolism on Baker's yeast fermentation	Substrate feed rate	Maximize biomass	Maintain ethanol at low level	Hocalar and Türker, (2010)
Optimal	Mammalian cells cultures	Glucose and glutamine feed rates	Maximization of biomass productivity	Avoiding the waste of substrates	(Amribt *et al*., 2013)

				(glucose and glutamine) via the overflow metabolism	
Type of Control	**Application**	**Manipulated Variables**	**Objective(s)**	**Actions / Method**	**Reference**
Optimal	Optimization of penicillin from glucose	Substrate feed rate	Maximize Penicillin productivity	Orthogonal collocation on finite elements	(Riascos and Pinto, 2004)
	PHA maximization from glucose and fructose with limiting propionic acid	Glucose, Xylose, and nitrogen feed rates	Maximize PHAs concentration		
Optimal	Biodiesel production from triacylglycerol in batch cultures	Temperature	Maximize concentration of methyl ester (biodiesel)	Temperature control by maximum principle	(Benavides and Diwekar, 2012);
Optimal	Xylitol production by *Candida mogii*	Glucose and Xylose feed rates	Maximization of xylitol production	Find optimal feed rate by Genetic algorithms	(Tochampa *et al*., 2015)
Multi objective optimization	Ethanol production from glucose	Glucose feed rate	Maximization of ethanol productivity and yield	dynamic flux balance models	(Hjersted and Henson, 2006).
Optimal Adaptive	Penicillin production	Substrate flow rate	Maximization of penicillin concentration	Biphasic control with state estimation	(Van Impe and Bastin, 1995)
Optimal Adaptive	Baker's yeast fermentation	Glucose flow rate	Maximization of biomass	Low ethanol concentration	(Valentinotti *et al*., 2003).

Model predictive	Penicillin fermentation	Water flow rate Glucose flow rate Acid and base additions	Maximize penicillin concentration	Temperature and pH control with an optimal feeding rate	(Ashoori *et al*., 2009)
Type of Control	**Application**	**Manipulated Variables**	**Objective(s)**	**Actions / Method**	**Reference**
Model predictive	Fermentation by *Escherichia coli* for vaccines	Glucose flow rate	Maximize biomass production	Maximize glucose consumption rate to avoid acetate (overflow metabolism)	(Santos *et al*., 2012)
Multi objective optimization	Producing α-amylase by *Zymomonas mobilis*	Optimal temperature, pH, time, and inoculum size	Maximize the amylase production and its activity	MOPSO MO Genetic Algorithms	(Rajulapati and Narasu, 2011)
Robust adaptive	Aerobic cultures of *Saccharomyces cerevisiae*	Glucose feed rate	Maximize biomass productivity	Low ethanol concentration	(Renard and Vande Wouwer, 2008)

III.1. Model based Control

The most applied model based control approaches include adaptive control, optimal control, optimal adaptive control, and model predictive control, which are briefly described in the following subsections.

III.1.1. Adaptive control (Linearizing control)

Adaptive control is used when a series of parameters is variable at any time of the process. Based on a well identified model, this control assumes that the process is linear in a short period of time. Hence, the principles of linearizing control are used to find a tracking control law. The

tracking error is governed by a pre-specified stable linear differential equation called a reference model (Bastin and Dochain, 1990). The main objective is thus to track a reference output signal.

This controller has two main steps: parameter estimation and control law (Williams *et al*., 1986). Parameter estimation is performed to adapt the system to the real operation.

For the sake of illustration, let us express the mass balances system (II.1) in a simplified from as,

$$\frac{d(\xi.V)}{dt} = K.r(\xi,t) + F_{IN}\xi_{IN} - F_{OUT}\xi \tag{II.34}$$

Here, the uptake rate vector $r(\xi,t)$ is function of the state vector ξ and the time *t*. Without loss of generality, the adaptive controller of a continuous culture can be deduced from the dynamical system (II.1), where the objective is to control a scalar output variable *y* which is a measured linear combination of state variables

$$y = \sum_{i=1}^{N} c_j.\xi_j = C^T\xi \tag{II.35}$$

where $C^T = [c_1, c_2, \ldots, c_N]$ is a vector of known constants. The control input *u* is usually the feed rate of one external substrate of the process which is denoted as,

$$\begin{aligned} &u = F_{IN} \\ &F = b.u + f \quad b^T = [b_1, b_2, \ldots, b_N] \;\; b_i = 1,\; b_j = 0 \quad \forall j \neq i \\ &\qquad\qquad f^T = [f_1, f_2, \ldots, f_N] \;\; f_i = 1,\; f_j = F_j \;\; \forall j \neq i \end{aligned} \tag{II.36}$$

Then the dynamics of the output variable *y* can be expressed as follows,

$$\frac{dy}{dt} = C^T\frac{d\xi}{dt} = C^T K.r(\xi,t) + C^T b.u + C^T f. - F_{OUT}\xi \tag{II.37}$$

In order to track the error, a reference linear model can be proposed in the form

$$\frac{d}{dt}(y^* - y) + \lambda_1(y^* - y) = 0 \tag{II.38}$$

where y^* is the desired value of the output variable. In this context, the input value *u* can be calculated from (II.37) and (II.38) to obtain,

$$u(t) = (C^T b)^{-1}\left[\frac{dy^*}{dt} + \lambda_1(y^* - y) - C^T K.r(\xi,t) + F_{OUT}.\xi - C^T.f\right] \tag{II.39}$$

Equation (II.39) assumes that the vector $K.r(\xi,t)$ is known, which normally, is not the case. Therefore, (II.39) is transformed into an adaptive control law by using single perturbation methods and/or parameter estimators (Bastin and Dochain, 1990). In this case, the uptake rate vector $r(\xi,t)$ is replace by an on line estimation $\hat{r}(\xi,t)$.

Notwithstanding, when disturbances are presented on the systems high errors between the setpoint and the measured values could exist. These errors are related to the assumed linearity of the process that could lead to the tracking of failures instead of the adaptation to disturbances (Van Impe and Bastin, 1995).

Adaptive control has been applied to alcoholic fermentations (Williams *et al*., 1986) to control CO_2 levels, RQ, alcohol concentration with PI controller with variable gains (Maher *et al*., 1995), *etc*. Recently, Hocalar and Türker, (2010) proposed an adaptation with a feedback action to the overflow metabolite on baker yeast's fermentation. The authors used a set point value of 0.1% (1g.L^{-1}) of ethanol (measured in the gas phase), which was efficiently followed by the control law coupled to a PID with variable gains.

III.1.2. Optimal control

Optimal control provides useful information for designing and controlling reaction process. The most common cases of optimal control search to find the best yield of conversion or productivities at a given final time (Riascos and Pinto, 2004).

A general optimal control problem can be defined from a given system

$$\frac{d\xi}{dt} = f\left(\xi(t), u(t), \theta, t\right) \tag{II.40}$$

where $\xi(t) \in R^n$ represent the state variables at time *t*, $u(t) \in R^m$ the system input or control variables and θ the vector of kinetic parameters. For the sake of simplicity $\xi(t)$ will be denoted as ξ, and $u(t)$ as u. There exists though an objective function *J* to be minimized subject to several constraints as displayed in the following equations.

$$\min_{u(t)} J \tag{II.41}$$

$$\text{Subject to} \begin{cases} \dfrac{d\xi}{dt} = f(\xi, u, \theta, t) \\ g_i(\xi) \le 0 \qquad i = 1,2,\ldots,m \\ h_i(\xi) = 0 \qquad i = 1,2,\ldots,p \\ \xi^L \le \xi \le \xi^U \\ u^L \le u \le u^U \end{cases} \tag{II.42}$$

where $g_i(\xi)$ represents the *m* inequality and $h_i(\xi)$ the *p* equality constraints. The super index *L* and *U* hold for the lower and upper limits of the states ξ and the inputs u of the system. Some examples of the used objective functions in fermentation are displayed on Table II.8, where it is seen that productivity and final concentration, where the most used objectives. However, some of them could be enhance together thanks to multi-objective optimization (see III.3)

Table II.8 Comparison of objective functions.

Objective function	Expression	Reference
Yield	$J = \dfrac{P(t_f)}{\int_0^{t_f} F(t) S_F \, dt}$	(Sarkar and Modak, 2005)
	$J = \dfrac{P(t_f)}{F_0 S_0 + \int_0^{t_f} F(t) S_F \, dt}$	(Hjersted and Henson, 2006)
Productivity	$J = \dfrac{P(t_f)}{t_f}$	(Sarkar and Modak, 2005)
	$J = \dfrac{1}{t_f} \left\| \dfrac{V_f X_f - V_0 X_0}{S_{IN}(V_f - V_0)} \right\|$	(Valentinotti et al., 2003)
	$J = \dfrac{P(t_f) V(t_f)}{t_f}$	(Tochampa et al., 2015)
Final concentration of Product	$J = P(t_f)$	(Smets et al., 2004)
Final amount of product	$J = P(t_f) V(t_f)$	(Riascos and Pinto, 2004; Van Impe and Bastin, 1995)
Maximize actual growth rate	$J = \dfrac{\mu(t) - \mu^{\max}}{\mu^{\max}}$	(Giuseppin and van Riel, 2000)

There exists two methods to solve (II.41), the indirect and direct search methods. The indirect methods rely on an appropriate mathematical representations of the systems where second order differential equation or partial differential equations are required for the optimization at every time step (Diego *et al*., 2000; Riascos and Pinto, 2004). Some attempts have been used to solve this hurdle, such as the Pontryagins minimum principle to minimize the Hamiltonian over all admissible control functions. The minimization of the Hamiltonian is simple since the control variable *u* appears linearly on it (Benavides and Diwekar, 2012).

On the other hand, direct methods are methods that do not use derivatives or approximations to solve the optimization problem (Kirgat and Surde, 2014). Two examples of this methods are the Pattern Search algorithm (Lewis and Torczon, 1996) and Particle Swarm Optimization (Eberhart and Kennedy, 1995) which will be explained in the next subsection.

Optimal control has been preferably applied to fed-batch cultures, where the definition of an optimal substrate feeding is aimed. Examples include the optimization of biodiesel production from triacylglycerol (Benavides and Diwekar, 2012); the maximization of biomass cell productivity of hybridoma cells under overflow metabolism of glutamate and glucose (Amribt *et al*., 2013); the optimization of penicillin from glucose, and PHA maximization from glucose and fructose with limiting propionic acid (Riascos and Pinto, 2004); the optimization of xylitol production with two input variables (glucose –xylose mixtures) (Tochampa *et al*., 2015); ethanol productions from glucose with dynamic flux balance models (Hjersted and Henson, 2006).

The performance of this type of control depends on the model quality, by example the well identification of the kinetic parameters θ of the model. This control approach is hampered in real life by the availability of sensors for process monitoring.

III.1.3. Optimal adaptive control

The optimal adaptive control approach was created to solve some hurdles generated by adaptive control, where the model uncertainties could provoke a failure in the controller. If models are not available, optimality is not guaranteed (Van Impe and Bastin, 1995).

This approach is implemented in three steps. The first assumes that a well identified model is available, and thus an optimal solution is found as described in the previous subsection. In a second step, a nearly optimal heuristic controller is proposed based on the knowledge of the process and the mathematical analysis of the optimal solution from the optimal control (Bastin *et al*., 1995; Smets *et al*., 2004). The solution of this optimal controller is a reference trajectory that can be tracked by an adaptive controller through nonlinear programming. Depending on the available online measurements, this type of control has considered the inclusion of state observers as soft sensors (Smets *et al*., 2004).

The optimal adaptive control has been used for the optimization of feeding profiles of two-phase fed-batch cultures regarding a growth phase and a production phase (Yamuna Rani and Ramachandra Rao, 1999). That is the case of penicillin production (Van Impe and Bastin, 1995) studied when the volume, and the concentrations of substrate, biomass, an penicillin were available. Another case was the maximization of baker's yeast concentration where the overflow metabolite phenomenon (ethanol production) occurred (Valentinotti *et al*., 2003).

III.1.4. Model Predictive Control

The main idea of Model predictive control (MPC) is to choose control actions by solving an online constrained optimization problem at each sample time. MPC can be seen as the compilation of three components (Christofides *et al*., 2013):

1. The Model. This is used to predict the future evolution of the system. The accuracy of the optimization highly depends on the well identification of the model.
2. A performance index over a finite horizon. This will be minimized subject to constraints imposed by the system and other considerations at each sampling time to obtain a trajectory of future control inputs.
3. A receding horizon scheme. This introduces feedback into the control law to compensate for disturbances and modeling errors.

MPC aims at minimizing a performance index over a finite prediction horizon *N* based on prediction obtained by a system model for $\xi(t_k)$. A standard MPC can be formulated as follows,

$$\min_{u(t)} J = \sum_{i=1}^{m} \int_{t_k}^{t_{k+N}} \left(\xi^T . Q . \xi + u_i^T . R . u_i \right) d\tau \qquad \text{(II.43)}$$

$$\text{Subject to} \quad \begin{cases} \dfrac{d\xi}{dt} = f(\xi, u, \theta, t) \\ u^L \le u \le u^U \end{cases} \tag{II.44}$$

Q and *R* are strictly positive definite symmetric weighing matrices used to trade-off setpoint tracking and manipulated variable movement, respectively. The control use the knowledge of the system model $\xi(t_k)$ at a moment t_k to design an input sequence $u_i(t|t_k)$, i=1,…m, for $t \in [t_k, t_k + N]$. The first step values are applied to the closed-loop system for $t \in [t_k, t_k + 1]$. At the next sampling time $t_k + 1$, a new value of $x(t_k + 1)$ is obtained and the process is repeated (Ashoori *et al*., 2009).

This approach is useful for handling nonlinear and time varying problems since the controller is a function of the model that can be modified in real time. MPC has been applied to the control of chemical reactors (Christofides *et al*., 2013), to the optimization of penicillin fermentation (Ashoori *et al*., 2009), overflow metabolite on *Escherichia coli* to prevent acetate formation (Santos *et al*., 2012), *etc*.

III.1.5. Piecewise functions to parametrize control

Piecewise functions are generally used to approximate the inputs of control, here denoted by *u*. Each component u_j is approximated by spline functions that are defined on a series of differently spaced knots with duration τ_j in time interval (0,*T*]. Therefore, the control component can be written as,

$$u(t) = \sum_{j=1}^{n} \sigma_j B_j^{(m)}(t) \tag{II.45}$$

where $B_{ij}^{(I)}$ is the indicator function of order m, with a set of j^{th} knots, which can be defined recursively. For $m = 0$, the system is defined by piecewise constants functions

$$B_j^{(0)}(t) = \begin{cases} 1 & \tau_j \le t < \tau_{j+1} \\ 0 & otherwise \end{cases} \tag{II.46}$$

Otherwise, for $m = 1$ the system is said to be piecewise linear (Schlegel et al., 2005) with

$$B_j^{(m)}(t) = \frac{t-\tau_j}{\tau_{j+m-1}-\tau_j} B_j^{(m-1)}(t) + \frac{\tau_{j+m}-t}{\tau_{j+m}-\tau_{j+1}} B_{j+1}^{(m-1)}(t) \quad \text{(II.47)}$$

The term σ_j corresponds to the parameters. The dimension the required splines is equal to *n*. For a piecewise constant representation with *l* interval, we need $n = l$ splines with $m = 0$, whereas a piecewise linear parametrization would require $n = l + 1$ splines with $m = 1$.

III.2. Optimization Algorithms

Optimization algorithms can be classified in two groups: local and global algorithms. The first assumes the convexity of the cost function, which is not often fulfilled, and thus the algorithms are trapped into local minima (Donoso-Bravo *et al*., 2011). To overcome this hurdle, a multi-start strategy was proposed to decide when to allocate additional resources to a particular starting condition (György and Kocsis, 2011).

Many local methods determine their direction of search in the parameter space based on the gradient and Hessian of the objective function at the present point in parameter space (Almquist *et al*., 2014). These methods are called gradient based methods. Some of these methods are: (i) The Levenberg-Marquardt method (Marquardt, 1963), which approximates the hessian through the iterative search of the Jacobian matrix. (ii) the sequential quadratic programing algorithm (SQP) (Diego *et al*., 2000) that solves a quadratic problem for each iteration; and (iii) Multiple shooting (Peifer and Timmer, 2007) that discretizes the time into time intervals. However, the use of these optimization algorithms is complicated in practice because the calculation of second order derivatives at each optimization time is time consuming and the methods do not guarantee global convergence to the solution. Even if some research has been done to improve the hessian matrix, these methods has been overpassed by the direct search methods (Kirgat and Surde, 2014).

Direct search methods are derivative free algorithms mostly based on populations (Kirgat and Surde, 2014; Lewis and Torczon, 1996).

Some popular methods based on global minimization routines are Simulated Annealing (SA), Evolutionary Algorithms (EA), Particle Swarm Optimization (PSO) (Almquist *et al*., 2014), and Pattern Search algorithms (Vaz and Vicente, 2007). Additionally to the global convergence, these algorithms do not require a differentiable objective function, and they can explore a large space of solutions. Although these algorithms produce better solutions, they do not guarantee that an optimal solution will ever be found at the price of large amounts of computation (Donoso-Bravo *et al*., 2011). Insights of these algorithms will be provided hereafter.

Particle Swarm Optimization and Pattern Search were employed along this thesis to perform parameter estimation and control because of their simplicity, their faster localization of optima, and their high directional search (Angeline, 1998). Therefore, these two optimization algorithms are described in more detail in the next subsections.

III.2.1. PSO

Particle swarm optimization (PSO) is a population-based stochastic technique, developed by Eberhart and Kennedy in 1995 (Eberhart and Kennedy, 1995). This algorithm is based on the behavior of biological communities that exhibit both individual and social behavior; examples of these communities are flocks of birds, schools of fishes and swarms of bees. Members of such societies share common goals that are realized by exploring its environment while interacting among them.

The popularity of PSO is due in part to the simplicity of the algorithm, but mainly to its effectiveness for producing good results at a very low computational cost (Trelea, 2003). In PSO each particle *p* represents a potential solution of the optimization problem. Moreover, all the particles have a fitness value which is evaluated by the fitness function to be optimized and have velocities *v* to direct their fly (Eberhart and Yuhui Shi, 2001).

PSO is initialized with a group of random solutions that are swarm through the hyper dimensional search space. For each iterate, the particles are updated to have two best values. The first is the best solution of the problem (*pBest*). The second correspond to the best value obtained by any particle in the entire swarm (*gBest*). Usually, the *gBest* is selected as the leader particle to direct the swarm. In addition, when a particle takes part of the population as its topological neighbor, the best value is a local best (*lbest*).

The changes of the position of particles follow the leader in a way of social tendency of the individuals to emulate the success of other individuals. The velocity vector of the search is calculated as,

$$v(i) = w \times v(i) + C_1 \times r_1 \times (pBest(i) - p(i)) + C_2 \times r_2 \times (gBest(i) - p(i)) \quad \text{(II.48)}$$

Where r_1 and $r_2 \in [0,1]$ are random values. C_1 and C_2 are the learning factors representing the attraction that a particle has towards its own success. The term *w* is the inertia weight employed to control the impact of the previous history of velocities on the current velocity (Eberhart and Shi, 2000). However, this velocity vector is usually bounded to a maximum velocity V_{max} (Tuppadung and Kurutach, 2011). The position of each particle is changed according to its own experience and that of its neighbors. If *p*(*i*) is the position of the particle *p*. The position of *p*(*i*) is then changed by adding a velocity *v*(*i*) to the current position.

$$p(i) = p(i) + v(i) \quad \text{(II.49)}$$

The algorithm then performs a series of steps to find the best position *gBest* that satisfy the fitness criteria or that was found after a fixed number of iterations.

III.2.2. Pattern search

Pattern search algorithms are feasible point methods with global convergence (Vaz and Vicente, 2007). The search is based on a series of exploratory moves around the current iterate x_j to choose a new iterate $x_{j+1} = x_j + s_j$ for some feasible steps s_j. It is worth noting that pattern search does not need the calculation of derivatives at each iterate x_j. The selection of the steps $s_j = \Delta_j P_j$ is determined by exploratory moves comprising a pattern P_j, and a step length Δ_j.

The search begins with a point $x_j \in R^n$ in the j^{th} iteration that satisfies the bounds and maintains feasibility throughout the search. The search for an optimum is done successively at the points $x_{j+1} = x_j + s_j$ until a x_{j+1} is found for which $f(x_j + s_j) < f(x_j)$. Otherwise, Δ_j is reduced by λ, so that $\Delta_{j+1} = \lambda.\Delta_j$ and $x_{j+1} = x_j$. These exploratory moves must produce a feasible step that also gives simple decrease on the function value at the current iterate (Lewis and Torczon, 1996).

III.3. Multi-objective Optimization

Even if the control approaches presented in the last section have been successfully applied, biotechnological processes optimization often involved multiple objectives, *i.e.* minimize cost, maximize performance, maximize reliability, *etc.* Nonetheless, most of the objectives could be competing to each other hampering the optimal operation of the processes. In fed-batch fermentation, one of the most common conflicts regards the maximization of product-substrate yield and the productivity at a final time.

According to equation (II.41), the objective function can be extended for multi-objective optimization as,

$$\min_{u(t)} J = \left[f_1(\xi, u, \theta, t),\ f_2(\xi, u, \theta, t), \cdots,\ f_n(\xi, u, \theta, t) \right] \tag{II.50}$$

where *n* is the number of objectives to be optimized. Without loss of generality, all objectives are of the minimization type (maximization can be converted to minimization by multiplying by negative one) (Marler and Arora, 2004). The space in which the objective vector belongs is called the objective space, and the image of the feasible objective vector belongs is called the objective space, and the image of the feasible set under *J* is called the attained set.

There are two general approaches used to solve multiple objective optimization (Coello Coello, 1999). One is to combine the individual objective functions into a single composite function or move all but one to the constraint set. The determination of a single objective is possible with methods such as utility theory, weighted sum method, *etc.* (Marler and Arora, 2004). However, the problem lies on the proper selection of the weights or utility functions to characterize the decision maker's preference which is complicated in practice. Moving the objective functions to the constraint set need the definition of a constraining value for each of these former objectives. The solution is then arbitrary and unique leading to the optimization without trade-offs between objectives.

A second general approach is to determine a set of optimal solutions comprising different trade-offs between the objective functions (Zitzler *et al.*, 2000). The optimal solution set is considered

to be a Pareto optimal set (Coello Coello, 1999). A solution point is considered to be Pareto optimal if it is not dominated by any other solution in the solution space. In other words, a solution that cannot move to improve at least on objective function without a detriment in other objective functions, is Pareto optimal (Marler and Arora, 2010, 2004). The set of all non-dominated solutions is referred as to the Pareto optimal set, and for a given Pareto optimal set, the corresponding objective function values in the objective space are called the Pareto front (Patil and Dangewar, 2014). The concept of Pareto front can be illustrated in Figure II.5, where the solution of two objective functions $f_1(x)$ and $f_2(x)$ is compared and the solid black line represents the Pareto front.

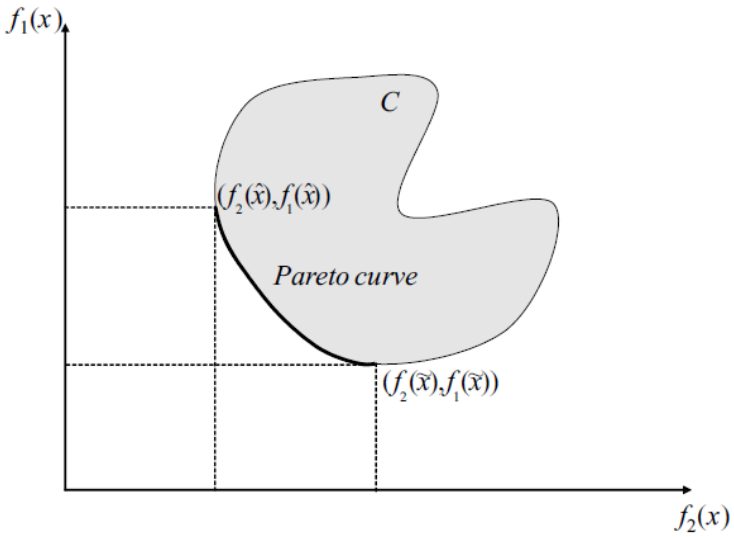


Figure II.5. Example of Pareto front.

Pareto optimal solutions are more realistic in practice since the final solution of the decision is a set of trade-offs. Notwithstanding, the determination of the entire Pareto optimal set is practically impossible due to its size (Lalwani *et al*., 2013). Therefore, a practical approach is to investigate a set of solutions that represent the Pareto optimal set as well as possible. In order to accomplish that the methods promoting diversity of solutions are highly desirable.
The most common algorithms to solve these problems are based on population such as evolutionary algorithms and particle swarm optimization (Angeline, 1998; Coello *et al*., 2007).

In the next subsections, several methods for multi-objective optimization are reviewed beginning with a weighted method to highlight the advantages of the utilization of Pareto-based algorithms.

III.3.1. Scalarization techniques

A multi objective problem is solved by combining multiple objectives into one single objective scalar function. This approach is known as the weighted sum or scalarization method (Marler and Arora, 2004). In more detail, this minimizes a positively weighted convex sum of the objectives as in equation (II.51)

$$\min \sum_{i=1}^{n} \gamma_i . f_i(\xi, u, \theta, t)$$
$$\sum_{i=1}^{n} \gamma_i = 1 \qquad \text{(II.51)}$$

This represents a new optimization problem with a unique objective function. It can be proved that minimizing this single objective function is an efficient solution for the original multi objective problem. However, this technique does not allow the definition of weight vectors iteratively to reach a certain portion of the Pareto curve. Performing several optimizations is necessary since the weight vector conduces to only one point of the Pareto curve. Moreover, the weighted sum method in two dimensions show that the solution on the space is a line, lacking of estimation of the convex zones (Marler and Arora, 2010).

III.3.2. Evolutionary Algorithms for Multi-objective optimization

Evolutionary algorithms are subdivided in three types (Reyes-Sierra and Coello Coello, 2006): aggregating functions; population based approaches; and Pareto based approaches.

Aggregating functions combines all the objectives in a single objective by any arithmetical operation. Population based approaches, use evolutionary algorithms to diversify the search. At each generation sub populations are generated by proportional selection. These approaches are simple to employ but their main limitation is the selection scheme, which is not based on Pareto optimality.

Pareto based approaches are the most popular approaches divided in two generations (Lalwani *et al*., 2013). In this type, the most applied Evolutionary algorithms are the Pareto Archived

Evolutionary Strategy (PAES) (Knowles and Corne, 2000) and Non-dominated Sorting Genetic Algorithm (NSGA) (Srinivas and Deb, 1994).

Although PAES holds important features, such as the external archive to store the non-dominated solution, it is a local search method whose performance depends on the population size. In the other hand, NSGA, even if improved by (Deb *et al*., 2002), uses a high computational complexity of non-dominated sorting which is inconvenient for large population sizes (Konak *et al*., 2006).

III.3.3. Multi-Objective Particle Swarm Optimization

Coello-Coello and Salazar-Lechuga (2002) (Coello-Coello et al., 2002) proposed a Multi-Objective Particle Swarm Optimization (MOPSO) to maximize the number of elements in the Pareto optimal set, minimize the distance with the Pareto front, and promote a high diversity of the solutions. Notwithstanding, some of the disadvantages of MOPSO are the static analysis of trade-offs and the handling of constraints.

Differently from evolutionary algorithms, in MOPSO each particle share information with others to select a leader (Coello *et al*., 2004) that will guide the swarm with a determined directional mutation (Lalwani *et al*., 2013).

In addition, as PAES, this algorithm accounts with a repository to save the non-dominated solutions (Reyes-Sierra and Coello Coello, 2006). Nonetheless, the repository has a limited size, and if it is full, new solutions are inserted based on the retention criterion, that is giving priority to solutions located in less crowded areas of the objective space.

MOPSO has been applied to different areas of knowledge (Lalwani *et al*., 2013), but there is only one application on fermentations to optimize the conditions of the production of α-amylase (Rajulapati and Narasu, 2011). The results showed a high conversion of starch to glucose and a further high ethanol yield in the fermentation.

IV. STATE ESTIMATION

Actually, the existing sensor technology for monitoring bioprocess does not allow the accurate and on-line readable measurement of significant indicators of fermentations (Dochain, 2008). These measurements are instead performed off-line providing delayed and infrequent information (Luttmann *et al*., 2012).

Software sensors have been proposed as an alternative to reconstruct the measurements of the process from the available measurements. Soft-sensors are process analytical devices that grant access to important non-measured process variables by mathematical transformations of available data. Two classes of soft sensors can be distinguished at a very general level: Model-based, and Data-based soft sensors (Kadlec *et al*., 2009).

Model based approaches are quite interesting to treat the transient behaviors of bioreactors to ensure the performance and efficiency of bioprocesses. The most common are the state observers, such as the Luengberger observer type (Zeitz, 1987), the Kalman Filter type (Bornard *et al*., 1989), and the asymptotic observers (Bastin and Dochain, 1990). This later was developed specifically for biological process. Usually, there is not enough phenomenological knowledge for an accurate description of the process modeling, and thus these approaches can cause severe errors of the on-line estimations. In addition to model uncertainties, observers take long times for convergence, and in some cases the matrix $K.r(\xi,\theta,t)$, relating the uptake vector and the matrix of yields, is not invertible which can cause observability problems (Dochain, 2003).

Data-based soft-sensors (Kadlec *et al*., 2009) are based on empirical knowledge that is processed to be expressed as rules, data or learning models. Their applicability depends only on the available process data (Zhao *et al*., 2008). Although simple on application, the interpretation of the relationships between the inputs and the estimated output is not always explicit and thus difficult to understand.

Luttmann *et al*., (2012) presented a review on the application of soft-sensors for bioprocesses, which are mainly based on on-line measurements from off gas analyzers. Quite common is the simple calculation of the derived variables such as oxygen uptake, carbon dioxide evolution rate, the respiratory quotient, among others.

In a general overview, soft-sensors provide online variables generating new insights about the processes (model-based) or granting simple classifications (data-based) of the measured variables. Table II.9 compiles interesting applications of software sensors for the online estimation of fermentation variables. The following subsections describe some common soft-sensors of the two classes before mentioned.

Table II.9. Application of software sensors to fermentations

Algorithm	Application	Input Variables	Output Variable	Reference
GSSA– SVM (AIC)	Erythromycin Fermentation	pO2, pH, Pressure, Rotation speed, Bean oil volume, Propyl alcohol, Water volume , Air flow rate, Biomass concentration at time (*t*-1)	Biomass concentration	(Liu *et al*., 2010)
SVM	Ivermectin fermentation	CO2 rate	Biomass concentration	(Lei et al., 2005)
		Sugar concentration	Ivermectin titer	
SVM (simulation)	Invertase process	Glucose, Biomass Ethanol, Volume	Invertase concentration	(Desai *et al*., 2006)
	Streptokinase process	Biomass, Glucose, Volume	Streptokinase concentration	
Least squared SVM	Penicillin Fermentation	Dissolve oxygen, Biomass concentration (*t*-1), penicillin concentration, substrate, Fermentation time	Growth rate + Biomass (t) (with model)	(Wang *et al*., 2010)
Improved PSO IPSO-SVM	Glutamate fermentation	Dissolved oxygen, Fermentation time, Biomass, Sugar,	Glutamate concentration (*t*)	(Zhang and Pan, 2015)

		Glutamate concentration (*t*-1)		
Algorithm	**Application**	**Input Variables**	**Output Variable**	**Reference**
SVM LSSVM	Glutathione production	Temperature, pH, Dissolved oxygen, CO2, Partial pressure of exhausted gas, Ethanol conc.	Biomass concentration	(Jianlin *et al*., 2006)
Multilayer NN	Iturin fermentation	Concentrations of: Asparagine, Glutamate, Proline	Iturin titer	(Chen *et al*., 2015)
Gen. Reg. NN				
SVM				
Genetic programming	Extracellular production of lipase	Fermentation time, Concentrations Soy oil, Ammonium nitrate, Corn Steep liquor,	Lipase activity (U/ml)	(Sharma and Tambe, 2014)
	Production of a bacterial copolymer	Acetate concentration, Propionate, Incubation period, pH	Accumulated Polyhydroxyalkanoates (% dcw)	
NARX,	Production of proteins by *Pichia pastoris*	pH, temperature, dissolved oxygen, stirrer speed,	heterologous protein concentration	(Geethalakshmi and Pappa, 2010)
Feed forward NN				
Continuous discrete Kalman Filter	Enzymatic Biodiesel Production	TAG (% w/w) DAG MAG FAME FFA	Glycerol Water Methanol Penetrated Enzyme Enzyme Complex Polar Volume Volume	(Price *et al*., 2015)

Extended Kalman Filter	Acetone Butanol Ethanol fermentation	CO_2 concentration	Concentrations of: Biomass, Butanol, Butyric acid, Acetone, Ethanol, Acetic acid, Sugar	(Jahanmiri and Rasooli, 2005)

Algorithm	**Application**	**Input Variables**	**Output Variable**	**Reference**
Non-linear observer	Ethanol production by *Saccharomyces cerevisiae* from glucose (Simulation)	Glucose concentration	Concentrations of: Biomass, Ethanol	(Rodriguez-Nunez *et al*., 2015)
Sliding observer				
Extended Kalman	Fed-batch fermentations of *Escherichia coli*	Dissolved CO_2, Dissolved Oxygen, CTR, OTR, Glucose feed rate, Weight,	Concentrations of: Biomass, Glucose, Acetate	(Veloso *et al*., 2009)
Asymptotic Observer				
Asymptotic Observer	Production of biodiesel from grease trap water	Temperature	Free fatty acids, Organic Esters, Water content	(Aguilar-Garnica *et al*., 2009)
Fuzzy Reset Observer				
Continuous discrete time observer	Ethanol production by *Saccharomyces cerevisiae* from sweet sorghum (Simulation)	Glucose concentration, Ethanol concentration	Growth rate, Production rate	(Bouraoui *et al*., 2015)
Robust Luengerger Observer	Microalgae cultures	Biomass concentration	Substrate concentration, Nitrogen quota	(Benavides *et al*., 2015)

IV.1. Model-based soft-sensors

One type of model-based soft-sensors are the so-called observers (Kadlec *et al*., 2009). Two classes of observers are found in the literature: the classical (Wells, 1971) and the asymptotic observers (Dochain, 2003). The former are applicable when the model is well identified and considered to lack of uncertainties. Nonetheless, it is difficult to develop error bounds for bioprocesses and there is often a large uncertainty on the parameters. The asymptotic observers

are based on the idea that the uncertainties in the bioprocess are within the process kinetics. The drawback of this observer is the rate of convergence of the estimation which fully depends on the operating conditions.

Observers, however, present interesting features to reconstruct the state variables. Therefore, in the following subsections, the classic and asymptotic observers are described in detailed.

IV.1.1. Classical observers

The theory of the classical observers was developed using perfect knowledge of the system parameters, in particular of the process kinetics (Dochain, 2003). A general schema of the application of observers as displayed in Figure II.6.

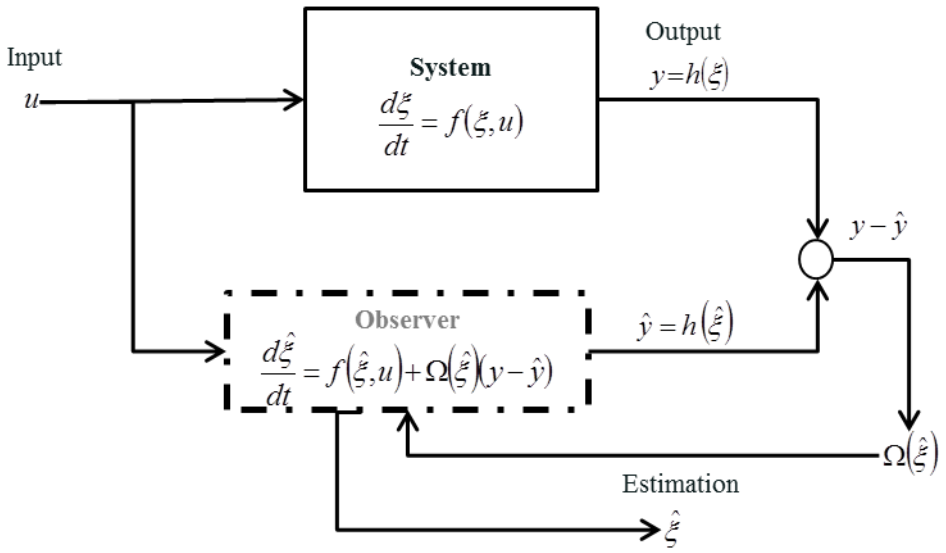


Figure II.6. General Schema of an observer.

In this figure, the system is described by a nonlinear state space model:

$$\frac{d\xi}{dt} = f(\xi, u) \tag{II.52}$$

where ξ represents the states of the model, and u the inputs of the system. The measured output of the system is considered as,

$$y = h(\xi) \tag{II.53}$$

which is related to the process dynamical model. Then the structure of the observer is

$$\frac{d\hat{\xi}}{dt} = f\left(\hat{\xi},u\right) + \Omega\left(\hat{\xi}\right)\left(y - \hat{y}\right) \tag{II.54}$$

where Ω is the gain matrix dependent on the online estimation of states $\hat{\xi}$. The estimated outputs are represented by $\hat{y}$. The design of the observer is reduced to the computation of the gain Ω. The design of the gain matrix is based on a linearized version of the process dynamics observation error. Therefore, let us define the observation error as,

$$e = \xi - \hat{\xi} \tag{II.55}$$

Whose dynamics are given by

$$\frac{de}{dt} = f\left(\hat{\xi} + e,u\right) - f\left(\hat{\xi},u\right) + \Omega\left(\hat{\xi}\right)\left(h\left(\hat{\xi} + e\right) - h\left(\hat{\xi}\right)\right) \tag{II.56}$$

This dynamic equation can only be calculated for linear systems. Although bioprocess models are non-linear systems, they are assumed to be linear around some equilibrium point. In this context, the system can be linearized by Taylor series around an equilibrium point (e = 0) and thus equation (II.56) is rewritten as,

$$\frac{de}{dt} = \left[A\left(\hat{\xi}\right) - \Omega\left(\hat{\xi}\right)L\left(\hat{\xi}\right)\right]e \tag{II.57}$$

$$A\left(\hat{\xi}\right) = \left[\frac{\partial f}{\partial \xi}\right]_{\xi=\hat{\xi}}, \qquad L\left(\hat{\xi}\right) = \left[\frac{\partial h}{\partial \xi}\right]_{\xi=\hat{\xi}} \tag{II.58}$$

A necessary condition for the applicability of state observers is the exponential observability condition stating that the error is exponentially observable if $rank\left(L, LA\left(\hat{\xi}\right), \ldots, L.A\left(\hat{\xi}\right)^{N-1}\right)^{T} = N$ (Bastin and Dochain, 1990).

The definition of state observers is thus dependent on the manner of computing the gain Ω. For instance, the Luenberger observer calculates the gain to guaranty asymptotic stability and a good convergence speed (Soroush, 1998). Otherwise, the Kalman Filter selects Ω by minimizing the mean square observation error (Bastin and Dochain, 1990). However, these two classical observers failed when parametric uncertainties are presented (Dochain, 2003), which is inherent on bioprocesses.

IV.1.2. Asymptotic observers

Asymptotic observers were developed on the idea that model uncertainties are only due to the kinetics. Let us consider a continuous culture, where a stoichiometric coefficient matrix *K* is introduced into the system considered in equation (II.52), with its respective reaction rate vector *r* as follows,

$$\frac{d\xi}{dt} = K.r(\xi,t) - D\xi + F \tag{II.59}$$

where *D* is the dilution rate, and *F* the input flow rate. If we consider that certain state variables are measured, the system can be partitioned in two as,

$$\begin{aligned} \frac{d\xi_a}{dt} &= K_a.r(\xi,t) - D\xi_a + F_a \\ \frac{d\xi_b}{dt} &= K_b.r(\xi,t) - D\xi_b + F_b \end{aligned} \tag{II.60}$$

where ξ_a holds for the measured stated and ξ_b for the non-measured states. Moreover, if K_a is full rank, then it is possible to define a transformation.

$$Z = A_0.\xi_a + \xi_b \tag{II.61}$$

Where A_0 is calculated to fulfill the following condition

$$A_{0.}K_a + K_b = 0 \tag{II.62}$$

Therefore, the system in the transformation variables is redefined as,

$$\frac{dZ}{dt} = -DZ + A_0.F_a + F_b \tag{II.63}$$

which is independent from reaction kinetics. Finally, the reconstruction of the states is made by:

$$\hat{\xi}_b = Z - A_0\xi_a \tag{II.64}$$

This observer is based on several assumptions where the dilution rate *D* is known and bounded between D_{min} and D_{max}. The matrix of stoichiometric coefficients *K* is constant and known, and the input flow rate *F* is also known.

One characteristic of this observer is that its stability and convergence depend on the dilution rate *D* being positive, which means that *D* must not remain equal to zero for long time (Dochain, 2003). Therefore, the asymptotic observers guaranty its asymptotic stability even if the dynamical model is unstable. However, their convergence depends on the initial value of *Z* which affects the performance of the observer.

Different types of observers have been employed for biological process (Dochain, 2008) such as interval observers for anaerobic digestion with a tunable convergence rate (Alcaraz-González *et al*., 2005, 2002); parameter estimation in fermentations (Jahanmiri and Rasooli, 2005); and the extensions of classical observers such as the Kalman Filter for biodiesel production (Price *et al*., 2015), and the Luenberger observer for microalgae cultures (Benavides *et al*., 2015). Although their utilization, in the industry they have been replaced by data-based soft sensors (Yin *et al*., 2015).

IV.2. Data-based soft-sensors

The data-based methods are based on the data measured within the processing plants, and thus describe the real process conditions. These algorithms are built from "experience" or historic process data of the phenomenon under study. These data can be pre-processed and expressed in a set of rules, or serve as training data for statistical and machine learning models. A complete review of data-driven methods is reported by (Kadlec *et al*., 2009), where the authors present five of the most popular techniques: Principle Component Analysis (PCA), Partial Least Squares (PLS), Neuro-fuzzy systems (NFS), Artificial Neural Networks (ANN), and Support Vector Machines (SVM). PCA and varieties of ANN were the most popular in use to the knowledge of those authors. However, the employment of SVM has gain in applications because of its simple representation and better generalization ability than ANN avoiding local minima and over-fitting (Dreiseitl and Ohno-Machado, 2002).

In the following subsections ANN and SVM are briefly described. Even if widely used, PCA is not revised in this thesis due to its limitations on handling nonlinear correlations of data, lack of information of the selection of the optimal number of principal components, and the missing relation between input and output data.

IV.2.1. Artificial Neural Networks

Artificial Neural Networks (ANN) is a multi-layer perceptron that learn the nonlinear relationships existing between the inputs $x \in R^m$ and outputs o_N to make quick predictions of

the outputs. ANN is based on the probability that the outputs are into the *m*-dimensional input data space *x*.

ANN provide a function to explicitly express that probability depending on several parameters (β, β_0) and the outputs o_H of the perceptron (hidden neurons).

$$P\left(x\middle|\beta,\beta_0,o_H\right)=\frac{1}{1+e^{-(\beta.o_H+\beta_0)}} \quad \text{(II.65)}$$

The parameters β and β_0 are determined based on the data set by maximum-likelihood estimation. However, the learnt knowledge is distributed in the weights between the particular neurons and is not available in terms of understandable representation. That is the resultant model and its parameters are difficult to interpret from the view point of gaining an insight into the data used in the model (Desai *et al*., 2006). ANN neither guarantees a high convergence speed, nor the avoidance of local minima and the over-fitting phenomenon. Furthermore, there are no general methods to choose the number of hidden units for general networks 'topology of the networks'. The topology of the ANN is critical for their performance because their generalization power is dependent on the complexity of the networks. These disadvantages on ANN have encouraged the search of other techniques such as SVM, and the coupling of the former with construction theory (models).

Nevertheless, recent examples of ANN have reported for the monitoring of lipases production by the fungus *Geotrichum candidum*, and PHA accumulation by *Wautersia eutropha* (Sharma and Tambe, 2014); protein production by *Pichia pastoris* (Geethalakshmi and Pappa, 2010); and amino acids on the iturin fermentation (Chen *et al*., 2015). ANN have been also coupled with models (Liu *et al*., 2014) for the estimation of mycelial and sugar concentrations in the erythromycin fermentation showing good agreement in the results.

IV.2.2. Support Vector Machines

Support vector machines (SVMs) is a supervised non-parametric statistical learning technique, therefore there is no assumption made on the underlying data distribution. The principle of SVM consists to bring the problem of discrimination to the search of an optimal hyperplane that separates the data in classes consistent with the training data sets (Hsu *et al.*, 2008; Vapnik *et al.*, 1996). The hyperplane is defined as solution of a constraint optimization problem:

$$\min_{w,\zeta,\zeta^*} J = \frac{1}{2} w^T.w + C_R.\sum_{i=i}^{N}\left(\zeta_\iota + \zeta_i^*\right) \tag{II.66}$$

$$\text{Subject to : } \begin{cases} y_i - w^T.\varphi(x_i) - b \le \varepsilon + \zeta_i \\ w^T.\varphi(x_i) + b - y_i \le \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \ge 0 \end{cases} \tag{II.67}$$

where $\varphi(x): R^N \to R^M$ is the high dimensional feature space which is non-linearly mapped from the input space *x*. The objective function is expressed in function of scalar products between vectors *w*, in which the number of active constraints or support vectors control the complexity of the model. In equation (II.66) the first term is the regularization term, and the second term is the empirical error (risk) measured by the insensitive ε-loss function. The variables ξ_i and ξ_i^* are the slack variables that measure the deviation of the support vectors from the boundaries of the ε-zone. C_R is the regularization constant that determines the trade-off between the empirical risk and the regularized term.

The representation of support vectors can be seen in Figure II.7, where the term ε represents the tube size around the solutions. The subsets of points that lie on the margin of the tube (called support vectors) are the only ones that define the hyperplane. Therefore, this approach is penalized by the number of potential support vectors and no by the number of variables.

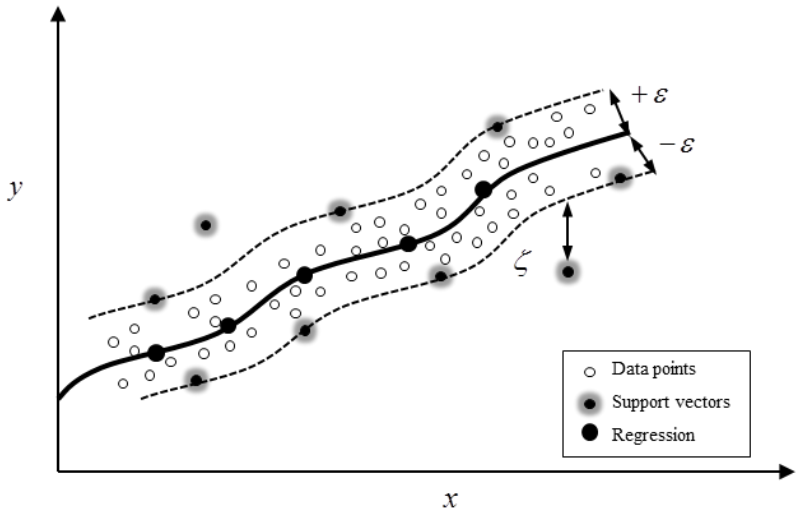


Figure II.7. The soft margin loss setting for SVM (Adapted from Desai *et al*. (2006)**).**

SVM build optimal separating boundaries between data sets by solving a constrained quadratic optimization problem (Huang and Dun, 2008). In order to simplify the minimization problem, Lagrange multipliers are introduced as follows

$$L = J - \sum_{i=1}^{N} \delta_i \left\{ \zeta_i + \varepsilon - y_i + w^T . \varphi(x_i) + b \right\} - \sum_{i=1}^{N} \delta_i^* \left\{ \zeta_i^* + \varepsilon + y_i - w^T . \varphi(x_i) - b \right\} - \sum_{i=1}^{N} \left(\eta_i . \zeta_i - \eta_i^* . \zeta_i^* \right) \quad \text{(II.68)}$$

where the parameters δ_i, δ_i^*, η_i and η_i^* are the Lagrange multipliers. According to the Karush-Kuhn-Tucker (KKT) of quadratic programming, the dual equation that can be obtained (Liu *et al*., 2010) is:

$$\min_{\alpha,\alpha^*} J_\delta = \frac{1}{2} \sum_{i,j=i}^{N} \left(\delta_\iota - \delta_i^* \right) . \left(\delta_j - \delta_j^* \right) . K(x_i, x_j) + \varepsilon . \sum_{i=i}^{N} \left(\delta_\iota + \delta_i^* \right) - \sum_{i=i}^{N} \left(\delta_\iota - \delta_i^* \right) . y_i \quad \text{(II.69)}$$

$$\text{Subject to : } \begin{cases} \sum_{i=i}^{N} \left(\delta_\iota - \delta_i^* \right) = 0 \\ 0 \le \delta_\iota, \delta_i^* \le C_R ; i = 1,2,\ldots,N \end{cases} \quad \text{(II.70)}$$

Therefore, the final regression function given in equation is rewritten as:

$$y = \sum_{i=i}^{N} \left(\delta_\iota - \delta_i^* \right) . K(x_i, x_j) + b \quad \text{(II.71)}$$

where $K(x_i, x_j) = \varphi(x_i) . \varphi(x_j)$ is the kernel function which allows to inducing a higher flexibility in handling nonlinearities (Dreiseitl and Ohno-Machado, 2002). In addition, the kernel transforms the mapping of the system from a sum of R^M dimensions into a sum of N points. Only a small proportion of the Lagrange multipliers (δ_i, δ_i^*) will be non-zero, and only the store of the corresponding training points from the input space x is needed. These points are the support vectors (Üstün *et al*., 2005).

One of the most important drawbacks of this approach is the determination of the regularization parameter C_R, the tube size ε, and the kernel related parameters which change the overall performance of the soft-sensor. Nonetheless, some authors have proposed the link with an optimization algorithm such as Genetic algorithms (Liu *et al*., 2010), or improved PSO (Zhang and Pan, 2015) to overcome the drawback before mentioned.

SVM has been scarcely applied to bioprocesses (Table II.9). In most of the reported works the used kernel was the Radial Basis Function (RBF or Gaussian) kernel because it can classify

multi-dimensional data, and it has fewer parameters to be set in comparison with other kernels. The exception is the work of (Zhang and Pan, 2015), who used a combination of linear and RBG kernel.

Further investigations are required to prove the experimental efficiency of SVM because all the reported works were merely simulation results. In addition, most of them take into account the concentration of certain state variables that are unfrequently measured. The exceptions are the works of (Liu *et al*., 2010) and (Lei et al., 2005) for the estimation of biomass concentration.

V. CONCLUSIONS

The existing models of lipid accumulation in yeasts and fungi were reviewed, focusing on the kinetics of the primary products (*i.e.* lipids, citric acid). All of these models were classified as unstructured models where only one model described the production of lipids and citric acid. However, these unstructured models cannot accurately predict the intracellular mechanisms that cause metabolic shifts such as overflow.

The dynamic metabolic modeling frameworks were also reviewed to include the interaction of metabolites on the cell metabolism, and specially to analyze the interactions in yeasts. All the frameworks rely on the balance growth hypothesis or quasi steady state assumption (QSSA). However, some metabolites could be considered as accumulated by modifying some equations (see section II.3.3). This fact is of great interest for modeling lipids. Therefore, this modeling approach could be used for coupling metabolism to the dynamic behavior of the primary metabolites towards the optimization of lipid production.

Regarding process optimization, most of the applications of the control strategies aimed at maximizing productivities of different products of interest by manipulating the feeding rate. Fermentations require the optimization of several objectives (*e.g.* productivity, yield, time). Therefore, multi-objective optimization is used in this work (see CHAPTER III.V). However, this type of optimization could cause some issues specially when weighting the objectives. Pareto based algorithms were presented as good alternatives, where MOPSO was highlighted due to its fast convergence and great distribution of the swarm. This state-of-the-art encourages

the proposition of multi-objective control for optimizing lipid productivity and lipid content fraction.

Finally, the review on state estimation highlighted the importance of monitoring the cultures towards optimization. The model-based soft sensors also known as observers were described with their several applications to bioprocesses. Although their utilization, they depend on a well identified and observable model, and their convergence speed is defined on the operating conditions. On the other hand, data-based soft sensors had gained popularity on industrial application, where SVM was depicted as an interesting algorithm for its simple representation and better generalization. It is therefore possible to believe that this method could be straightforward applicable for online estimation of the most important variables involved in lipid fermentations.

VI. REFERENCES

Aggelis, G., Komaitis, M., Papanikolaou, S., Papadopoulos, G., 1995a. A mathematical model for the study of lipid accumulation in oleaginous microorganisms. I: Lipid accumulation during growth of Mucor circinelloides CBS 172-27 on a vegetable oil. Grasas y aceites 46, 169–173.

Aggelis, G., Komaitis, M., Papanikolaou, S., Papadopoulos, G., 1995b. A mathematical model for the study of lipid accumulation in oleaginous microorganisms. II. Study of cellular lipids of Mucor circinelloides during growth on a vegetable oil. Grasas y Aceites 46, 245–250. doi:10.3989/gya.1995.v46.i4-5.932

Aggelis, G., Sourdis, J., 1997. Prediction of lipid accumulation-degradation in oleaginous micro-organisms growing on vegetable oils. Antonie Van Leeuwenhoek 72, 159–165.

Aguilar-Garnica, E., Dochain, D., Alcaraz-González, V., González-Álvarez, V., 2009. A multivariable control scheme in a two-stage anaerobic digestion system described by partial differential equations. J. Process Control 19, 1324–1332. doi:10.1016/j.jprocont.2009.04.003

Alcaraz-González, V., Harmand, J., Rapaport, A., Steyer, J.P., González-Alvarez, V., Pelayo-Ortiz, C., 2002. Software sensors for highly uncertain WWTPs: a new approach based on interval observers. Water Res. 36, 2515–24.

Alcaraz-González, V., Salazar-Pena, R., Gonzalez Alvarez, V., Gouzé, J.-L., Steyer, J.-P.,

2005. A tunable multivariable nonlinear robust observer for biological systems. C.R.Biol 328, 317–325.

Alford, J.S., 2006. Bioprocess control: Advances and challenges. Comput. Chem. Eng. 30, 1464–1475. doi:10.1016/j.compchemeng.2006.05.039

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., Jirstrand, M., 2014. Kinetic models in industrial biotechnology – Improving cell factory performance. Metab. Eng. 24, 38–60. doi:10.1016/j.ymben.2014.03.007

Amribt, Z., Niu, H., Bogaerts, P., 2013. Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. Biochem. Eng. J. 70, 196–209. doi:10.1016/j.bej.2012.11.005

Anastassiadis, S., Aivasidis, A., Wandrey, C., S., A., A., A., C., W., Anastassiadis, S., Aivasidis, A., Wandrey, C., 2002. Citric acid production by Candida strains under intracellular nitrogen limitation. Appl. Microbiol. Biotechnol. 60, 81–7. doi:10.1007/s00253-002-1098-1

Angeline, P.J., 1998. Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences. Lect. Notes Comput. Sci. Evol. Program. VII 1447, 601–610. doi:10.1007/BFb0040811

Arzumanov, T.T.E., Sidorov, I.I.A., Shishkanova, N.V.N., Finogenova, T.T. V., 2000. Mathematical modeling of citric acid production by repeated batch culture. Enzyme Microb. Technol. 26, 826–833. doi:http://dx.doi.org/10.1016/S0141-0229(00)00178-2

Ashoori, A., Moshiri, B., Khaki-Sedigh, A., Bakhtiari, M.R., 2009. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. J. Process Control 19, 1162–1173. doi:10.1016/j.jprocont.2009.03.006

Bailey, J.E.J.E., 1998. Mathematical Modeling and Analysis in Biochemical Engineering : Past Accomplishments and Future Opportunities. Biotechnol. Prog. 1973, 8–20. doi:10.1021/bp9701269

Balagurunathan, B., Jonnalagadda, S., Tan, L., Srinivasan, R., 2012. Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis. Microb. Cell Fact. 11, 27. doi:10.1186/1475-2859-11-27

Baroukh, C., Muñoz-Tamayo, R., Steyer, J.-P., Bernard, O., 2014. DRUM: a new framework for metabolic modeling under non-balanced growth. Application to the carbon metabolism of unicellular microalgae. PLoS One 9, e104499. doi:10.1371/journal.pone.0104499

Bastin, G., Van Impe, J.F., Bastin, G., Van Impe, J.F., Bastin, G., Van Impe, J.F., 1995. Nonlinear and Adaptive Control in Biotechnology: A Tutorial. Eur. J. Control 1, 37–53.

doi:10.1016/S0947-3580(95)70006-1

Bastin, G., Dochain, D., 1990. On-line Estimation and Adaptive Control of Bioreactors. Elsevier.

Beard, D.A., Liang, S.C., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. Biophys. J. 83, 79–86. doi:10.1016/S0006-3495(02)75150-3

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.Ø., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–38. doi:10.1038/nprot.2007.99

Benavides, M., Coutinho, D., Hantson, A.-L., Van Impe, J., Vande Wouwer, A., 2015. Robust Luenberger observers for microalgal cultures. J. Process Control 36, 55–63. doi:10.1016/j.jprocont.2015.09.005

Benavides, P.T., Diwekar, U., 2012. Optimal control of biodiesel production in a batch reactor. Fuel 94, 218–226. doi:10.1016/j.fuel.2011.08.033

Bideaux, C., Montheard, J., Cameleyre, X., Molina-Jouve, C., Alfenore, S., 2016. Metabolic flux analysis model for optimizing xylose conversion into ethanol by the natural C5-fermenting yeast Candida shehatae. Appl. Microbiol. Biotechnol. 100, 1489–99. doi:10.1007/s00253-015-7085-0

Bornard, G., Couenne, N., Celle, F., 1989. Regularly persistent observers for bilinear systems, in: New Trends in Nonlinear Control Theory. Springer-Verlag, Berlin/Heidelberg, pp. 130–140. doi:10.1007/BFb0043023

Bouraoui, I., Farza, M., Ménard, T., Ben Abdennour, R., M'Saad, M., Mosrati, H., 2015. Observer design for a class of uncertain nonlinear systems with sampled outputs—Application to the estimation of kinetic rates in bioreactors, Automatica. doi:10.1016/j.automatica.2015.02.036

Çelik, E., Çalik, P., Oliver, S.G., 2010. Metabolic flux analysis for recombinant protein production by Pichia pastoris using dual carbon sources: Effects of methanol feeding rate. Biotechnol. Bioeng. 105, 317–329. doi:10.1002/bit.22543

Chen, F., Li, H., Xu, Z., Hou, S., Yang, D., 2015. User-friendly optimization approach of fed-batch fermentation conditions for the production of iturin A using artificial neural networks and support vector machine. Electron. J. Biotechnol. 18, 273–280. doi:10.1016/j.ejbt.2015.05.001

Christofides, P.D., Scattolini, R., Muñoz de la Peña, D., Liu, J., 2013. Distributed model predictive control: A tutorial review and future research directions. Comput. Chem. Eng. 51, 21–41. doi:10.1016/j.compchemeng.2012.05.011

Coello-Coello, C.A.C.A., Salazar-Lechuga, M.S., Coello Coello, C. a., Lechuga, M.S., 2002. MOPSO: A proposal for multiple objective particle swarm optimization. Proc. 2002 Congr. Evol. Comput. CEC'02 (Cat. No.02TH8600) 2, 1051–1056. doi:10.1109/CEC.2002.1004388

Coello, C.A.C. a C., Pulido, G.T.T., Lechuga, M.S.S., 2004. Handling multiple objectives with particle swarm optimization. Evol. Comput. IEEE Trans. 8, 256–279. doi:10.1109/TEVC.2004.826067

Coello, C.C., Lamont, G.B., Veldhuizen, D.A. van, 2007. Evolutionary Algorithms for Solving Multi-Objective Problems.

Coello Coello, C. a., 1999. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. Knowl. Inf. Syst. 1, 269–308. doi:10.1007/BF03325101

Crown, S.B., Ahn, W.S., Antoniewicz, M.R., 2012. Rational design of 13C-labeling experiments for metabolic flux analysis in mammalian cells. BMC Syst. Biol. 6, 43. doi:10.1186/1752-0509-6-43

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197. doi:10.1109/4235.996017

Desai, K., Badhe, Y., Tambe, S.S., Kulkarni, B.D., 2006. Soft-sensor development for fed-batch bioreactors using support vector regression. Biochem. Eng. J. 27, 225–239. doi:10.1016/j.bej.2005.08.002

Diego, S., Jolla, L., Gill, P.E., Jay, L.O., Leonard, M.W., Petzold, L.R., 2000. An SQP method for the optimal control of large-scale dynamical systems. J. Comput. Appl. Math. 120, 197–213.

Dobson, P.D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W.B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N.J., Kell, D.B., King, R.D., Oliver, S.G., Stevens, R.D., Mendes, P., 2010. Further developments towards a genome-scale metabolic model of yeast. BMC Syst. Biol. 4, 145. doi:10.1186/1752-0509-4-145

Dochain, D., 2008. Bioprocess Control, First edit. ed. John Wiley & Sons, Inc., London.

Dochain, D., 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. J. Process Control 13, 801–818. doi:10.1016/S0959-1524(03)00026-X

Donoho, D., Stodden, V., 2004. When does non-negative matrix factorization give a correct decomposition into parts? Proc. Adv. Neural Inf. Process. Syst. 16 1141–1148.

Donoso-Bravo, A., Mailier, J., Martin, C., Rodríguez, J., Aceves-Lara, C.A., Vande Wouwer, A., 2011. Model selection, identification and validation in anaerobic digestion: a review.

Water Res. 45, 5347–64. doi:10.1016/j.watres.2011.08.059

Dreiseitl, S., Ohno-Machado, L., 2002. Logistic regression and artificial neural network classification models: a methodology review. J. Biomed. Inform. 35, 352–359. doi:10.1016/S1532-0464(03)00034-0

Duarte, N.C., Herrgård, M.J., Palsson, B.Ø., 2004. Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. Genome Res. 14, 1298–1309. doi:10.1101/gr.2250904

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. MHS'95. Proc. Sixth Int. Symp. Micro Mach. Hum. Sci. 39–43. doi:10.1109/MHS.1995.494215

Eberhart, R.C., Shi, Y., 2000. Comparing inertia weights and constriction factors in particle swarm optimization, in: Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512). IEEE, pp. 84–88. doi:10.1109/CEC.2000.870279

Eberhart, R.C., Yuhui Shi, 2001. Particle swarm optimization: developments, applications and resources, in: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546). IEEE, pp. 81–86. doi:10.1109/CEC.2001.934374

Economou, C.N., Vasiliadou, I.. A., Aggelis, G., Pavlou, S., Vayenas, D. V., 2011. Modeling of oleaginous fungal biofilm developed on semi-solid media. Bioresour. Technol. 102, 9697–704. doi:10.1016/j.biortech.2011.08.003

Economou, C.N.C.N., Aggelis, G., Pavlou, S., Vayenas, D.V. V., 2011. Modeling of single-cell oil production under nitrogen-limited and substrate inhibition conditions. Biotechnol. Bioeng. 108, 1049–55. doi:10.1002/bit.23026

Edwards, J.S., Covert, M., Palsson, B., 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. 4, 133–140. doi:10.1046/j.1462-2920.2002.00282.x

Fakas, S., Stamatina, M., Aggelis, G., 2009. Single cell oil and gamma-linolenic acid production by Thamnidium elegans on raw glycerol, in: NY: Nova Science Publishers (Ed.), Aggelis G, Editors. Microbial Conversions of Raw Glycerol. pp. 85–100.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 163–195. doi:10.1038/msb4100155

Feist, A.M., Palsson, B.O., 2010. The biomass objective function. Curr. Opin. Microbiol. 13, 344–349. doi:10.1016/j.mib.2010.03.003

Forster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-Scale Reconstruction of

the Saccharomyces cerevisiae Metabolic Network. Genome Res. 13, 244–253. doi:10.1101/gr.234503

Franz, A., Song, H.-S., Ramkrishna, D., Kienle, A., 2011. Experimental and theoretical analysis of poly(β-hydroxybutyrate) formation and consumption in Ralstonia eutropha. Biochem. Eng. J. 55, 49–58. doi:10.1016/j.bej.2011.03.006

Fredrickson, A.G., Megee, R.D., Tsuchiya, H.M., 1970. Mathematical Models in Fermentation Processes. Adv. Appl. Microbiol. 13, 419.

Galluzzo, M., Cosenza, B., 2010. Adaptive type-2 fuzzy logic control of a bioreactor. Chem. Eng. Sci. 65, 4208–4221. doi:10.1016/j.ces.2010.04.023

Geethalakshmi, S., Pappa, N., 2010. Artificial Neural Network Based Soft Sensor for Fermentation of Recombinant Pichia Pastoris. Adv. Comput. Eng. (ACE), 2010 Int. Conf. 0–4. doi:10.1109/ACE.2010.56

Geng, J., Bi, J.X., Zeng, A.P., Yuan, J.Q., 2013. Application of hybrid cybernetic model in simulating myeloma cell culture co-consuming glucose and glutamine with mixed consumption patterns. Process Biochem. 48, 955–964. doi:10.1016/j.procbio.2013.03.019

Geng, J., Song, H.-S., Yuan, J., Ramkrishna, D., 2012. On enhancing productivity of bioethanol with multiple species. Biotechnol. Bioeng. 109, 1508–17. doi:10.1002/bit.24419

Giuseppin, M.L., van Riel, N. a, 2000. Metabolic modeling of Saccharomyces cerevisiae using the optimal control of homeostasis: a cybernetic model definition. Metab. Eng. 2, 14–33. doi:10.1006/mben.1999.0134

György, A., Kocsis, L., 2011. Efficient Multi-Start Strategies for Local Search Algorithms. J. Artif. Intell. Res. 41, 407–444.

Hanly, T.J., Henson, M.A., 2013. Dynamic metabolic modeling of a microaerobic yeast co-culture: predicting and optimizing ethanol production from glucose/xylose mixtures. Biotechnol. Biofuels 6, 44. doi:10.1186/1754-6834-6-44

Hanly, T.J., Henson, M.A., 2011. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol. Bioeng. 108, 376–85. doi:10.1002/bit.22954

Herrgård, M.J., Lee, B.-S., Portnoy, V., Palsson, B.Ø., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Res. 16, 627–35. doi:10.1101/gr.4083206

Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Büthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E.,

Smallbone, K., Spasié, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H. V, Kürdar, B., Penttilä, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S.G., Mendes, P., Nielsen, J., Kell, D.B., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155–1160. doi:10.1038/nbt1492

Herwig, C., von Stockar, U., 2002. A small metabolic flux model to identify transient metabolic regulations in Saccharomyces cerevisiae. Bioprocess Biosyst. Eng. 24, 395–403. doi:10.1007/s00449-001-0277-2

Hisbullah, Hussain, M.A., Ramachandran, K.B., 2003. Design of a Fuzzy Logic Controller for Regulating Substrate Feed to Fed-Batch Fermentation. Food Bioprod. Process. 81, 138–146. doi:10.1205/096030803322088279

Hjersted, J.L., Henson, M.A., 2009. Steady-state and dynamic flux balance analysis of ethanol production by Saccharomyces cerevisiae. IET Syst. Biol. 3, 167–79. doi:10.1049/iet-syb.2008.0103

Hjersted, J.L., Henson, M.A., 2006. Optimization of fed-batch Saccharomyces cerevisiae fermentation using dynamic flux balance models. Biotechnol. Prog. 22, 1239–1248. doi:10.1021/bp060059v

Hocalar, A., Türker, M., 2010. Model based control of minimal overflow metabolite in technical scale fed-batch yeast fermentation. Biochem. Eng. J. 51, 64–71. doi:10.1016/j.bej.2010.04.014

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2008. A Practical Guide to Support Vector Classification. BJU Int. 101, 1396–400. doi:10.1177/02632760022050997

Huang, C.-L., Dun, J.-F., 2008. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. Appl. Soft Comput. 8, 1381–1391. doi:10.1016/j.asoc.2007.10.007

Jahanmiri, A., Rasooli, H., 2005. On-line states and parameter identification of acetone–butanol–ethanol fermentation process. Biochem. Eng. J. 24, 115–123. doi:10.1016/j.bej.2005.02.006

Jianlin, W., Tao, Y.U., Cuiyun, J.I.N., 2006. On-line estimation of biomass in fermentation process using support vector machine. Chinese J. Chem. Eng. 14, 383–388.

Jin, H., Chen, X., Yang, J., Wu, L., Wang, L., 2014. Hybrid intelligent control of substrate feeding for industrial fed-batch chlortetracycline fermentation process. ISA Trans. 53, 1822–1837. doi:10.1016/j.isatra.2014.08.015

Jol, S.J., Kümmel, A., Terzer, M., Stelling, J., Heinemann, M., 2012. System-level insights into

yeast metabolism by thermodynamic analysis of elementary flux modes. PLoS Comput. Biol. 8, e1002415. doi:10.1371/journal.pcbi.1002415

Jouhten, P., Wiebe, M., Penttilä, M., 2012. Dynamic flux balance analysis of the metabolism of Saccharomyces cerevisiae during the shift from fully respirative or respirofermentative metabolic states to anaerobiosis. FEBS J. 279, 3338–3354. doi:10.1111/j.1742-4658.2012.08649.x

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven Soft Sensors in the process industry. Comput. Chem. Eng. 33, 795–814. doi:10.1016/j.compchemeng.2008.12.012

Kamp, A. v., Schuster, S., von Kamp, A., Schuster, S., 2006. Metatool 5.0: Fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931. doi:10.1093/bioinformatics/btl267

Kasperski, A., Miśkiewicz, T., 2002. An adaptive fuzzy logic controller using the respiratory quotient as an indicator of overdosage in the baker's yeast process. Biotechnol. Lett. 24, 17–21. doi:10.1023/A:1013896930645

Kavšček, M., Bhutada, G., Madl, T., Natter, K., 2015. Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. BMC Syst. Biol. 9, 72. doi:10.1186/s12918-015-0217-4

Kim, J. Il, Varner, J.D., Ramkrishna, D., 2008. A Hybrid Model of Anaerobic E. coli GJT001: Combination of Elementary Flux Modes and Cybernetic Variables. Biotechnol. Prog. 24, 993–1006. doi:10.1021/bp.73

Kirgat, G.S., Surde, A.N., 2014. Review of Hooke and Jeeves Direct Search Solution Method Analysis Applicable To Mechanical Design Engineering. Int. J. Innov. Eng. Res. Technol. 1, 1–14.

Klamt, S., Saez-Rodriguez, J., Gilles, E.D., 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2. doi:10.1186/1752-0509-1-2

Knowles, J.D., Corne, D.W., 2000. Approximating the nondominated front using the Pareto Archived Evolution Strategy. Evol. Comput. 8, 149–172. doi:10.1162/106365600568167

Koffas, M., Stephanopoulos, G., 2005. Strain improvement by metabolic engineering: lysine production as a case study for systems biology. Curr. Opin. Biotechnol. 16, 361–366. doi:10.1016/j.copbio.2005.04.010

Konak, A., Coit, D.W., Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: A tutorial. Reliab. Eng. Syst. Saf. 91, 992–1007. doi:10.1016/j.ress.2005.11.018

Koutinas, M., Kiparissides, A., Pistikopoulos, E.N., Mantalaris, A., 2012. Bioprocess systems

engineering: Tansferring traditional process engineering principles to industrial biotechnology. Comput. Struct. Biotechnol. J. 3, 1–9. doi:10.5936/csbj.201210022

Kuepfer, L., Sauer, U., Blank, L.M., 2005. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. Genome Res. 15, 1421–30. doi:10.1101/gr.3992505

Lalwani, S., Singhal, S., Kumar, R., Gupta, N., 2013. A comprehensive survey: Applications of multi-objective particle swarm optimization (MOPSO) algorithm. Trans. Comb. 2, 39–101.

Lei, L., Sun, Z., Liang-Yu Lei, Zong-Hai Sun, Lei, L., Sun, Z., 2005. Soft Sensor Based on Generalized Support Vector Machines. Mach. Learn. 7, 18–21. doi:10.1109/ICMLC.2005.1527694

Lewis, R.M., Torczon, V., 1996. Pattern Search algorithms for bound constrained minimization. SIAM J. Optim.

Liu, G., Yu, S., Mei, C., Ding, Y., 2014. A novel soft sensor model based on artificial neural network in the fermentation process. African J. Biotechnol. doi:10.4314/ajb.v10i85.

Liu, G., Zhou, D., Xu, H., Mei, C., 2010. Model optimization of SVM for a fermentation soft sensor. Expert Syst. Appl. 37, 2708–2713. doi:10.1016/j.eswa.2009.08.008

Llaneras, F., Picó, J., 2008. Stoichiometric modelling of cell metabolism. J. Biosci. Bioeng. 105, 1–11. doi:10.1263/jbb.105.1

Loira, N., Dulermo, T., Nicaud, J.-M., Sherman, D.J., 2012. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Syst. Biol. 6, 35. doi:10.1186/1752-0509-6-35

Luttmann, R., Bracewell, D.G., Cornelissen, G., Gernaey, K. V, Glassey, J., Hass, V.C., Kaiser, C., Preusse, C., Striedner, G., Mandenius, C.-F., 2012. Soft sensors in bioprocessing: a status report and recommendations. Biotechnol. J. 7, 1040–8. doi:10.1002/biot.201100506

Maertens, J., Vanrolleghem, P. a., 2010. Modeling with a view to target identification in metabolic engineering: A critical evaluation of the available tools. Biotechnol. Prog. 26, 313–331. doi:10.1002/btpr.349

Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys. J. 83, 1331–40. doi:10.1016/S0006-3495(02)73903-9

Maher, M., Dahhou, B., Zeng, F.Y., 1995. Experimental results in model reference adaptive estimation and control of a fermentation process. Control Eng. Pract. 3, 313–320. doi:10.1016/0967-0661(95)00002-C

Marler, R.T., Arora, J.S., 2010. The weighted sum method for multi-objective optimization:

New insights. Struct. Multidiscip. Optim. 41, 853–862. doi:10.1007/s00158-009-0460-7

Marler, R.T., Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. 26, 369–395. doi:10.1007/s00158-003-0368-6

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11, 431–441. doi:10.1137/0111030

Meeuwse, P., Akbari, P., Tramper, J., Rinzema, A., 2012. Modeling growth, lipid accumulation and lipid turnover in submerged batch cultures of Umbelopsis isabellina. Bioprocess Biosyst. Eng. 35, 591–603. doi:10.1007/s00449-011-0632-x

Meeuwse, P., Tramper, J., Rinzema, A., 2011. Modeling lipid accumulation in oleaginous fungi in chemostat cultures: I. Development and validation of a chemostat model for Umbelopsis isabellina. Bioprocess Biosyst. Eng. 34, 939–49. doi:10.1007/s00449-011-0545-8

Mo, M.L., Palsson, B.Ø., Herrgård, M.J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst. Biol. 3, 37. doi:10.1186/1752-0509-3-37

Nissen, T.L., Schulze, U., Nielsen, J., Villadsen, J., 1997. Flux Distributions in Anaerobic, Glucose-Limited Continuous Cultures of Saccharomyces Cerevisiae. Microbiology 143, 203–218. doi:10.1099/00221287-143-1-203

Nookaew, I., Jewett, M.C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., Bhumiratana, S., 2008. The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. BMC Syst. Biol. 2, 71. doi:10.1186/1752-0509-2-71

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–8. doi:10.1038/nbt.1614

Pan, P., Hua, Q., 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS One 7, e51535. doi:10.1371/journal.pone.0051535

Papanikolaou, S., Aggelis, G., 2003a. Modeling lipid accumulation and degradation in Yarrowia lipolytica cultivated on industrial fats. Curr. Microbiol. 46, 398–402. doi:10.1007/s00284-002-3907-2

Papanikolaou, S., Aggelis, G., 2003b. Modelling aspects of the biotechnological valorization of raw glycerol: production of citric acid by Yarrowia lipolytica and 1,3-propanediol by Clostridium butyricum. J. Chem. Technol. Biotechnol. 78, 542–547. doi:10.1002/jctb.831

Papanikolaou, S., Aggelis, G., 2002. Lipid production by Yarrowia lipolytica growing on industrial glycerol in a single-stage continuous culture. Bioresour. Technol. 82, 43–49.

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Papin, J. a, Stelling, J., Price, N.D., Klamt, S., Schuster, S., Palsson, B.O., 2004. Comparison of network-based pathway analysis methods. Trends Biotechnol. 22, 400–405. doi:10.1016/j.tibtech.2004.06.010

Patil, D.D., Dangewar, B.D., 2014. Multi-Objective Particle Swarm Optimization (MOPSO) based on Pareto Dominance Approach. Int. J. Comput. Appl. 107, 13–15. doi:10.5120/18738-9983

Peifer, M., Timmer, J., 2007. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. IET Syst. Biol. 1, 78–88. doi:10.1049/iet-syb

Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15, 251–7.

Price, J., Nordblad, M., Woodley, J.M., Huusom, J.K., 2015. Real-time model based process monitoring of enzymatic biodiesel production. Biotechnol. Prog. 31, 585–595. doi:10.1002/btpr.2030

Provost, A., 2006. Metabolic Design of Dynamic Bioreaction Models. Universite catholique de louvain.

Provost, A., Bastin, G., Agathos, S.N., Schneider, Y.-J.-J., 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. Bioprocess Biosyst. Eng. 29, 349–66. doi:10.1007/s00449-006-0083-y

Provost, A., Bastin, G., Schneider, Y.-J., 2007. From Metabolic Networks to Minimal Dynamic Bioreaction Models F, in: IFAC Proceedings Volumes. pp. 1–6. doi:10.3182/20070604-3-MX-2914.00002

Puchałka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., Papin, J.A., Martins dos Santos, V.A.P., 2008. Genome-Scale Reconstruction and Analysis of the Pseudomonas putida KT2440 Metabolic Network Facilitates Applications in Biotechnology. PLoS Comput. Biol. 4, e1000210. doi:10.1371/journal.pcbi.1000210

Rajulapati, S.B., Narasu, L., 2011. Studies On A-Amylase And Ethanol Production From Spoiled Starch Rich Vegetables And Multi Objective Optimization By P . S . O And Genetic Algorithm. IEEE Trans. Automat. Contr. 978, 37–41.

Raman, K., Chandra, N., 2009. Flux balance analysis of biological systems: applications and

challenges. Brief. Bioinform. 10, 435–449. doi:10.1093/bib/bbp011

Ramkrishna, D., 2003. On modeling of bioreactors for control. J. Process Control 13, 581–589.

Renard, F., Vande Wouwer, A., 2008. Robust adaptive control of yeast fed-batch cultures. Comput. Chem. Eng. 32, 1238–1248. doi:10.1016/j.compchemeng.2007.05.008

Reyes-Sierra, M., Coello Coello, C.A., 2006. Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art. Int. J. Comput. Intell. Res. 2, 287–308. doi:10.5019/j.ijcir.2006.68

Riascos, C., Pinto, J.M., 2004. Optimal control of bioreactors: a simultaneous approach for complex systems. Chem. Eng. J. 99, 23–34. doi:10.1016/j.cej.2003.09.002

Robles-Rodriguez, C.E., Bideaux, C., Gaucel, S., Larouche, B., Aceves-Lara, C.A., 2014. Reduction of metabolic models by polygons optimization method applied to Bioethanol production with co-substrates, in: IFAC 2014.

Rodriguez-Nunez, J.R., Castillo-Baltazar, O.S., Pena-Caballero, V., Quintana, F.A.O., Castro, H.B., 2015. Sofware sensors design for monitoring of variables in reactors: A biological processes application, in: 2015 IEEE 2nd Colombian Conference on Automatic Control (CCAC). IEEE, pp. 1–7. doi:10.1109/CCAC.2015.7345219

Ruppin, E., Papin, J. a, de Figueiredo, L.F., Schuster, S., 2010. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. Curr. Opin. Biotechnol. 21, 502–510. doi:http://dx.doi.org/10.1016/j.copbio.2010.07.002

Santos, L.O., Dewasme, L., Coutinho, D., Wouwer, A. Vande, 2012. Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. Comput. Chem. Eng. 39, 143–151. doi:10.1016/j.compchemeng.2011.12.010

Sarkar, D., Modak, J.M., 2005. Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. Chem. Eng. Sci. 60, 481–492. doi:10.1016/j.ces.2004.07.130

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.Ø., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–307. doi:10.1038/nprot.2011.308

Schlegel, M., Stockmann, K., Binder, T., Marquardt, W., 2005. Dynamic optimization using adaptive control vector parameterization. Comput. Chem. Eng. 29, 1731–1751. doi:10.1016/j.compchemeng.2005.02.036

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for

predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. 3, 5. doi:10.1038/msb4100162

Schuster, S., Hilgetag, C., Woods, J.H.H., Fell, D.A.A., 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–81. doi:10.1007/s002850200143

Schwartz, J.-M., Kanehisa, M., 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7, 186. doi:10.1186/1471-2105-7-186

Sharma, S., Tambe, S.S., 2014. Soft-sensor development for biochemical systems using genetic programming. Biochem. Eng. J. 85, 89–100. doi:10.1016/j.bej.2014.02.007

Simutis, R., Lübbert, A., 2015. Bioreactor control improves bioprocess performance. Biotechnol. J. 10, 1115–1130. doi:10.1002/biot.201500016

Sitepu, I.R., Garay, L.A., Sestric, R., Levin, D., Block, D.E., German, J.B., Boundy-Mills, K.L., 2014. Oleaginous yeasts for biodiesel: current and future trends in biology and production. Biotechnol. Adv. 32, 1336–60. doi:10.1016/j.biotechadv.2014.08.003

Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P., Van Impe, J.F., 2004. Optimal adaptive control of (bio)chemical reactors: past, present and future. J. Process Control 14, 795–805. doi:10.1016/j.jprocont.2003.12.005

Song, H.-S., Cannon, W., Beliaev, A., Konopka, A., 2014. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. Processes 2, 711–752. doi:10.3390/pr2040711

Song, H.-S., Morgan, J.A., Ramkrishna, D., 2009. Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol. Bioeng. 103, 984–1002. doi:10.1002/bit.22332

Song, H.-S., Ramkrishna, D., 2012. Prediction of dynamic behavior of mutant strains from limited wild-type data. Metab. Eng. 14, 69–80. doi:10.1016/j.ymben.2012.02.003

Song, H.-S., Ramkrishna, D., 2011. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. Biotechnol. Bioeng. 108, 127–140. doi:10.1002/bit.22922

Song, H.-S., Ramkrishna, D., 2010. Prediction of metabolic function from limited data: Lumped hybrid cybernetic modeling (L-HCM). Biotechnol. Bioeng. 106, 271–284. doi:10.1002/bit.22692

Song, H.-S., Ramkrishna, D., 2009a. When is the Quasi-Steady-State Approximation Admissible in Metabolic Modeling? When Admissible, What Models are Desirable? Ind.

Eng. Chem. Res. 48, 7976–7985. doi:10.1021/ie900075f

Song, H.-S., Ramkrishna, D., 2009b. Reduction of a set of elementary modes using yield analysis. Biotechnol. Bioeng. 102, 554–568. doi:10.1002/bit.22062

Song, H.-S., Ramkrishna, D., Pinchuk, G.E., Beliaev, A.S., Konopka, A.E., Fredrickson, J.K., 2013. Dynamic modeling of aerobic growth of Shewanella oneidensis. Predicting triauxic growth, flux distributions, and energy requirement for growth. Metab. Eng. 15, 25–33. doi:10.1016/j.ymben.2012.08.004

Song, H.-S.S., DeVilbiss, F., Ramkrishna, D., 2013. Modeling metabolic systems: the need for dynamics. Curr. Opin. Chem. Eng. 2, 373–382. doi:10.1016/j.coche.2013.08.004

Soroush, M., 1998. State and parameter estimations and their applications in process control. Comput. Chem. Eng. 23, 229–245. doi:10.1016/S0098-1354(98)00263-4

Srinivas, N., Deb, K., 1994. Muiltiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. Evol. Comput. 2, 221–248. doi:10.1162/evco.1994.2.3.221

Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24, 2229–35. doi:10.1093/bioinformatics/btn401

Tochampa, W., Sirisansaneeyakul, S., Vanichsriratana, W., Srinophakun, P., Bakker, H.H.C., Wannawilai, S., Chisti, Y., 2015. Optimal Control of Feeding in Fed-Batch Production of Xylitol. Ind. Eng. Chem. Res. 54, 1992–2000. doi:10.1021/ie5032937

Trelea, I.C., 2003. The particle swarm optimization algorithm: convergence analysis and parameter selection. Inf. Process. Lett. 85, 317–325. doi:10.1016/S0020-0190(02)00447-7

Trinh, C.T., Wlaschin, A., Srienc, F., 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl. Microbiol. Biotechnol. 81, 813–826. doi:10.1007/s00253-008-1770-1

Tuppadung, Y., Kurutach, W., 2011. Comparing nonlinear inertia weights and constriction factors in particle swarm optimization. Int. J. Knowledge-Based Intell. Eng. Syst. 15, 65–70. doi:10.3233/KES-2010-0211

Üstün, B., Melssen, W.J., Oudenhuijzen, M., Buydens, L.M.C., 2005. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Anal. Chim. Acta 544, 292–305. doi:10.1016/j.aca.2004.12.024

Valentinotti, S., Srinivasan, B., Holmberg, U., Bonvin, D., Cannizzaro, C., Rhiel, M., von Stockar, U., 2003. Optimal operation of fed-batch fermentations via adaptive control of overflow metabolite. Control Eng. Pract. 11, 665–674. doi:10.1016/S0967-0661(02)00172-7

van Gulik, W.M.M., Heijnen, J.J.J., 1995. A metabolic network stoichiometry analysis of microbial growth and product formation. Biotechnol. Bioeng. 48, 681–698. doi:10.1002/bit.260480617

Van Impe, J.., Bastin, G., 1995. Optimal adaptive control of fed-batch fermentation processes. Control Eng. Pract. 3, 939–954. doi:10.1016/0967-0661(95)00077-8

Vanrolleghem, P.A., de Jong-Gubbels, P., van Gulik, W.M., Pronk, J.T., van Dijken, J.P., Heijnen, S., 1996. Validation of a Metabolic Network for Saccharomyces cerevisiae Using Mixed Substrate Studies. Biotechnol. Prog. 12, 434–448. doi:10.1021/bp960022i

Vapnik, V., Golowich, S.E., Smola, A., 1996. Support vector method for function approximation, regression estimation, and signal processing. Annu. Conf. Neural Inf. Process. Syst. 281–287. doi:10.1007/978-3-642-33311-8_5

Varma, A., Palsson, B.O., 1994. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Bio/Technology 12, 994–998. doi:10.1038/nbt1094-994

Vaz, A.I.F., Vicente, L.N., 2007. A particle swarm pattern search method for bound constrained global optimization. J. Glob. Optim. 39, 197–219. doi:10.1007/s10898-007-9133-5

Veloso, A.C.A., Rocha, I., Ferreira, E.C., 2009. Monitoring of fed-batch E. coli fermentations with software sensors. Bioprocess Biosyst. Eng. 32, 381–388. doi:10.1007/s00449-008-0257-x

Wagner, C., Urbanczik, R., 2005. The geometry of the flux cone of a metabolic network. Biophys. J. 89, 3837–45. doi:10.1529/biophysj.104.055129

Wang, X., Chen, J., Liu, C., Pan, F., 2010. Hybrid modeling of penicillin fermentation process based on least square support vector machine. Chem. Eng. Res. Des. 88, 415–420. doi:10.1016/j.cherd.2009.08.010

Wells, C.H., 1971. Application of modern estimation and identification techniques to chemical processes. AIChE J. 17, 966–973. doi:10.1002/aic.690170433

Wiechert, W., 2001. 13C Metabolic Flux Analysis. Metab. Eng. 3, 195–206. doi:10.1006/mben.2001.0187

Williams, D., Yousefpour, P., Wellington, E.M., 1986. On-line adaptive control of a fed-batch fermentation of Saccharomyces cerevisiae. Biotechnol. Bioeng. 28, 631–45. doi:10.1002/bit.260280502

Wong, W.C., Song, H.-S., Lee, J.H., Ramkrishna, D., 2010. Hybrid cybernetic model-based simulation of continuous production of lignocellulosic ethanol: Rejecting abruptly changing feed conditions. Control Eng. Pract. 18, 177–189. doi:10.1016/j.conengprac.2009.09.002

Yamuna Rani, K., Ramachandra Rao, V.S., 1999. Control of fermenters - A review. Bioprocess Eng. 21, 77–88. doi:10.1007/s004490050644

Yin, S., Li, X., Gao, H., Kaynak, O., 2015. Data-based techniques focused on modern industry: An overview. IEEE Trans. Ind. Electron. 62, 657–667. doi:10.1109/TIE.2014.2308133

Ykema, A., Verbree, E.C., Kater, M.M., Smit, H., 1988. Applied Microbiology Biotechnology Optimization of lipid production in the oleaginous yeast Apiotrichum curvature in wheypermeate. Growth (Lakeland) 211–218.

Ykema, A., Verbree, E.C., Van Verseveld, H.W., Smit, H., 1986. Mathematical modelling of lipid production by oleaginous yeasts in continuous cultures. Antonie Van Leeuwenhoek 52, 491–506.

Young, J.D., Ramkrishna, D., 2007. On the matching and proportional laws of cybernetic models. Biotechnol. Prog. 23, 83–99. doi:10.1021/bp060176q

Zadeh, L. a., 1965. Fuzzy sets. Inf. Control 8, 338–353. doi:10.1016/S0019-9958(65)90241-X

Zamboni, N., 2011. 13C metabolic flux analysis in complex systems. Curr. Opin. Biotechnol. 22, 103–108. doi:10.1016/j.copbio.2010.08.009

Zeitz, M., 1987. The extended Luenberger observer for nonlinear systems. Syst. Control Lett. 9, 149–156. doi:10.1016/0167-6911(87)90021-1

Zhang, X., Pan, F., 2015. Parameters Optimization and Application to Glutamate Fermentation Model Using SVM 2015.

Zhao, W., Bhushan, A., Santamaria, A.D., Simon, M.G., Davis, C.E., 2008. Machine Learning: A Crucial Tool for Sensor Design. Algorithms 1, 130–152. doi:10.3390/a1020130

Zitzler, E.E., Deb, K., Thiele, L., 2000. Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. 8, 173–95. doi:10.1162/106365600568202

Zomorrodi, A.R., Maranas, C.D., 2010. Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. BMC Syst. Biol. 4, 178. doi:10.1186/1752-0509-4-178

CHAPTER III.

RESULTS

CHAPTER III. RESULTS

SUMMARY OF THE CHAPTER

In this chapter, the work developed during this thesis is presented as publications. The first section is an introduction to the three articles and the two papers to make some initial comments, summary and highlights of each work.

A methodology to reduce metabolic networks into few pathways is described in the first article (section II). This methodology is based on the concept of convex hull to reduce the number of Elementary modes by optimization methods. Modeling lipid accumulation is addressed in two papers regarding different approaches. The first paper (section III) presents three structured models called Dynamic Metabolic Models that link metabolism with dynamics by the inclusion of a stoichiometric matrix resulted from the reduction of a metabolic network for *Yarrowia lipolytica* accomplished by the methodology from the first paper.

The second modeling approach is introduced in the paper in section IV, which deals with two unstructured models. One of them is incorporated to a control strategy to optimize lipid accumulation. The article in section V describes a multi-objective control attaining the maximization of two objectives: lipid content and lipid productivity. Static Pareto fronts are coupled to a dynamic optimization. This control strategy uses one Dynamic Metabolic Model to simulate the system and to predict the new control states.

The article in section VI introduces the development of soft-sensors to reproduce non-measured state variables of lipid fermentations that are necessary for the control strategy. Three model independent soft-sensors are developed to avoid the addition of model uncertainties to the control loop. These were built on the support vector machines that were optimized by a PSO algorithm.

Finally, section VII comments and highlights the conclusions of the five presented publications.

I. INTRODUCTION

Biotechnological processes require the use of models able of bringing some new insights and knowledge of the systems. In fermentations, when pure cultures are used, models can be employed to describe the metabolism of the interest microorganisms in order to understand the transient states of those processes (Ramkrishna and Song, 2012). The use of metabolism implies the utilization of a metabolic network that is composed by a large amount of reactions relating external metabolites (*i.e.* substrates, and products) through internal metabolites.

In order to simplify the number of reactions, metabolic models (CHAPTER II) are defined under the quasi-steady state approximation stating that the internal reactions occurred faster in comparison to the uptake and production reactions. Under this assumption, it is possible to analyze the metabolic networks to compute all the possible pathways to convert substrates into products. Those pathways are called Elementary Modes (CHAPTER II, section II.2.3). However, the number of elementary modes depends on the number of the considered reactions on the metabolic network.

The article in section II of this chapter proposes a methodology that aimed at reducing the number of Elementary Modes (EMs) being able to trade-off quantity and quality of the reduction. To do so, it was proposed to use polygons with three, four, and five elementary modes, which were plotted into the yield space. Different cases were considered to deal with the available experimental data. The different polygons were tested with literature data (Geng *et al*., 2012) to validate the method. Our results proved that no significant differences were shown when using the proposed n-vertices polygons. Therefore, the smallest polygon, which in this case was the triangle, was preferable for the extension to other applications.

From a mathematical point of view the base of optimization is the development of models able to accurately reproduce the system's dynamics. In a manner to describe lipid fermentations by *Yarrowia lipolytica* and to provide new insights of how the metabolism evolves through the culture, three Dynamic Metabolic Models (DMM) (CHAPTER II, section II) were proposed in the paper from section III. These models are known as structured models, which bring together the metabolic networks with the dynamics of the culture. This was done by the inclusion of a

reduced matrix of elementary modes obtained with the method presented in the article of the section II. The metabolic network used in this case was shown on CHAPTER I section II.7.

The three models are based on the quasi-steady state assumption with a variant, where some internal metabolites have slow dynamics and are allow to accumulate (*i.e.* lipids). Three phases were defined in order to model the phenomena occurring in lipid production: growth phase, lipid accumulation phase, and lipid and citric acid production phase (See the Appendix). The three models were assumed to be sharing a common structure described by the mass balances and the uptake rates, but they diverged in the shifting of metabolic phases. The models were: the Hybrid Cybernetic Model (HCM) which uses cybernetic enzymes to account for regulation between phases; the Macroscopic Bioreaction Model (MBM) considering phase shifting by a manually activation-deactivation; and a new approach defined as Fuzzy MBM that performs the shifts by fuzzy rules based on the uptake ratio of N/C. The three models were calibrated, validated and statistically compared by the *AIC* criterion, proving Fuzzy MBM and HCM as pertinent models. MBM was not compared due to its non-dynamic phase shifting.

Even if the DMM were efficient and could be applied for control purposes, two other models were developed in parallel with the Dynamic Metabolic Models from last section. These two models corresponded to macroscopic models described in a particular paper (section IV). The first model was proposed based on Monod-type and inhibition kinetics (unstructured) and the second on the quota model from Droop (Quota model) to account for time-dependent yields. Instead of using phases, these two models considered overflow metabolism to take into account the concomitant production of citric acid and lipids. This phenomenon was modeled by a Boolean indicator, which was calculated based on the glucose uptake and the nitrogen quota. The quota model was statistically better with a good management of nitrogen concentration.

The optimization of lipid accumulation could be seen as the maximization of two criteria: lipid fraction content and lipid productivity. Based on the bibliographic review (see CHAPTER I, section III.1) managing the N/C ratio could help the optimization. Therefore, the use of two feeds, one of carbon and other of nitrogen source were searched to optimize the two objectives. The quota model was used for the optimization, where the solution was approximated by piecewise linear functions. The results proved good efficiency for optimizing lipid contents and the productivities of this process.

Optimizing the two above mentioned objectives (lipid fraction content and lipid productivity) is often conflicting and complex due to the assignment of adequate weights to each objective, as presented in section III.3 from CHAPTER II. One option to deal with weighting the two-folded objective functions is the use of non-dominated Pareto solutions. The calculation of Pareto fronts can be achieved by the use of population based algorithms such as Particle Swarm Optimization (PSO). However, the calculation of Pareto fronts is static, and moreover PSO does not account with the use of constraints which are necessary for a realistic implementation. Hence, the article in section V couples a Multi-Objective PSO (CHAPTER II, section III.3.3) with a dynamic optimization solver such as pattern search algorithm (CHAPTER II, section III.2.2) to optimize lipids.

The maximization was performed thanks to the Hybrid Cybernetic Model (presented in section II.3.2 of this chapter), which was used to simulate the dynamics of the control strategy. Furthermore, a reduced model was developed to reduce the time of Pareto front computation. The solution of this control strategy aimed at finding the optimal nitrogen and carbon sources feeding policies to maximize lipid fraction and lipid productivity. Three cases were evaluated to prove the efficiency of our control law, where remarkable productivities were obtained.

Sections IV and V prove that lipid optimization could be achieved by different based-model control strategies. Those results were based on simulations where all the state variables were available. Nevertheless, in real life applications, the majority of the experimental data are measured off-line after the fermentation time is ended. The use of sensors, and moreover, the development of soft-sensors is thus required, as presented in section IV from CHAPTER II.

In the article of section VI, three soft-sensors are developed being model independent to avoid the addition of model uncertainties to the control loop. The soft-sensors were built on the Support Vector Machines that were optimized by a PSO algorithm. The use of on-line data was classified and used to define the support vectors that would reconstruct the state variables. SVM is based on learning machines, and they thus depend on the quality and the availability of data. In order to prove their effectiveness, two data sets were used: one and a half were used for the training, while the other was used for the validation. Results proved their fitness quality on the on-line prediction of lipids, biomass, and citric acid. Even if experiments are needed to validate this method, their results proved good efficiency for monitoring this process.

II. REDUCTION OF METABOLIC MODELS BY POLYGONS OPTIMIZATION METHOD APPLIED TO BIOETHANOL PRODUCTION WITH CO-SUBSTRATES

C.E. Robles-Rodriguez[1, 2, 3], C. Bideaux[1, 2, 3], S. Gaucel [4], B. Laroche[5], N. Gorret[1, 2, 3], C.A. Aceves-Lara[1, 2, 3]

[1]Université de Toulouse ; UPS, INSA, INP, LISBP ; F-31077 Toulouse, France
[2]INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France
[3]CNRS, UMR5504, Toulouse, France 135 Avenue de Rangueil, Toulouse Cedex, F-31077
[4]UMR 782 Génie et Microbiologie des Procédés Alimentaires, AgroParisTech – INRA, BP 1, 1 Avenue Lucien Brétignières, 78850 Thiverval-Grignon, France
[5] UMR341 Mathématiques et Informatique Appliquées MIA-Jouy INRA Domaine de Vilvert 78352 Jouy en Josas Cedex

Accepted article for Oral presentation on the 19th World Congress of the International Federation of Automatic Control (IFAC 2014). August 24-29 2014, Cape Town, South Africa

Abstract

In literature metabolic stoichiometric matrix reduction is based on convex analysis by choosing the greatest triangle. This paper proposes a new methodology for the reduction of metabolic networks based on the concept of convex hull by optimization methods. Different polygons are tested to conjointly minimize the squared error (convex hull - experimental data) and maximize the convex hull area in order to reduce the set of metabolic reactions involved in the model. The advantage of this method relies on its ability to select different geometries in a simple manner with the knowledge of the elementary modes. A cybernetic model implementing the proposed optimization method is tested with data for bioethanol production by *Saccharomyces cerevisiae* growing on four substrates. Parameter estimation and model validation allow comparing the performance of the chosen polygons for reduction of metabolic pathways.

II.1. Introduction

Mathematical modelling of biological processes has confronted an ample variety of difficulties that have motivated the study of biological kinetics and its analysis in different manners.

Macroscopic modelling provides dynamical models which have proven enormous interest in bioengineering for the design of the on-line algorithms for bioreactor monitoring, control, and optimization (Bastin and Dochain, 1990). These models have been extended for their use for metabolic modelling.

The kinetic approach relates metabolites concentrations with their corresponding rates. Nonetheless, they require a detailed understanding of reaction mechanisms and regulatory interaction leading to an increasing set of adjustable parameters as models grow more sophisticated (Edwards *et al*., 2002). To overcome these difficulties, stoichiometric models assume that all intracellular concentrations were at steady state, which led to algebraic equations. The inclusion of pseudo-stoichiometric matrix has permitted to lump together the set of intracellular metabolic reactions of the involved microbial species (Bernard and Bastin, 2005). The necessary condition to reach steady state is that the rates of the initial and final reactions (or, equivalently, the concentrations of the initial and final metabolites) must be constant simplifying the mass balance of metabolites (Stephanopoulos *et al*., 1998).

Another approach is cybernetic modelling (Kompala *et al*., 1984; Ramkrishna, 1983) which have the aim to include regulatory effects at the level of enzymes in a way to enrich kinetic and stoichiometric models (Young *et al*., 2008). The complexity of this approach relies on decomposing metabolic networks into elementary pathways for which the standard cybernetic control laws can be applied.

Under the steady state condition, the metabolic network can be decomposed into a set of sub-networks called Elementary Modes (EMs) which are a set of non-decomposable pathways consisting of a minimal set of reactions that function in steady state (Schuster *et al*., 2002). Even the computation of EMs due to external metabolites reduces the metabolic network; the complete set can still be large to be employed in metabolic modelling, so that other considerations must be applied (Song and Ramkrishna, 2009b).

Useful tools based on convex mathematics have drawn to the introduction of convex hull, which main characteristic is its ability to reconstruct any data point inside the convex hull data points (Thurau *et al*., 2009). The points (modes) located in the convex hull are commonly named as generating modes.

Several efforts for reduction of elementary flux modes have been presented by Wagner and Urbanczik (2005) by using Flux Balance Analysis (FBA), which represents an important step towards making full quantitative use of stoichiometry for maximizing biomass yield with information about the metabolic state of the organism (Geng *et al*., 2012).

In addition to FBA, Song and Ramkrishna (2009) reported a reduction method of generating modes (GMs) based on yield analysis (YA) allowing the simplification of the convex hull and its 2-D visualization. YA implies the reduction of convex hull by (i) selecting the active modes (AMs) that form the greatest triangle and (ii) increasing the area of the convex hull by adding one vertex (AMs) at a time to achieve the 99% of the total area of the convex hull. Other works based on YA have been tested to find the 99% of the area (*e.g.* Ant Colony Algorithms) (Aceves-Lara *et al*., 2011).

Following the approach of YA, the objective of this paper is the proposition of an optimization method to reduce the set of GMs by minimizing error with data and maximizing the area. The reduced set of GMs, now called AMs, will be used into the cybernetic model. The methodology is evaluated in yield space to compare different 2-D geometrical configurations (polygons) constructed by the selection of GMs, which only requires the knowledge of EMs, and the calculation of surfaces instead of regarding the distances between them.

II.2. Model Reduction Method

II.2.1. Cybernetic Model

Cybernetic approach is concerned with modelling regulatory processes. It views metabolic regulation as an attempt by cells to make optimal adjustments continually in response to changes in the environment by controlling both the synthesis and the activities of enzymes (Young *et al*., 2008).

Dynamic mass balances of extracellular metabolites in a batch process can be given as:

$$\frac{dx}{dt} = S_x rc \qquad \text{(III.1)}$$

where x is the vector of the n_x concentrations of extracellular components including biomass c. S_x is the n_x x n_r stoichiometric matrix, and r is the vector of n_r intracellular and exchange fluxes expressed per gram of biomass. Under the quasi-steady state approximation, the flux vector r can be represented by a convex combination of EMs (Schuster *et al*., 2002).

$$r = Zr_M \qquad \text{(III.2)}$$

where Z is the n_r x n_Z EMs matrix, and r_M is the vector of the n_Z elementary flux modes, such that (III.1) can be rewritten as:

$$\frac{dx}{dt} = S_x Z r_M c \qquad \text{(III.3)}$$

The expression of r_M fluxes are based on the product of (i) a cybernetic variable $v_{M,j}$ controlling the enzyme activity, (ii) a term of relative level $e_{M,j}^{\text{rel}}$ of enzyme in relation with its maximum value, (iii) and the kinetic term $r_{M,j}^{kin}$ considered for the j^{th} elementary mode. In most of the cases, the latter term is given by the Michaelis-Menten kinetics,

$$r_{M,j}^{kin} = k_j^{\max} \prod_i \frac{x_i}{K_{j,i} + x_i} (i = 1,\ldots,n_x;\ j = 1,\ldots,n_Z) \qquad \text{(III.4)}$$

where $k_j^{\max}$ corresponds to the reaction constant of substrates, and x_i is the concentration of external metabolites. The consumption constant of Michaelis-Menten is represented by $K_{j,i}$ for external metabolites. The kinetic term for enzymes $r_{ME,j}^{kin}$ can be computed from equation (III.4) by replacing the reaction and consumption constants for $k_{E,j}^{\max}$ and $K_{E,j,i}$ respectively.

The enzyme level is determined considering the following dynamic equation,

$$\frac{de_{M,j}}{dt} = \alpha_{M,j} + u_{M,j}.r_{ME,j}^{kin} - (\beta_{M,j} + \mu)e_{M,j} \qquad \text{(III.5)}$$

The parameters $\alpha_{M,j}$ and $\beta_{M,j}$ represent the constitutive synthesis and degradation rate respectively, meanwhile $u_{M,j} r_{ME,j}^{kin}$ is the inducible synthesis rate term regulated by the cybernetic variable $u_{M,j}$, and μ is the dilution rate due to the growth of enzymes.

Following the description of hybrid cybernetic modeling (Song *et al*., 2009), this work focus on seeking different options to reduce the matrix $S_xZ = Z_y$ expressed in function of external metabolites.

II.2.2. Reduction method

The optimization aims at finding the convex hull of the normalized EMs to one compound of the metabolic pathways (*e.g.* biomass/substrate yield). Following the methodology described by (Song *et al*., 2009) the yield vector can be represented by:

$$y = Z_y h,\ h \geq 0,\ \|h\|_1 = 1 \qquad \text{(III.6)}$$

where Z_y is the normalized S_xZ expressed as yields, and h indicates the weight vector implying the contribution of each GM to the total area of the convex hull.

Assuming that data (experimental yield) is available, two cases are taken into account in our method.

[case 1] Data inside the convex hull. The method takes all possible combinations of GMs for the predetermined geometry which include experimental points inside the polygons. In this work we consider polygons of 3, 4, or 5 vertexes

[case 2] Data outside the convex hull. The method looks for the projection of the experimental data into the two closest GMs. Those modes are taken as reference fixed points to construct the set of combinations for the predetermined geometry reducing the number of possible combinations with respect to case 1.

Once all the possibilities of geometric configurations are known, the next optimization problem is established in order to find the best values of Z_yh that,

$$\min_{Z_y h} \sum \left|Z_y h - \bar{y}\right|^T W \left|Z_y h - \bar{y}\right| \quad s.t. \max\ area \qquad \text{(III.7)}$$

From (III.7), Z_yh and $\bar{y}$ are the N_y polygon generated and experimental yield vector respectively, and W is an N_y *x* N_y weighting matrix. *area* is the area of the current polygon. For the sake of evaluating the different polygons, the maximum possible area was considered to be sufficient

to represent all the set of elementary modes. A case of study is presented in the following section to evaluate whether it is possible or not to use this assumption.

If experimental data are not available, the minimization of the sum of square residuals is computed by replacing $\bar{y}$ in equation (III.7) for $t_{theoretical}$ that represents the theoretical yield generated for all the GMs ensuring surface maximization.

II.3. Case of Study

II.3.1. Bio-ethanol production from co-substrates

In the last decades, bioethanol production from lignocellulosic biomass arouses increasing the interest to avoid the use of food crops. After the hydrolysis of lignocellulosic biomass, the remaining materials are glucose, xylose, mannose, galactose, and arabinose (Taherzadeh *et al*., 1997). For this work, the metabolic network considered correspond to the proposed by (Geng *et al*., 2012) to produce ethanol from four substrates: glucose, xylose, mannose and galactose. Notice that oxygen concentration has been omitted for model simplification. Table III.1 shows the 40 reactions involved in the fermentation. The network is mainly composed by glycolytic and pentose phosphate pathways, citric acid cycle, pyruvate metabolism, xylose metabolism, mannose metabolism and galactose metabolism.

- Four substrates: glucose (GLC), xylose (XYL), mannose (MAN), and galactose (GAL).
- Five products: ethanol (ETH), glycerol (GOLx), xylitol (XOLx), carbon dioxide (CO_2), and acetate (ACTx).
- Biomass (Biom) and consumption of excess ATP for maintenance (MAINT). All of them considered as external metabolites.
- Thirty one compounds are considered as internal metabolites.

Saccharomyces cerevisiae is the most traditional microorganism used for bioethanol production. Even its popularity this yeast is not able to ferment xylose (Kotter and Ciriacy, 1993). So in this work, xylose will not be considered for calculations.

The first step was to determine the Elementary modes, which were calculated with METATOOL 2005 (Kamp et al., 2006). A total set of 602 EMs were obtained.

Table III.1. Metabolic network reactions for ethanol production.

v1: GLC + ATP => G6P	v21: X5P + E4P = F6P + GAP
v2: G6P = F6P	v22: S7P + GAP = F6P + E4P
v3: F6P + ATP = DHAP + GAP	v23: PYR => AcCoAm + CO2 + NADHm
v4: DHAP = GAP	v24: OAA + NADH = OAAm + NADHm
v5: DHAP + NADH => GOL	v25: OAAm + AcCoAm => ICT
v6: GOL => GOLx	v26: ICT => AKG + CO2 + NADHm
v7: GAP => PG3 + NADH + ATP	v27: ICT => AKG + CO2 + NADPHm
v8: PG3 = PEP	v28: AKG => SUC + ATP + CO2 + NADHm
v9: PEP = PYR + ATP	v29: SUC = MAL + 0.5 NADHm
v10: PYR => ACD + CO2	v30: MAL = OAAm + NADHm
v11: ACD + NADH => ETH	v31: XYL + 0.5 NADH + 0.5 NADPH => XOL
v12: ACD + NADHm => ETH	v32: XOL => XOLx
v13: ACD => ACT + NADPH	v33: XOL => XUL + NADH
v14: ACT => ACTx	v34: XUL + ATP = X5P
v15: ACT + 2 ATP => AcCoA	v35: 1.04 AKG + 0.57 E4P + 0.11 GOL + 2.39 G6P + 1.07 OAA + 0.99 PEP + 0.57 PG3 + 1.15 PYR + 0.74 R5P + 2.36 AcCoA + 0.31 AcCoAm + 11.55 NADPH + 1.51 NADPHm + 30.48 ATP + 0.43 CO2 => BIOM + 2.68 NADH + 0.53 NADHm
v16: PYR + ATP + CO2 => OAA	v36: ATP => MAINT
v17: G6P => Ru5P + CO2 + 2 NADPH	v37: NADH =>
v18: Ru5P = X5P	v38: MAN + ATP => M6P
v19: Ru5P = R5P	v39: M6P => F6P
v20: R5P + X5P = S7P + GAP	v40: GAL + ATP => G6P

Three groups were identified for the consumption of each substrate individually consisting of 33 EMs for glucose, 33 EMs for mannose, and 33 EMs for galactose, as it was mentioned by (Geng *et al*., 2012).

II.3.2. Method implementation

Each group of elementary modes is normalized for yield analysis according to their respective substrate. Experimental yield data for ethanol and biomass are taken from individual experiments for each substrate reported by (Rouhollah *et al*., 2007) for the implementation of the method. Seven out of thirty tree elementary modes were identified as GMs for each substrate (Figure III.1).

Minimization of square errors and maximization of area is implemented for each substrate. Figure III.2, Figure III.3, and Figure III.4 show the resulting optimized polygons and provide the value of the sum of square errors and the normalized area as a percentage of the total area of the convex hull. In order to appreciate the simplification of this method in 2-D yield space, the phase plane (adding GOLx/GLC yield) is presented in Figure III.2 - Figure III.4 (d) for each geometry. Results for other substrates reported similar structures (Results not shown).

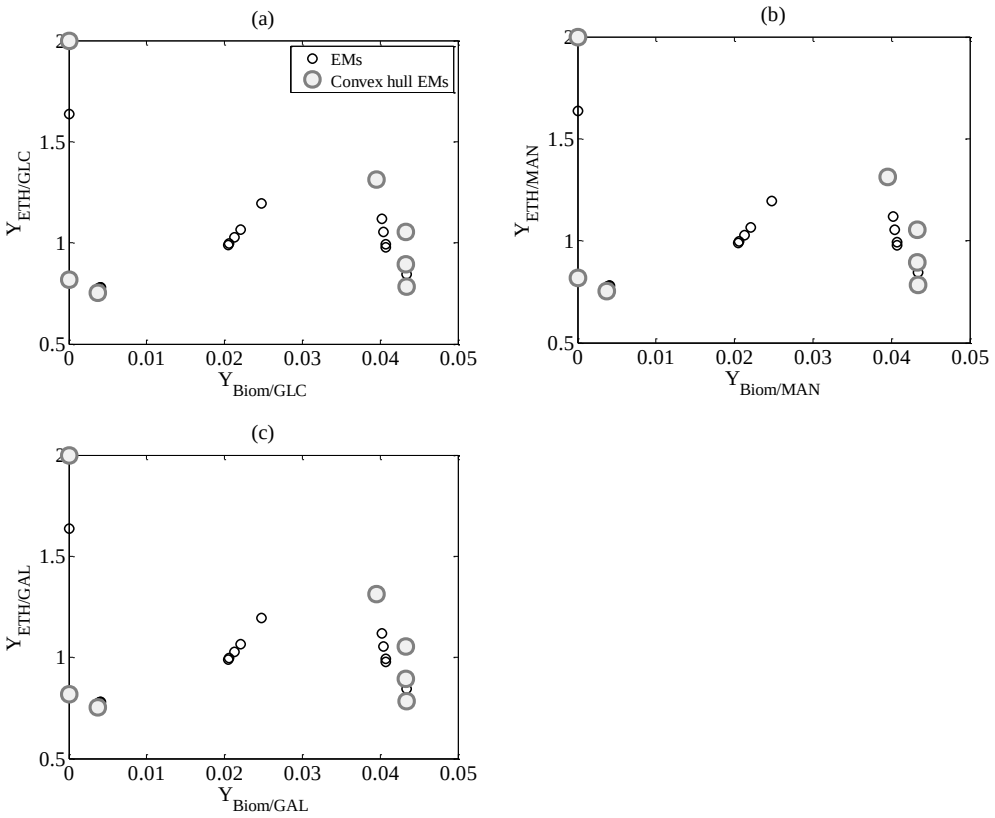


Figure III.1. Elementary Modes representing the metabolic network for (a) Glucose, (b) Mannose, and (c) Galactose.

The calculation of the selected geometry is made based on [case 1] for glucose and galactose where the yield experimental point is observed inside the convex hull, meanwhile [case 2] is applied for mannose.

Regarding the number of possible polygons that can be computed for polygons, it is possible to find 35, 35, and 21 different combinations considering three, four, and five vertexes. In contrast, our optimization method calculates only (8, 5, and 5) polygons with three, (16, 10, and 10) with four, and (14, 10, and 10) with five AMs or vertexes for glucose, mannose, and galactose respectively (Results not shown).

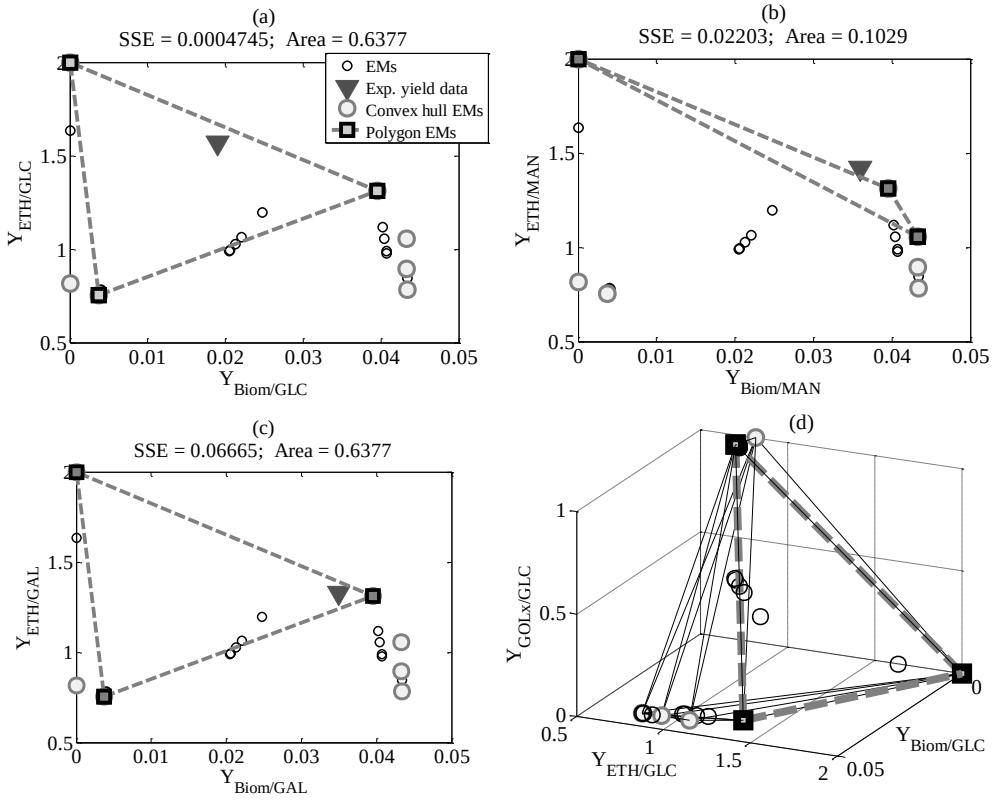


Figure III.2 Optimized three AMs polygon (a) Glucose, (b) Mannose, and (c) Galactose. (d) Phase plane of yields.

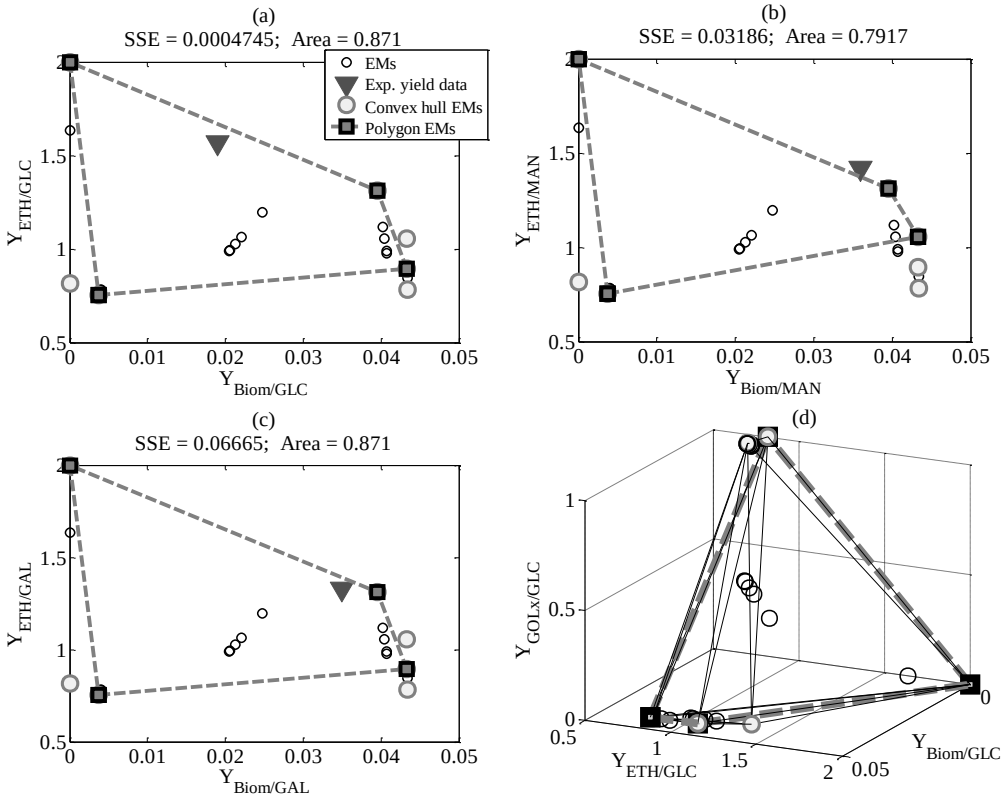


Figure III.3. Optimized four AMs polygon for (a) Glucose, (b) Mannose, and (c) Galactose. (d) Phase plane of yields

As it is stated in Figure III.2, Mannose is not optimized reflecting an area of 10.3% out of the total; however this result can be explained because the experimental yield point is placed outside the convex hull. Similar results are presented by (Geng *et al.*, 2012) where they only chose two elementary extreme modes to describe mannose.

Even though, glucose and galactose are represented by the 63.8% of the area which reflects the well applicability of the method here proposed.

Regarding Figure III.3 and Figure III.4, areas are maximized up to more than 79% and 87% for polygons formed out of four and five points respectively. Glucose reflects the best optimization for both criteria.

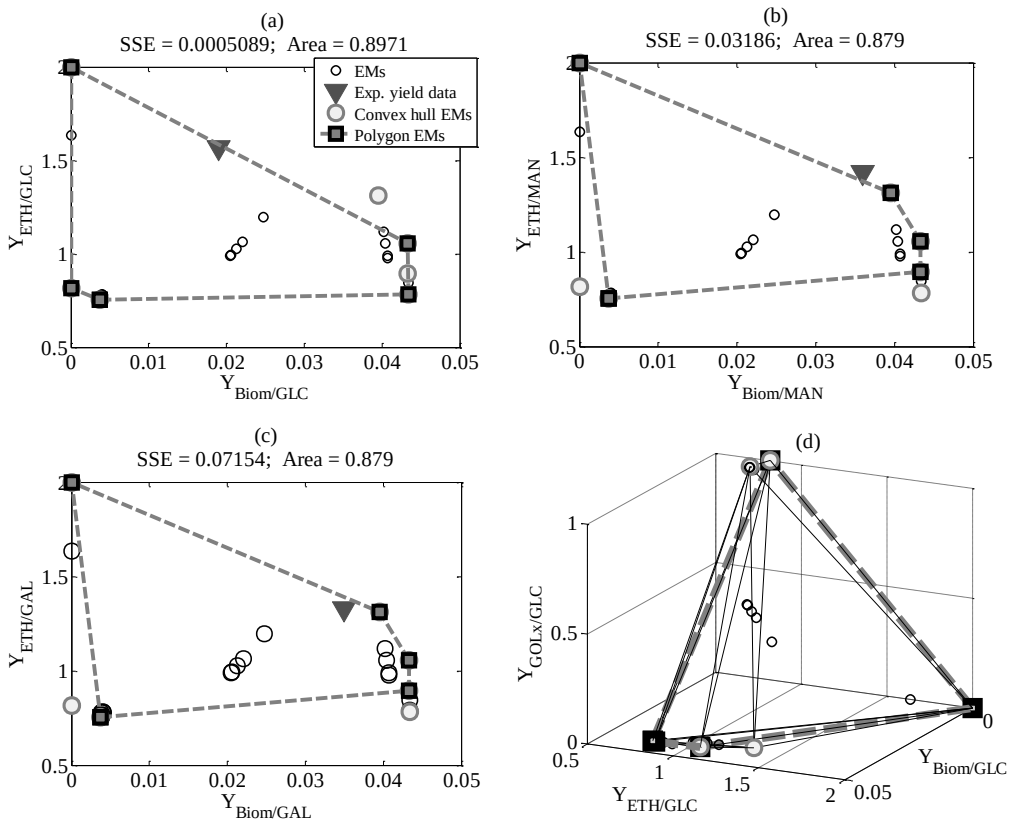


Figure III.4. Optimized five AMs polygon for (a) Glucose, (b) Mannose, and (c) Galactose. (d) Phase plane of yields.

II.3.3. Parameter estimation and validation

The reduced model consists of 9, 12, and 15 reactions instead of 602 (full model) for polygons of three, four, and five vertexes respectively. These reactions are employed to compute the $Z_y = S_x Z$ matrix of equation (III.3).

The experiments reported by (Rouhollah *et al*., 2007) for co-substrates using *Saccaromyces cerevisiae* are taken as data for model validation. (Rouhollah *et al*., 2007) proposed a batch experiment initiated with four substrates glucose, xylose, mannose, and galactose with initial concentrations of 30, 30, 12, 8 $g.L^{-1}$, respectively. As it was mentioned before, xylose is neither considered in the polygons nor for parameter estimation.

Equation (III.4) is taken from (Geng *et al*., 2012) describing metabolites and enzymes as,

$$r_M^{kin} = k_j^{\max} \frac{x_i}{K_{j,i} + x_i} \frac{1}{1 + x_{ETH} / K_{I,i}} \quad \text{(III.8)}$$

$$r_{ME}^{kin} = k_{E,j}^{\max} \frac{x_i}{K_{j,i} + x_i} \frac{1}{1 + x_{ETH} / K_{I,i}} \quad \text{(III.9)}$$

$$(i = 1,\ldots,n_x;\ j = 1,\ldots,n_Z)$$

The Michaelis-Menten constants $K_{j,I}$ and $K_{E,j,i}$ for equations (III.8) and (III.9) have the corresponding values employed by (Geng *et al*., 2012). Notice that as all the parameters are available, thus model is fully identifiable. The optimal values of $k_j^{\max}$ and $k_{E,j}^{\max}$ (Table III.2) are estimated by the Rosenbrock method implemented in MATLAB® to fit the model to the experimental data set described above (Table III.1).

Table III.2. Parameters for the different polygons

AMs	$k_j^{\max}$[1/h]					$k_{E,j}^{\max}$[1/h]				
3	43.63	18.02	36.75	25.31	26.92	0.94	1.45	0.94	1.03	1.03
	20.56	1.02	0.95	0.95		0.95	1.52	1.05	1.32	
4	15.00	15.00	43.63	43.63	24.51	1.25	1.25	0.89	0.89	1.25
	9.40	13.09	144.66	0.95	1.27	1.25	2.46	0.89	1.25	0.89
	0.9	0.9				1.25	1.25			
5	11.28	13.63	13.98	38.44	64.37	1.33	0.91	1.12	0.80	1.06
	24.55	8.03	10.71	11.15	144.98	2.48	2.48	2.48	1.25	1.00
	0.89	1.57	0.89	1.26	1.26	1.25	0.89	1.25	1.25	1.25

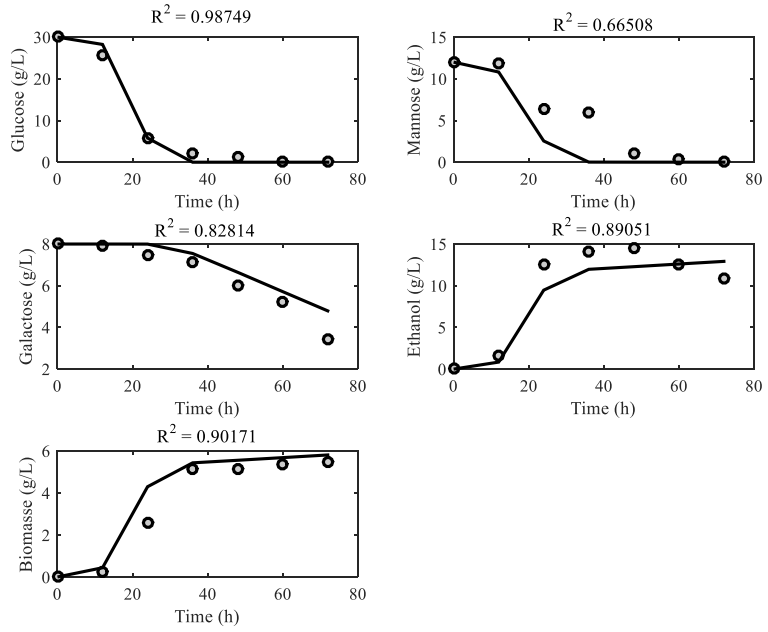


Figure III.5. Performance of the polygon with three vertexes reduced metabolic model fitted to experimental data.

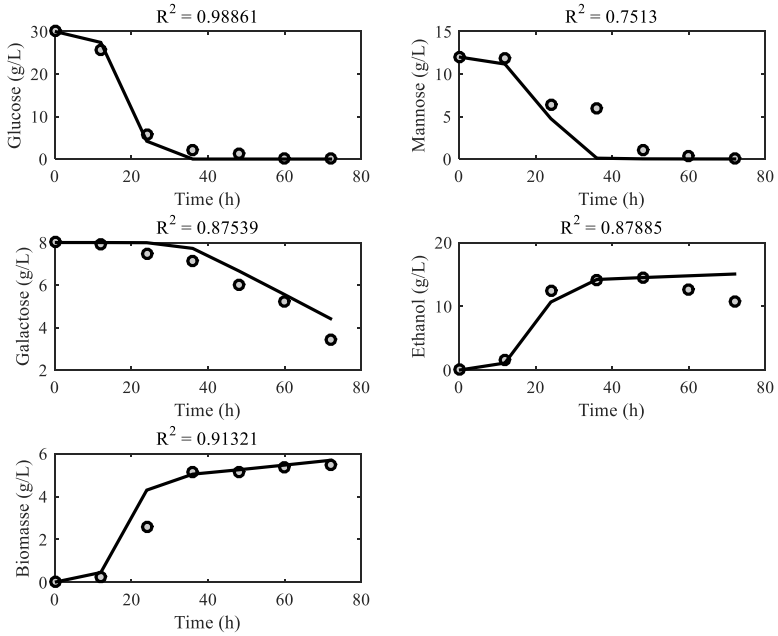


Figure III.6. Performance of the polygon with four vertexes reduced metabolic model fitted to experimental data.

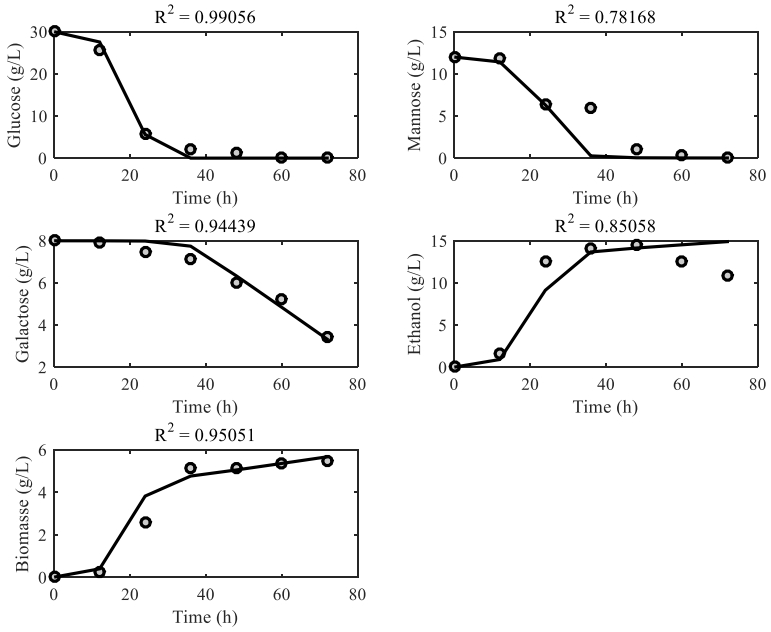


Figure III.7. Performance of the polygon with five vertexes reduced metabolic model fitted to experimental data.

Similar results were found for the estimated values of $k_{E,j}^{max}$ for each vertex of the convex hull (AM) reflecting the conservation of these values within a low threshold of variation. The values of k_j^{max} present large standard deviations due to the coupling of different substrates, nonetheless the order of variation is analogous and intrinsically related to the chosen AMs.

Validation of the model with experimental data is displayed in Figure III.5, Figure III.6, and Figure III.7, which are obtained after the estimation of 18, 24, and 30 parameters for polygons with three, four and five vertexes respectively. The coefficient of determination (R^2) is determined for each substrate and product of the metabolic pathway.

Concerning validity, the Root Mean Square Residuals (RMSE) is evaluated taking into account the number of the *p* estimated parameters as reported by (Dochain and Vanrrolleghem, 2001),

$$RMSE = \frac{(x-\bar{x})^T W (x-\bar{x})}{(N-p)} \quad \text{(III.10)}$$

N represents the number of data, *W* the weighting matrix, *x* and $\bar{x}$ are the data and estimated data respectively. Besides, a coefficient of determination (R^2_T) is calculated by involving all the set of experimental data. Table III.3 reports the performance evaluation of each polygon

reduced metabolic model where the polygon with 5 vertexes (larger area) demonstrate the better R^2_T, but its well-fitting capacity is questioned by the great number of parameters estimated as it is reflected in the RMSE. However, the different polygon configurations exhibit a similar behavior in Figure III.5 - Figure III.7 and analogous results in enforcement evaluation. Notice that Mannose is the substrate with the inferior coefficient of determination due to the position of its experimental yield placed outside of the convex hull. Even though, the maximization of surface improved the fitting capacity of its parameters in the cybernetic model.

Table III.3. Enforcement evaluation of the different polygon based models

Reduced Models by Polygons	RMSE	R^2_T
3 vertexes	0.1970	0.9406
4 vertexes	0.2235	0.9489
5 vertexes	0.3098	0.9508

Even the enforcement evaluation makes difficult to establish the best polygon configuration for model reduction, it is observed that all of them can represent properly the experimental data if parameter estimation is well executed.

II.4. Conclusions

Regarding the optimization of bioprocesses, robust models are needed to guaranty their proper use for several applications (*e.g.* process control). Great number of modelling approaches used metabolic networks information which involves many Elementary Modes that should be reduced. Several methods are proposed in literature, considering complicated calculations and the knowledge of experimental data and the metabolic network. The optimization method presented in this work is a useful tool to simply identify the elementary modes through polygons representing a part of the convex hull of Elementary Modes. This method can be implemented assuming that data is available or not by switching a simple parameter. Even though, it always requires the knowledge of the metabolic network.

The development of the method used in this work has been implemented on data published in the literature. In this application, the simplest polygon is the best choice. Indeed, it reproduces available data and provides the reduced model with a minimum number of parameters. Further

studies will contemplate searching for a trade-off between complexity of the reduced model (analysis of the reactions of the metabolic network chosen by the polygons) and quality of the fitness considering or not the availability of experimental yield data.

II.5. References

Aceves-Lara, C.A., Bideaux, C., Molina-Jouve, C., Roux, G., 2011. Determination of stoichiometric matrix for ethanol production from xylose by reduction of elementary modes with ant colony sustems, in: IFAC 2011. Milano, Italy.

Bastin, G., Dochain, D., 1990. On-line Estimation and Adaptive Control of Bioreactors. Elsevier.

Bernard, O., Bastin, G., 2005. On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes. Math. Biosci. 193, 51–77. doi:10.1016/j.mbs.2004.10.004

Dochain, D., Vanrrolleghem, P., 2001. Dynamical modelling and estimation in wastewater treatment processes. Padstow, Cornwall, UK.

Edwards, J.S., Covert, M., Palsson, B., 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. 4, 133–140. doi:10.1046/j.1462-2920.2002.00282.x

Geng, J., Song, H.-S., Yuan, J., Ramkrishna, D., 2012. On enhancing productivity of bioethanol with multiple species. Biotechnol. Bioeng. 109, 1508–17. doi:10.1002/bit.24419

Kompala, D.S., Ramkrishna, D., Tsao, G.T., 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnol. Bioeng. 26, 1272–81. doi:10.1002/bit.260261103

Kotter, P., Ciriacy, M., 1993. Xylose fermentation by Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 38, 776–783. doi:10.1007/BF00167144

Ramkrishna, D., 1983. Foundations of Biochemical Engineering, ACS Symposium Series. American Chemical Society, Washington, D.C. doi:10.1021/bk-1983-0207

Rouhollah, H., Iraj, N., Giti, E., Sorah, A., 2007. Mixed sugar fermentation by Pichia stipitis , Sacharomyces cerevisiaea , and an isolated xylose- fermenting Kluyveromyces marxianus and their cocultures. African J. Biotechnol. 6, 1110–1114.

Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A., 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–81. doi:10.1007/s002850200143

Song, H.-S., Morgan, J.A., Ramkrishna, D., 2009. Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol. Bioeng. 103, 984–1002. doi:10.1002/bit.22332

Song, H.-S., Ramkrishna, D., 2009. Reduction of a set of elementary modes using yield analysis. Biotechnol. Bioeng. 102, 554–568. doi:10.1002/bit.22062

Stephanopoulos, G.G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies, Metabolic Engineering. doi:10.1016/B978-0-12-666260-3.50019-4

Taherzadeh, M.J., Eklund, R., Gustafsson, L., Niklasson, C., Lidén, G., 1997. Characterization and Fermentation of Dilute-Acid Hydrolyzates from Wood. Ind. Eng. Chem. Res. 36, 4659–4665. doi:10.1021/ie9700831

Thurau, C., Kersting, K., Bauckhage, C., 2009. Convex Non-negative Matrix Factorization in the Wild. 2009 Ninth IEEE Int. Conf. Data Min.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: Fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931. doi:10.1093/bioinformatics/btl267

Wagner, C., Urbanczik, R., 2005. The geometry of the flux cone of a metabolic network. Biophys. J. 89, 3837–45. doi:10.1529/biophysj.104.055129

Young, J.D., Henne, K.L., Morgan, J.A., Konopka, A.E., Ramkrishna, D., 2008. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. Biotechnol. Bioeng. 100, 542–559. doi:10.1002/bit.21780

III. Dynamic Metabolic Modeling of Lipid Accumulation and Citric Acid production by *Yarrowia lipolytica*

Carlos Eduardo Robles-Rodriguez[1], Carine Bideaux[1], Stéphane E. Guillouet[1], Nathalie Gorret[1], Julien Cescut[1], Jean-Louis Uribelarrea[1], Carole Molina-Jouve[1], Gilles Roux[2], César Arturo Aceves-Lara[1*]

[1] LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

[2] LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France

Published in Computers and Chemical Engineering

Abstract

Yarrowia lipolytica has the capacity to accumulate large amounts of lipids triggered by a depletion of nitrogen in excess of carbon source. However, under similar conditions this yeast also produces citric acid decreasing the lipid conversion yield. Three dynamic metabolic models are presented to describe lipid accumulation and citric acid production by *Yarrowia lipolytica*. First and second models were respectively based on the Hybrid Cybernetic Modeling (HCM) and the Macroscopic Bioreaction Modeling (MBM) approaches. The third model was a new approach based on the coupling between MBM and fuzzy sets. Simulation results of the three models fitted acceptably the experimental data sets for calibration and validation. However, MBM is time-dependent to consider metabolic shifts, and thus impractical for further applications. HCM and Fuzzy MBM adequately managed and described metabolic shifts presenting highlighting features for control and optimization. HCM and Fuzzy MBM were statistically compared reflecting similar results.

Keywords: Dynamic metabolic modeling; Elementary modes (EM); Hybrid Cybernetic Model (HCM); Macroscopic Bioreaction Model (MBM); Lipid accumulation; *Yarrowia lipolytica*;

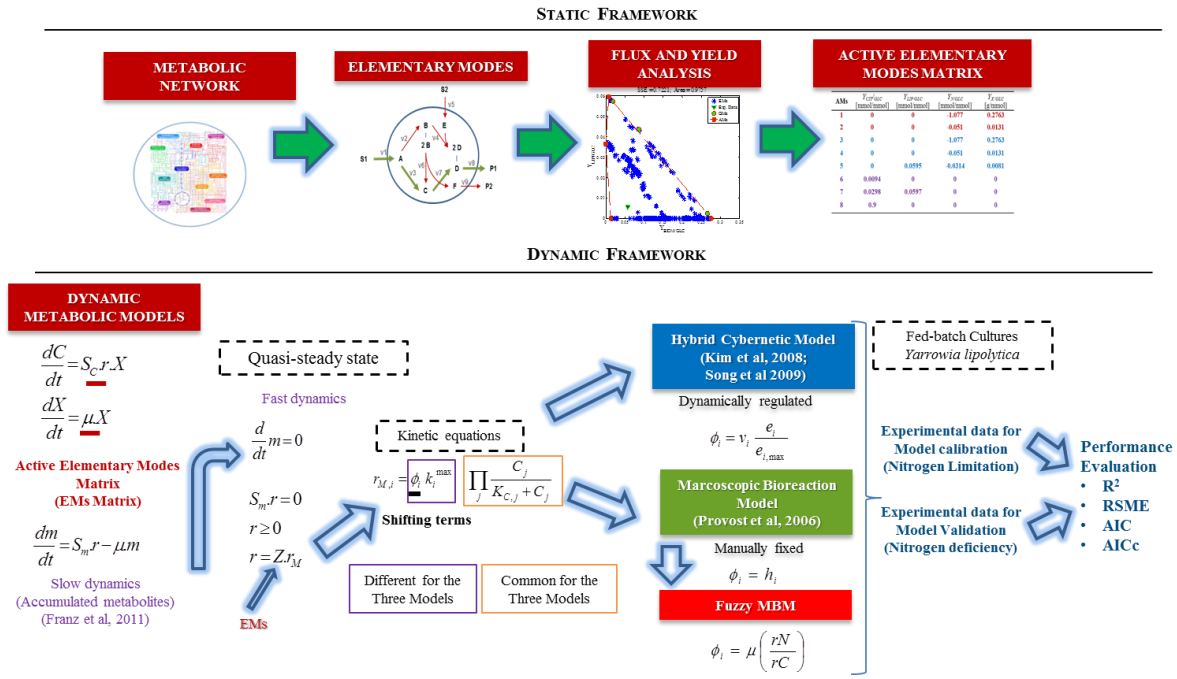


Figure III.8. Graphical Abstract.

III.1. Nomenclature

ATP	Adenosine Triphosphate [mol.mol^{-1}]
b	Fraction of catalytic biomass [-]
C	Concentration of extracellular metabolites [g.L^{-1}]
CIT	Citric acid concentration [g.L^{-1}]
CO_2	Carbon dioxide [mol.mol^{-1}]
e_i	Enzyme level [h^{-1}]
e_i^{max}	Maximum enzyme level [h^{-1}]
e_i^{rel}	Relative enzyme [-]
$f_{C,i}$	Carbon uptake units [Cmol]
F_{IN}	Input flow rate [L.h^{-1}]
F_{Sample}	Sampling flow rate [L.h^{-1}]
g	Weighing factor for convex hull surface [-]
GLC	Glucose concentration [g.L^{-1}]
GLC_{IN}	Glucose input concentration [g.L^{-1}]
h_i	Normalized weights of phases activities [-]
k_i^{max}	Maximum specific rate related to EMs [mmol*GLC*.g*X*$^{-1}$.h^{-1}]
$k_{E,i}$	Reaction rate constant of "enzyme" synthesis for each mode [mmol*GLC*.g*X*$^{-1}$.h^{-1}]
$K_{GLC,i}$	Saturation constant of glucose [g.L^{-1}]

$K_{N,i}$	Saturation constant of nitrogen [mol.L^{-1}]
k_i	Inhibition constant [mol.L^{-1}]
LIP	Total lipids concentration [g.L^{-1}]
LIP_A	Accumulated lipids concentration [g.L^{-1}]
LIP_S	Structured lipids concentration [g.L^{-1}]
m	Fraction intracellular metabolites [gm.gX^{-1}]
m_f	Fraction of intracellular metabolites for fast dynamics [gm_f.gX^{-1}]
m_s	Fraction of intracellular metabolites for slow dynamics [gm_s.gX^{-1}]
n	Number of experimental data
n_C	Number of extracellular metabolites
n_{LIP}	Number of intracellular lipids
n_m	Number of intracellular metabolites
n_P	Number of product
n_{P1}	Number of modes in phase *P1*
n_{P2}	Number of modes in phase *P2*
n_{P3}	Number of modes in phase *P3*
n_r	Number of specific reaction rates
n_S	Number of substrates
n_X	Number of biomass
n_y	Number of experimental yields
n_Z	Number of elementary flux modes
N	Nitrogen concentration [mol.L^{-1}]
N_{IN}	Nitrogen input concentration [mol.L^{-1}]
p	Number of parameters
p_i	Return on the investment
P	Concentration of Product [g.L^{-1}]
P1	Growth phase
P2	Lipid accumulation phase
P3	Citric acid production and lipid accumulation phase
qG_{IN}	Input flow rate of glucose [L.h^{-1}]
qN_{IN}	Input flow rate of nitrogen [L.h^{-1}]
r	Reaction rate vector [h^{-1}]
r_M	Uptake flux vector [h^{-1}]
$r_{M,i}^{kin}$	Kinetics in the uptake flux vector
$r_{EM,i}$	Inducible synthesis rate of cybernetic enzymes
rNC	Ratio of Nitrogen uptake rate over carbon uptake rate [mol$N$.Cmol^{-1}]

S	Concentration of Substrate [g.L^{-1}]
S_C	Stoichiometric matrix of extracellular metabolites
S_{CIT}	Stoichiometric vector of citrate
S_{GLC}	Stoichiometric vector of glucose
S_{LIP_A}	Stoichiometric matrix of intracellular accumulated lipids
S_m	Stoichiometric matrix of the *m* intracellular metabolites
S_{ms}	Stoichiometric matrix of the m_s intracellular metabolites with slow dynamics
S_N	Stoichiometric vector of nitrogen
$S_{X_{Cat}}$	Stoichiometric vector of catabolic biomass
$S_C \cdot Z$	Reduced yield matrix of AMs with respect to extracellular species
$S_{CIT} \cdot Z$	Reduced yield vector of AMs with respect to citrate
$S_{GLC} \cdot Z$	Reduced yield vector of AMs with respect to glucose
$S_{LIP_A} Z$	Reduced yield vector of AMs with respect to intracellular accumulated lipids
$S_N \cdot Z$	Reduced yield vector of AMs with respect to nitrogen
$S_{X_{Cat}} Z$	Reduced yield vector of catalytic biomass
t	Time [h]
t_{12}	Switching time from phase *P1* to phase *P2* in MBM [h]
t_{23}	Switching time from phase *P2* to phase *P3* in MBM [h]
u_i	Cybernetic variable [-]
v_i	Cybernetic variable [-]
V	Volume [L]
W	Weighing matrix for reduction method
X	Total biomass concentration [g.L^{-1}]
X_{Cat}	Catalytic biomass concentration [g.L^{-1}]
$\bar{y}$	Experimental yields for reduction methods [g.g^{-1}]
$Y_{CIT/GLC}$	Citrate/glucose yield [mmol*CIT*.mmol*GLC*$^{-1}$]
$Y_{CO_2/GLC}$	CO_2/glucose yield [mmolCO_2.mmol*GLC*$^{-1}$]
$Y_{LIP_A/GLC}$	Accumulated lipid/glucose yield [mmolLIP_A.mmolGLC^{-1}]
$Y_{N/GLC}$	Nitrogen/glucose yield [mmol*N*.mmol*GLC*$^{-1}$]
Y_{P1}	Experimental specific flux value for phase *P1* [mmol*GLC*.g*X*$^{-1}$.h^{-1}]
Y_{P2}	Experimental yield value of phase *P2* [mol*LIP*.g*X*$^{-1}$]
Y_{P3}	Experimental yield value of phase *P3* [mol*LIP*.mol*CIT*$^{-1}$]
$Y_{X_{Cat}/GLC}$	Catalytic biomass/glucose yield [gX_{Cat}.mmol*GLC*$^{-1}$]
Z	Elementary modes matrix
Z_y	Normalized yield matrix of intra and extracellular metabolites

Abbreviations

AIC	Akaike information criterion
AICc	Correction of Akaike information criterion
AMs	Active modes
area	Surface of the convex hull
EMs	Elementary modes
Fuzzy	Fuzzy sets model
GMs	Elementary modes of the convex hull
HCM	Hybrid Cybernetic model
MBM	Macroscopic Bioreaction model
ODE	Ordinary differential equation
PSO	Particle Swarm Optimization
QSSA	Quasi-steady state approximation
R^2	Determination coefficient
ROI	Return on the investment
RSME	Root Mean Square Errors
SSE	Sum of Square Errors
TAG	Triacylglycerol

Greek symbols

α_i	Constant / constitutive rate of "enzymes"
β_i	Constant / degradation rate of "enzymes"
μ	Growth rate [h^{-1}]
μ_i^{max}	Maximum growth rate [h^{-1}]
ϕ_i	Phase Shifting term
γ	Structural lipids content in catalytic biomass [gLIP_S.gX_{Cat}^{-1}]
λ	*rN/rC* for transition *P1* to *P1-P2* in Fuzzy MBM [mol*N*.Cmol^{-1}]
σ	*rN/rC* for transition *P1-P2* to *P2* in Fuzzy MBM [mol*N*.Cmol^{-1}]
ψ	*rN/rC* for transition *P2* to *P2-P3* in Fuzzy MBM [mol*N*.Cmol^{-1}]
ω	*rN/rC* for transition *P2-P3* to *P3* in Fuzzy MBM [mol*N*.Cmol^{-1}]
ξ_i	Slope for *rN/rC* shifting

Subscripts

Cat	Catalytic
i	Number of elementary mode

Superscripts

kin	Kinetic

max	Maximum
rel	Relative
Operators	
$\frac{d}{dt}$	Derivative of a variable with respect to time
min	Minimization
$\| \ \|_1$	Manhattan norm
$\| \ \|_\infty$	Maximum norm

III.2. Introduction

The oleaginous yeast *Yarrowia lipolytica* accumulates lipids up to 36% of its dry cell weight in the form of triacylglycerol (TAG) (>90%) with a fatty acid profile rich in linoleic acid (C18:2) (56%), oleic acid (C18:1) (28%), and palmitic acid (C16:0) (11%) (Beopoulos et al., 2009a; Meng et al., 2009). This lipid profile is promising aiming at the use of these lipids as precursors for the transesterification of TAGs into biofuels (*e.g.* biodiesel). Other produced lipids are used as membrane constituents (structural lipids), which represent another 5-10% w/w of dry cell weight (Papanikolaou and Aggelis, 2010). In addition to the lipid profile and accumulation, *Yarrowia lipolytica* is a versatile strain for genetic and metabolic engineering approaches for the development of new strains aiming the storage of larger amounts of lipids with a specific fatty acid composition (Nicaud, 2012).

Lipid accumulation is triggered by the depletion of a nutrient in the medium (usually nitrogen) and the excess of the carbon source (*i.e.* glucose) (Cescut, 2009; Ochoa-Estopier, 2012; Ochoa-Estopier and Guillouet, 2014). After nitrogen depletion, this excess of carbon is still assimilated by the microorganisms to be converted into TAGs. Catalytic biomass concentration (Free-TAG Biomass) can also change due to the accumulation of non-lipid materials such as polysaccharides (Beopoulos et al., 2009a). For *Yarrowia lipolytica* growing from glucose, citric acid (Anastassiadis et al., 2002a) is concomitantly produced with lipids, in certain conditions of nutrient depletion and carbon excess, which leads to a decrease of the carbon flux dedicated to TAGs synthesis. Furthermore, if the carbon source is exhausted, storage lipids could also be consumed to produce biomass and to assure the survival of the cell (Beopoulos et al., 2009a).

With the aim of understanding these phenomena, mathematical models have been proposed as straightforward alternatives to help optimizing lipid contents. Several macroscopic models have described lipid accumulation and growth by oleaginous yeasts and fungus (C. N. C. N. Economou et al., 2011; Meeuwse et al., 2012; Papanikolaou and Aggelis, 2003b, 2002; Ykema et al., 1986). These models fitted acceptably the dynamics of lipids and biomass concentrations in continuous and batch processes. Among these modeling approaches, it is worth noting the work from (C. N. C. N. Economou et al., 2011) that describes the growth of the fungus *Mortierella isabellina* as a function of carbon and nitrogen substrates. In this model, a term is included to take into account that lipid concentration is low when nitrogen concentration is high. This structure was assumed to work properly also on yeasts. Another interesting model was proposed to study the effect of the carbon to nitrogen ratio (C/N) of the medium in continuous cultures of the yeast *Apriotrichum curvatum* (Ykema *et al.*, 1986). The C/N ratio is a parameter of major importance to control and to induce the activation-deactivation of the pathways (metabolic shifts) in charge of accumulating lipids and, in the case of *Yarrowia lipolytica*, the further excretion of citric acid (Ochoa-Estopier, 2012; Ochoa-Estopier and Guillouet, 2014). These authors identified the range of the N/C ratio allowing accumulation of lipids without production of citric acid.

Other models (Arzumanov et al., 2000; Papanikolaou and Aggelis, 2003a) were proposed to study citric acid production by *Yarrowia lipolytica* in fed-batch and batch cultures by using ethanol and glycerol as substrates, which showed acceptable fittings with the modeled variables. However, only one model described the production of citric acid and lipids in batch cultures of *Yarrowia lipolytica* growing on a mixture of glucose/stearin (Papanikolaou *et al.*, 2006). Although this model reported good agreements with dynamics of citric acid and biomass concentrations, the results only reported structural lipids, where no accumulation was observed. Even if all the previous models obtained good results, these models neither considered internal regulation of metabolism, nor metabolic shifts, which has to be defined by new approaches integrating kinetics with metabolic models (Ramkrishna, 2003).

The dynamic metabolic models have been proposed in order to account for a metabolic description and regulation of metabolism (Kim *et al.*, 2008; Provost *et al.*, 2006; Ramkrishna and Song, 2012; Song *et al.*, 2009). These approaches allow the definition of strategies to improve culture performances through the identification of rate controlling steps and metabolic fluxes. The dynamic metabolic models are based on the quasi-steady state assumption (QSSA),

in which the intracellular metabolites consumption and production rates are assumed to be faster than the uptake and excretion rates of extracellular metabolites (Song and Ramkrishna, 2009a). This assumption can be used in order to decompose metabolic networks into a set of sub-networks called Elementary Modes (EMs). EMs are a set of non-decomposable pathways consisting of a minimal set of reactions in steady state (Schuster *et al.*, 2002). Based both on EMs decomposition, the Macroscopic Bioreaction Models (MBM) (Provost *et al.*, 2006) and the Hybrid Cybernetic Models (HCM) (Kim *et al.*, 2008; Song *et al.*, 2009) are two interesting modeling approaches. These two approaches have already been applied to model Chinese hamster ovary cells, and the growth of the yeast *Saccharomyces cerevisiae*, respectively. These two models fitted well the dynamics of biomass and substrates concentrations considering metabolic shifts.

In the present study, III.3 describes HCM, MBM, and a third model to consider a non-time dependent shifting of MBM through fuzzy sets. The dynamic metabolic models are implemented for lipid accumulation and citric acid production by *Yarrowia lipolytica* based on a simplified metabolic network. In III.4, three data sets were used to calibrate and validate the models. The data sets were performed in fed-batch and sequential batch cultures of *Yarrowia lipolytica* from glucose under nitrogen limitation and deficiency conditions. One data set and half of the second are employed for the identification of one set of parameters for each model. The models were further validated with the rest of the second data and the third independent data set. Finally, the performances of the three models were compared and evaluated by statistical criteria.

III.3. Methods

III.3.1. Experimental data of lipid accumulation

The microorganism of study was *Yarrowia lipolytica* wild type W29. This strain was obtained from the Microbiology and Molecular genetics laboratory (Paris-Grignon, France). The strain was cultivated on synthetic mineral medium, as shown by (Ochoa-Estopier and Guillouet, 2014).

III.3.1.1. Fed-batch culture under nitrogen limitation

This experimental data set (**Exp.1**) was reported by (Cescut, 2009), where *Yarrowia lipolytica* was grown in a 20 L bioreactor using the Braun Biostat E fermenting system (Braun, Melsungen, Germany) without oxygen limitation. Operating conditions were fixed and controlled to a temperature of 28 °C and a pH value of 5.5, which was regulated by the addition of a 10 M NH3 solution (growth phase) or KOH solution (after nitrogen limitation). Input and output airflow rates, and their composition were measured online as well as the agitation, which was considered homogenous. Glucose (730 $g.L^{-1}$) and ammonia (5 M) were fed continuously (Figure III.9 a). A custom-made software was used for online data acquisition and pumps commands.

At the beginning of **Exp.1**, biomass concentration was around 8 $gDW.L^{-1}$, which resulted from the pre-culture. The feeding of both, nitrogen and glucose, started when biomass concentration arose 23 $gDW.L^{-1}$. At this moment (4 h), nitrogen limitation was set. The maximum lipid concentration (18 $g.L^{-1}$) was achieved at 37 h, when citric acid was also produced. At the end of the culture, the maximal concentration of citric acid was 97 $g.L^{-1}$.

In this experiment, three metabolic phases were defined for modeling purposes:
(i) Growth phase (*P1*): All nutritional requirements were included in the initial medium allowing growth without any limitation. Nitrogen limitation was set at the end of this phase by shifting the pH corrector solution from the ammonia solution to the KOH solution.
(ii) Lipid accumulation phase (*P2*): Lipids accumulate due to an increase of carbon feeding rate and a decrease of nitrogen feeding rate.
(iii) Citric acid and lipid production phase (*P3*): The excess of carbon source with a severe nitrogen limitation leads to concomitant citric acid excretion and lipid accumulation.

III.3.1.2. Fed-batch cultures under nitrogen deficiency

Two data sets (Exp.2 and Exp.3) were obtained from a 5*L* stirred tank bioreactor using the Biostat B. Braun Biotech International (Sartorius AG, Germany) with the acquisition software MFCS/win 2.1®. Operating conditions were set to a controlled temperature of 28°C and a pH value of 5.6 regulated by a 5*M* KOH solution. Only two phases were identified in these cultures:

Growth phase (*P1*), and Citric acid and lipid production phase (*P3*). The working volume of the reactor was 3*L*.

Exp.2. Glucose was fed by pulses at 35*h* and 55*h* (Figure III.9 b). In this case, 0.031$mol.L^{-1}$ of $(NH_4)_2SO_4$ were added to the initial medium.

Exp.3. Glucose was fed by one pulse at 19.3*h* (Figure III.9 c). In this case, 0.068$mol.L^{-1}$ of $(NH_4)_2SO_4$ were added to the initial medium.

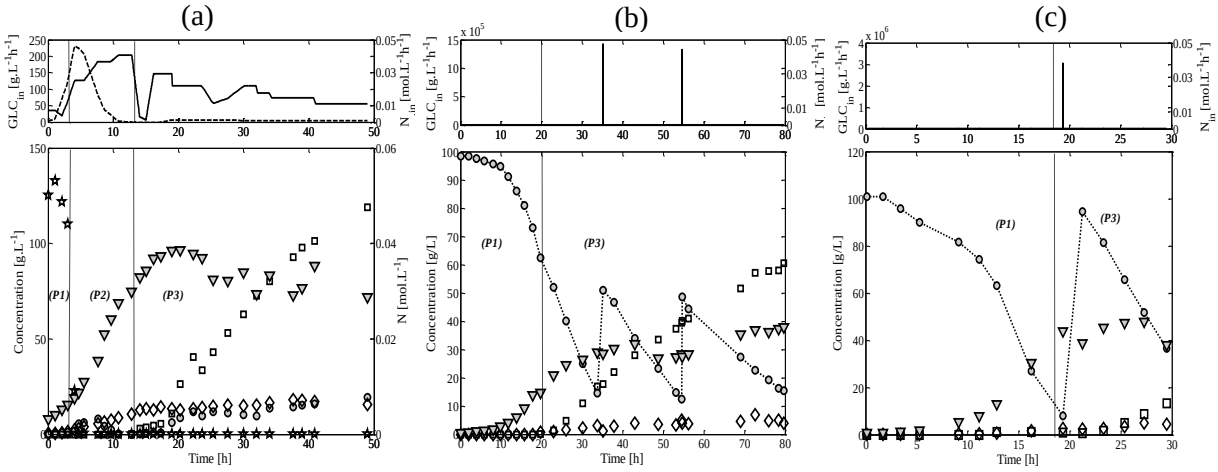


Figure III.9. Data of cultures of *Yarrowia lipolytica* growing on glucose. (a) Exp.1 Fed-Batch Culture under nitrogen limitation (Cescut, 2009)**. Sequential batch cultures under nitrogen deficiency: (b) Exp.2. (c) Exp.3. The solid line (-) represent the input flow rate of glucose, and the dashed line (--) the input flow rate of ammonia. (●) Glucose, (□) Citrate, (◊) Lipids, (▼) Biomass, (☆) Nitrogen.**

III.3.1.3. Analytical methods

Yeast concentration was determined by spectrophotometry at 600 nm in a spectrophotometer HITACHI U-1100. For dry cell weight determination, culture samples (5–10 mL) were harvested by filtration on a 0.45 μm membrane (Sartorius), and dried at 200 mm Hg and 60 °C during 48 h until a constant weight was achieved. Determination of the alcohols, organic acids, and sugars concentrations from supernatants was performed by HPLC using a column Aminex HPX-87H (300 mm*7.8 mm) with the following conditions: Temperature of 50 °C, with 5 mM H_2SO_4 as eluent (flow rate of 0.5 mL min^{-1}), and dual detection (refractometer and UV at 210 nm). Compounds were identified and quantified with respect to standards.

An ammonium ion electrode (PH/ISE meter model 710A + Ammonia Gas Sensing Electrode Model 95-12, Orion Research Inc. Boston USA) was used to quantify the residual ammonia concentration in the culture medium.

For total lipids extraction, the procedure consisted in three consecutive extractions in methanol/chloroform at different concentrations (2:1, 1:1, and 1:2 v/v), where 500 mg of dry mass cells in 15 mL represented one volume. Each extraction step consisted of incubation for 24 h at room temperature on roller mixer. As the extraction could "extract" non-lipid contaminants (sugars, amino acids, proteins and salts), the organic phase was rocking against a KCL solution (0.08 $g.L^{-1}$) for 15 min, and then a liquid/liquid separation technique was used after centrifugation (5000 g, 10 min). Finally, lipids were recovered as dry material after solvent evaporation in a Rota vapor (35 °C). Total lipids content was quantified by gravimetric methods. Lipid extracts were kept in chloroform at –20 °C. Fatty acids were methylated into Fatty Acid Methyl Esters (FAME) using trimethyl sulfonium hydroxide (TMSH, 0.2 M in methanol, Macherey-Nagel, Germany). GC analysis were carried out on a Hewlett-Packard 5890 gas chromatograph equipped with a 50 m*250 µm*25 µm WCOT Fused Silica column (VARIAN, USA) and a FID detector. The carrier gas was nitrogen with a 50 $mL.min^{-1}$ flow. The oven temperature was varied in different gradients: first 9 $°C.min^{-1}$ for a range between 50-75 °C, after 13 $°C.min^{-1}$ for 75-140 °C and 140-180 °C, then 1.5 $°C.min^{-1}$ for 180-240 °C, and finally 4.5 $°C.min^{-1}$ for an injection temperature of 140 °C with a detector temperature of 250 °C. Identification and quantification of methyl esters were based on the comparison with known standards.

III.3.2. Framework of Dynamic Metabolic Modeling

III.3.2.1. Dynamic Modeling description in batch

Let us consider a microorganism consuming substrates S (n_S x 1), converted into intracellular metabolites to synthesize biomass X (n_X x 1), and to excrete products P (n_P x 1). In batch mode, the mass balance equations on each compound within the bioreactor, which was assumed to be perfectly mixed with a constant volume, are written as follows,

$$\frac{dC}{dt} = S_C\, r\, X \qquad \text{(III.11)}$$

$$\frac{dm}{dt} = S_m\, r - \mu\, m \tag{III.12}$$

$$\frac{dX}{dt} = \mu\, X \tag{III.13}$$

where $C = [S, P]^T$ is the (n_C x 1) vector of extracellular metabolites concentrations ($n_C = n_S + n_P$). S_C (n_C x n_r) and S_m (n_m x n_r) are the stoichiometric matrices from the reactions involving the extracellular and intracellular metabolites respectively. The vector r (n_r x 1) represents the specific reaction rates (per unit of biomass) of each reaction in the metabolic network, whilst μ is of specific growth rate.

The quasi-steady state assumption (QSSA), also called balance-growth condition, establishes that the response time of the rates of intracellular metabolites are faster than the extracellular ones. Therefore, the time derivative of intracellular metabolites concentration (III.12) becomes zero,

$$\frac{dm}{dt} = 0 \tag{III.14}$$

Nevertheless, some of these internal metabolites (*i.e.* lipids) could be accumulated inside the cell during the cultures. (Franz *et al.*, 2011) assumed the separation of intracellular metabolites into two subgroups according to their dynamics: m_f and m_s, fast and slow dynamics, respectively. Thus, equation (III.12) can be redefined,

$$\frac{dm}{dt} = \begin{bmatrix} \frac{dm_s}{dt} \\ \frac{dm_f}{dt} \end{bmatrix} = \begin{bmatrix} S_{m_s}\, r - \mu\, m_s \\ 0 \end{bmatrix} \tag{III.15}$$

It is worth to note that for m_f, the quasi-steady state assumption holds, whereas for m_s the equation (III.12) is maintained.

If under the balance-growth condition ($\frac{dm}{dt} = 0$) the term of dilution with respect to biomass, $\mu\, m_s$, is considered negligible, it is possible to define

$$S_{m_s}\, r = 0 \tag{III.16}$$

Consequently, the n_r reaction rates will only depend on the stoichiometry of the metabolic network, and the uptake of substrates. Equation (III.16) has an admissible solution if each component of the flux vector r satisfies the constraint $r \geq 0$, which implies that all the fluxes

must be considered as irreversible. In this context, let us define the concept of elementary models (EMs), which describe the set of all possibilities of metabolic pathways to transform substrates into products (Schuster *et al*., 2002). Hence, the solution of equation (III.16) corresponds to a cone containing non-negative combinations of EMs. The flux vector *r* can be thus expressed as a combination of EMs as $r = Z.r_M$, where *Z* (n_r x n_Z) is the matrix of EMs normalized with respect to a substrate, and r_M is the kinetic (n_Z x 1) vector that describes the n_Z elementary flux modes. The set of EMs can be computed by different software implemented in MATLAB®. The most popular are METATOOL (Kamp et al., 2006), *efmtool* (Terzer and Stelling, 2008), and CellNetAnalyzer (Klamt *et al*., 2007).

III.3.2.2. Reduction of Elementary Modes

In a real metabolic scheme the number of reaction is high, and consequently the number of EMs (*e.g.* 100 reactions generate over 50 thousand EMs). A reduction of the solution space is thus needed. A selection must be done by identifying the EMs that best represent the system, also called active modes (AMs). This identification was based on a modified methodology of yield analysis (Robles-Rodriguez *et al*., 2014; Song and Ramkrishna, 2009b). This methodology implied an optimization based on the simultaneous minimization of the error between the centroid and the experimental data, and the maximization of the area of the convex hull. For instance, this allows the reduction of the convex hull by eliminating the trivial points with negligible contribution maintaining a set that can describe the original convex hull in an appropriate level (Song and Ramkrishna, 2009b).

$$\min_{\substack{Z_y \cdot g \\ s.t. \max area}} \sum |Z_y\, g - \overline{y}|^T W |Z_y\, g - \overline{y}| \qquad \text{(III.17)}$$

Here, the term Z_y represents the yield matrix of intra and extracellular metabolites with respect to the substrate, which is compared with the experimental yield $\overline{y}$. The weighting factor g is included to measure surface impact of each mode constrained to $g \geq 0$, and $\|g\|_1 = 1$. W (n_y x n_y) is a weighting matrix computed as the inverse of the square of the maximum value of the experimental yield $\overline{y}$, and *area* is the surface of the current polygon.

III.3.2.3. Metabolic network for lipid accumulation

The reduced metabolic network of *Yarrowia lipolytica* used in these models is displayed in Figure III.10. This network only considered two compartments (cytosol and mitochondria) in order to simplify transport reactions. Acetyl-Coenzyme A, and cofactors as NADH and NADPH, are described as being unable to freely cross the mitochondrial membrane. The network consists of 59 reactions with 43 internal metabolites, and 8 principal metabolites (Glucose, Citric acid, TAG, Nitrogen, CO_2, Maintenance, Biomass and Oxygen). In the network, biomass reaction was taken from (Cakir *et al.*, 2007). This reaction was completed to assume a constant fraction of structural lipids equivalent to 0.06 gLIP_S.gX_{Cat}^{-1}. This is consistent with the values reported by (Papanikolaou and Aggelis, 2010).

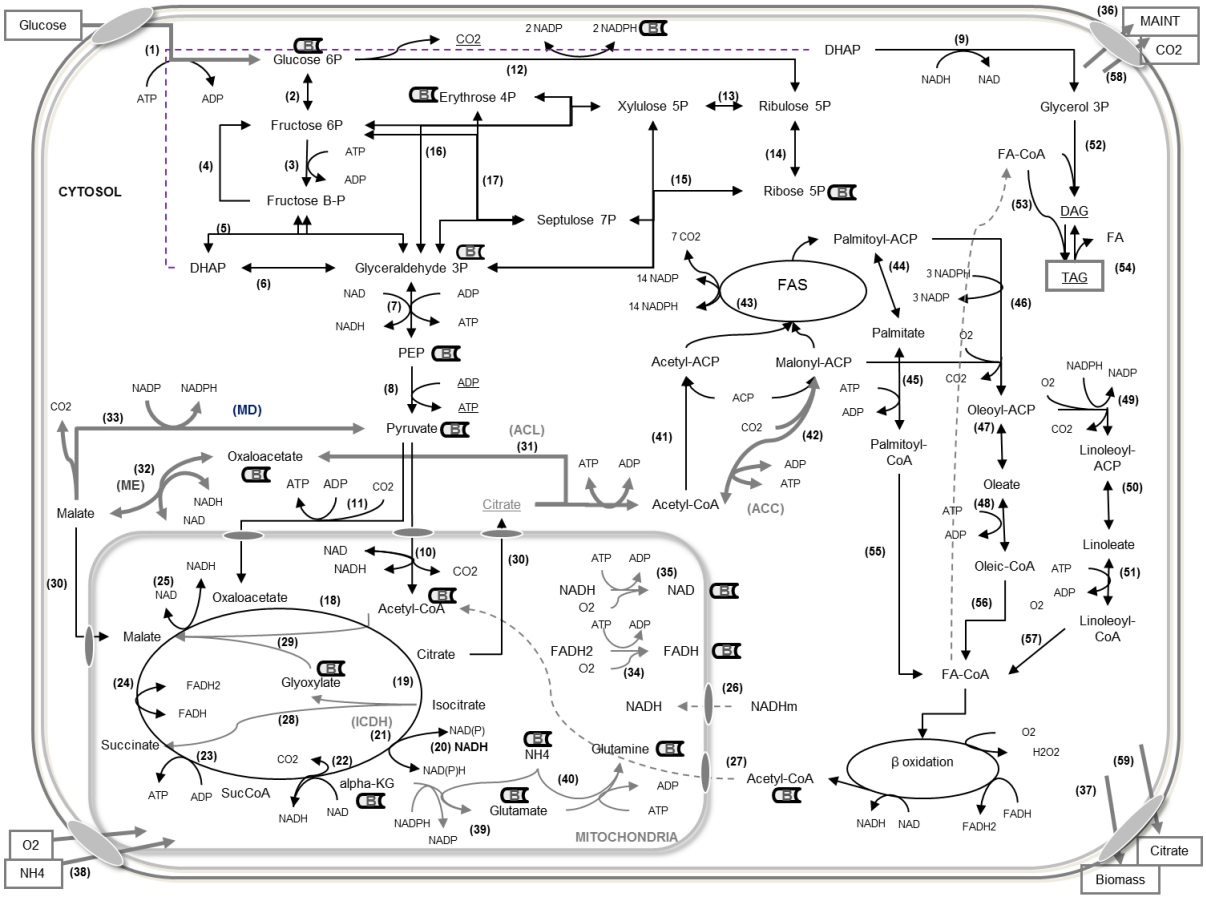


Figure III.10. Reduced Metabolic Network for *Yarrowia lipolytica*. (CBI) Biomass components.

III.3.2.4. Dynamic metabolic model for lipid accumulation

Based on the metabolic network presented in last section, and the available experimental data, 5 of the 8 principal metabolites (Glucose, Citric acid, TAG, Nitrogen, and Biomass) were taken into account for the modeling approach.

The model is based on the following assumptions:

- A carbon source (glucose *GLC*) and a nitrogen source (*N*) are essential for growth.
- Accumulated lipids (LIP_A) are assumed to be the only intracellular metabolite in the form of triacylglycerols (TAGs). Consequently, $m = \frac{LIP_A}{X}$ in equation (III.12). However, for the sake of comparison with experimental data, LIP_A are considered as external metabolites.
- Total lipids (*LIP*), also considered as external metabolites, correspond to the sum of accumulated lipids (LIP_A) and structural lipids (LIP_S)
  $LIP = LIP_A + LIP_S$
  For simplicity and consistency with experimental data, only Total lipids (*LIP*) are described in the mass balances.
- The structural lipids (LIP_S) are intrinsic to the catalytic biomass (X_{Cat}) in a constant fraction γ, which is considered in the biomass reaction of the metabolic network as 0.06 gLIP_S.gX_{Cat}^{-1},
  $LIP_S = \gamma X_{Cat}$.
- A new biomass is incorporated, called Catalytic Biomass (X_{Cat}), representing the biomass without accumulated lipids.
- Total biomass (*X*) is the sum of catalytic biomass (X_{Cat}) and accumulated lipids (LIP_A).
  $X = X_{Cat} + LIP_A$
- The specific growth rate μ of the total biomass *X* comprises the sum of catalytic biomass (X_{Cat}) and accumulated lipids (LIP_A) yield matrices.

Mass balance equations for the principal metabolites in lipid accumulation by *Yarrowia lipolytica* are set-up for fed-batch and sequential batch cultures with sampling volumes and no oxygen transfer limitations as follows,

$$\frac{dGLC}{dt} = S_{GLC}Z\, r_M\, X - \frac{GLC}{V}F_{IN} + \frac{GLC_{IN}}{V}\, qG_{IN} \quad \text{(III.18)}$$

$$\frac{dN}{dt} = S_N Z\, r_M\, X - \frac{N}{V}F_{IN} + \frac{N_{IN}}{V}\, qN_{IN} \quad \text{(III.19)}$$

$$\frac{dCIT}{dt} = S_{CIT}Z\, r_M\, X - \frac{CIT}{V}F_{IN} \quad \text{(III.20)}$$

$$\frac{dLIP}{dt} = \left(S_{LIP_A}Z + \gamma\, S_{X_{Cat}}Z\right) r_M\, X - \frac{LIP}{V}F_{IN} \quad \text{(III.21)}$$

$$\frac{dX}{dt} = \left(S_{X_{Cat}}Z + S_{LIP_A}Z\right) r_M\, X - \frac{X}{V}F_{IN} \quad \text{(III.22)}$$

$$\frac{dX_{Cat}}{dt} = S_{X_{Cat}}Z\, r_M\, X - \frac{X_{Cat}}{V}F_{IN} \quad \text{(III.23)}$$

$$\frac{dV}{dt} = F_{IN} - F_{Sample} \quad \text{(III.24)}$$

with $F_{IN} = qG_{IN} + qN_{IN}$

The terms $S_{GLC}Z$, S_NZ, $S_{CIT}Z$ and $S_{LIP_A}Z$ are the reduced yield vectors (1 x n_r) of AMs with respect to glucose of the principal metabolites (*GLC, N, CIT,* LIP_A). The concentrations of (*GLC, CIT, LIP*) are expressed in (g.L^{-1}), whereas the concentration of *N* is expressed in (mol.L^{-1}). The reduced yield matrix of catalytic biomass is $S_{X_{Cat}}Z$ (1 x n_r). The volume *V* in equation (III.24) accounts with a term of output flow rate, F_{Sample}, to deal with sampling volumes.

In order to model the physiological changes of metabolism throughout the fed-batch and sequential batch cultures, three phases were defined concerning the quasi-steady state of intracellular metabolites (see section III.3.1). The uptake flux vector $r_{M,i}$ was defined as,

$$r_{M,i} = \phi_i\, k_i^{max}\, r_{M,i}^{kin} \quad \text{(III.25)}$$

where ϕ_i is a shifting term used to differentiate the modeling approaches, where k_i^{max} is the maximum specific rate and $r_{M,i}^{kin}$ the reaction rate of each i^{th} AM. Kinetics was defined to be common for the three models and some inhibitions were contemplated as in (Song *et al*., 2009). Growth phase (*P1*) was described by a Monod-type equation with a double substrate limitation on nitrogen and glucose. Lipid accumulation phase (*P2*) was assumed to follow a Monod-type dependence on glucose. Nitrogen limitation was considered as in (C. N. C. N. Economou et al., 2011) to ensure that under high nitrogen concentrations, lipids were not accumulated. An inhibition by citric acid was added to express the competition of lipids and citric acid production. This inhibition was also considered for the citric acid and lipid production phase (*P3*), where experimental data proved that glucose was not fully consumed.

$$P1 \rightarrow r_{M,i}^{kin} = \left(\frac{GLC}{K_{GLC,1} + GLC}\right)\left(\frac{N}{K_{N,1} + N}\right) \qquad (i = 1, \dots, n_{P1}) \qquad \text{(III.26)}$$

$$P2 \rightarrow r_{M,i}^{kin} = \left(\frac{GLC}{K_{GLC,2} + GLC}\right)\left(\frac{1}{1 + CIT/k_1}\right) \qquad (i = 1, \dots, n_{P2}) \qquad \text{(III.27)}$$

$$P3 \rightarrow r_{M,i}^{kin} = \left(\frac{GLC}{K_{GLC,3} + GLC}\right)\left(\frac{1}{1 + CIT/k_2}\right) \qquad (i = 1, \dots, n_{P3}) \qquad \text{(III.28)}$$

$K_{GLC,i}$, and $K_{N,1}$ are the saturation constants for glucose and nitrogen respectively, and k_1 and k_2 represent the inhibition constants due to citric acid concentration. The n_{P1}, n_{P2}, and n_{P3} of AMs represent the number of modes in phases *P1*, *P2*, and *P3*, respectively ($n_r = n_{P1} + n_{P2} + n_{P3}$).

The dynamic metabolic models for lipid accumulation, here presented, are based on equations ((III.18) – (III.28)). However, the product of the shifting term and the maximum specific rate $\phi_i \, k_i^{max}$ allow us to differentiate between modeling approaches. Three dynamic metabolic models are described in the following subsections, where two were based on literature (HCM and MBM), and the third is a new proposition of the authors called Fuzzy MBM.

III.3.2.5. Hybrid Cybernetic model (HCM)

Hybrid Cybernetic Model (HCM) (Kim *et al.*, 2008; Song *et al.*, 2009) includes regulatory effects in the vector of uptake rates $r_{M,i}$ at the level of "enzymes" (one "enzyme" per EM). Therefore, the dimension of the model is bounded to the number of EMs. The shifting term ϕ_i of the uptake rate vector $r_{M,i}$ (equation(III.25)) is regulated for each i^{th} EM by a vector of cybernetic variables v_i, controlling enzyme activity,

$$\phi_i = v_i \, e_i^{rel} \qquad \text{(III.29)}$$

where $e_i^{rel} = \frac{e_i}{e_i^{max}}$, and $e_i^{max} = \frac{\alpha_i + k_{E,i}}{\beta_i + \mu_i^{max}}$.

The reaction rate constant of "enzyme" synthesis of the i^{th} mode is considered by $k_{E,i}$. The term e_i^{rel} is the relative "enzyme" concentration with respect to its maximum value $e_i^{\max}$ calculated at steady state. In the last, $\mu_i^{\max}$ is the maximum growth rate of the i^{th} EM (Franz *et al.*, 2011; Song *et al.*, 2009). The vector e_i represents the concentration of the cybernetic "enzymes" whose dynamics are,

$$\frac{de_i}{dt} = \alpha_i + r_{EM,i} - (\beta_i + \mu)\, e_i \qquad \text{(III.30)}$$

The parameters α_i and β_i are the constitutive synthesis and degradation rates respectively. The term $\mu.e_i$ is the dilution rate of enzymes due to growth and $r_{EM,i}$ is the inducible synthesis rate of "enzymes" regulated by the cybernetic variable u_i as,

$$r_{EM,i} = u_i \, k_{E,i} \, b \, r_{M,i}^{kin} \tag{III.31}$$

where $r_{M,i}^{kin}$ is the kinetic term as defined in equations ((III.26) – (III.28)). and $b = 1 - m = 1 - LIP_A/X$ represents the fraction of biomass without accumulated metabolites, which in this case correspond to the fraction of catalytic biomass.

The formulation of cybernetic variables, v_i and u_i, is stated by the proportional and matching laws, respectively, as

$$v_i = \frac{p_i}{\|p_i\|_\infty}, \; v_i \leq 1 \tag{III.32}$$

$$u_i = \frac{p_i}{\|p_i\|_1}, \; \sum_i u_i = 1 \tag{III.33}$$

$$p_i = f_{C,i}.\, e_i^{rel}.\, r_M^{kin} \tag{III.34}$$

where p_i is the return on investment (ROI) of the metabolic objective function seeking the maximization of the carbon uptake units $f_{C,i}$ (Franz *et al*., 2011; Song *et al*., 2009).

III.3.2.6. <u>*Macroscopic Bioreaction Model (MBM)*</u>

The Macroscopic Bioreaction Model (MBM), developed by (Provost *et al*., 2006), is also based on the QSSA. The computation of EM is used to express the model by macroscopic reactions. Even if HCM and MBM are similar in structure and both are based on EMs, MBM does not account for cellular regulation. This is reflected in the kinetic expression of the uptake rates $r_{M,i}$, in which the adaptation term is equal to one ($\phi_i = 1$). This straightforward methodology can be also applied to model metabolic shifts through the addition of kinetic equations and the inclusion of ϕ_i smooth switching functions (*e.g.* sigmoidal functions). However, these functions are bounded by time *t* to favor the transition of phases. For the lipid accumulation cultures, another kinetic term was included into phase *P2* to consider that under high nitrogen concentrations, lipids were not accumulated as reported by (C. N. C. N. Economou et al., 2011). Therefore, equation (III.27) was rewritten as,

$$P2 \rightarrow r_{M,i}^{kin} = \left(\frac{GLC}{K_{GLC,2} + GLC}\right)\left(\frac{1}{1 + CIT/k_1}\right)\left(\frac{1}{1 + N/k_3}\right) \qquad (i = 1, \ldots, n_{P2}) \tag{III.35}$$

where k_3 is the limitation constant for nitrogen. The phases shifting factor ϕ_i was defined by step functions as,

$$\phi_i = \begin{cases} h_1, & 1-h_1, & 0 & t < t_{12} \\ 1-h_2, & h_2, & 0 & t_{12} \le t \le t_{23} \\ 0, & 1-h_3, & h_3 & t > t_{23} \end{cases} \quad \text{(III.36)}$$

The variables h_i are the normalized weights of each phase within the [0,1] set along the culture. The limits t_{12} and t_{23} correspond to the experimental starting times of phases *P2* and *P3* to address the transitions on the lipid accumulation cultures.

As seen in equation (III.36), phase *P2* is always active complementing the weights either of phase *P1* or *P3*. Despite of the low number of parameters to outline transitions in MBM, it requires a manual selection of switching times, and thus other assumptions are needed before further uses of the model.

III.3.2.7. Fuzzy MBM

Since MBM is time dependent, a Fuzzy set adaptation is proposed to avoid the use of the switching times t_{12} and t_{23}. A new approach called Fuzzy MBM is this here presented. As this approach is based on MBM, the normalized weights of each phase h_i remain within the range [0,1]. The membership functions of the fuzzy sets are defined by the behavior of the consumption ratio of nitrogen over carbon rN/rC (rNC for simplicity) (Cescut, 2009; Ochoa-Estopier, 2012; Ochoa-Estopier and Guillouet, 2014), which is a fundamental parameter to identify metabolic shifts. The membership functions (Zadeh, 1970) for each phase were determined by trapezoidal functions. In this case as in MBM, it is also assumed that phase *P2* is never completely deactivated, and thus it exists a threshold allowing the values to approach the limits of the [0,1] range. The corresponding mathematical representation of $\phi_i = \mu(rN/rC) = \mu(rNC)$, is expressed in the following equation,

$$\phi_i = \begin{cases} h_1, & 1-h_1, & 0 & \lambda < rNC \\ \left(h_1 - (1-h_2)\right)\xi_1 + h_1, & (1-h_1-h_2)\,\xi_1 + (1-h_1), & 0 & \sigma \le rNC \le \lambda \\ 1-h_2, & h_2, & 0 & \psi \le rNC \le \sigma \\ 0 & \left(h_2 - (1-h_3)\right)\xi_2 + h_2, & -h_3.\xi_2 & \omega \le rNC \le \psi \\ 0 & 1-h_3 & h_3 & rNC \le \omega \end{cases} \quad \text{(III.37)}$$

where $\xi_1 = \left(\frac{rNC-\lambda}{\lambda-\sigma}\right)$ and $\xi_2 = \left(\frac{rNC-\psi}{\psi-\omega}\right)$. The variables h_i are normalized weights, as well as in MBM. The parameters λ, σ, ψ, and ω correspond to the limits of each phase and its transitions *P1*, *P1 – P2*, *P2*, *P2 – P3*, and *P3* respectively.

Equation (III.37) describes the adaptation for the cultures with three phases. Nevertheless, for the cultures under nitrogen deficiency only two phases were identified. The lack of nitrogen input neglects the existence of phase *P2* in this experimental data. In this context, it is hypothesized that when nitrogen feeding is null $h_2 = 1 - h_1$. This allows maintaining the maximal value of *P2* at a low level, which is reflected in a longer *P1*.

III.4. Results and Discussion

III.4.1. Computation of Elementary modes

The computation of elementary modes with the metabolic network was performed in METATOOL (Kamp et al., 2006), implemented in MATLAB®, leading to1944 elementary modes (EMs).

Figure III.11 displays the distribution of the 1944 EMs, where the diagonal depicts the histograms relating the quantity of EMs versus the metabolite yield. The maximum yield value of citric acid was found to be 1, whilst the maximum value for CO_2 is 6, which is consisted with the stoichiometry. The majority of EMs (more than 1000) produced neither citric acid nor lipids. Otherwise, the lower diagonal matrix shows yield plots of the metabolites on the axes with respect to consumed Glucose. The maximum yield values, for instance, could be identified for all the metabolites. In the case of the accumulated lipids, it attained 0.28 gLIP_A.gGLC^{-1}, whereas for catalytic biomass was 0.5 gX_{Cat}.gGLC^{-1}.

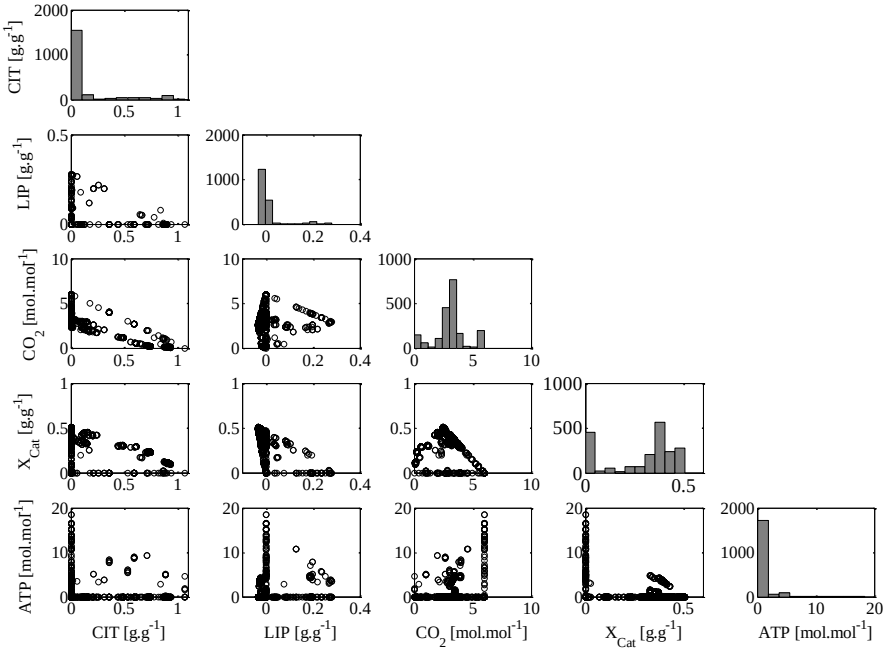


Figure III.11. Distribution of Elementary Modes on the reduced metabolic network. Histograms of elementary modes distribution (Diagonal). Yield plots (Lower diagonal).

Table III.4. Stoichiometry of active modes (AMs), expressed as yields with respect to glucose, which are necessary to reproduce experimental data corresponding to each phase. Units of yields are in given in (mmol.mmolGLC^{-1}) except for Biomass *X* given in (gX.mmolGLC^{-1}).

AMs	$Y_{CIT/GLC}$	$Y_{LIP/GLC}$	$Y_{CO2/GLC}$	$Y_{N/GLC}$	$Y_{X/GLC}$	*Phase*
1	0	0	2.8680	-0.5465	0.0738	*P1*
2	0	0	2.5797	-0.5968	0.0805	*P1*
3	0	0.0051	3.1240	-0.4526	0.0611	*P2*
4	0	0.0537	2.8576	-0.0252	0.0034	*P2*
5	0	0.0167	2.3679	-0.4711	0.0636	*P2*
6	0.7908	0.0153	0.4031	0	0	*P3*
7	1	0	0	0	0	*P3*
8	0.0375	0	5.7748	0	0	*P3*

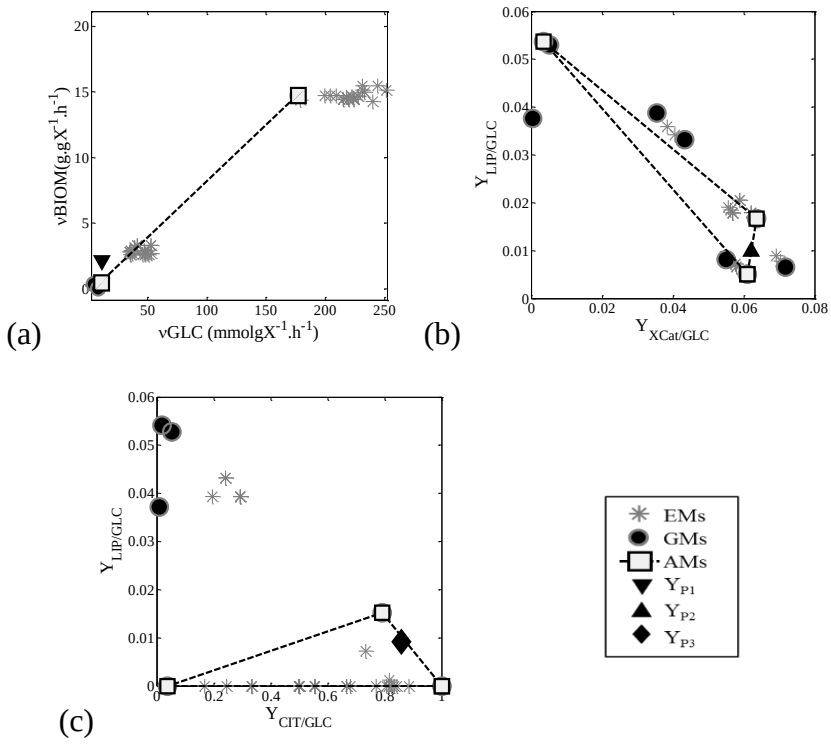


Figure III.12. Selection of active modes. (a) Flux analysis for phase *P1*. (b) Yield analysis for phase *P2*. (c) Yield analysis for phase *P3*. Generating Modes (GMs) represent the elementary modes in the convex hull. AMs are the selected elementary modes. Y_{P1} is the experimental specific flux value for phase *P1*. Y_{P2} and Y_{P3} represent the experimental yield values of phase *P2* and *P3*, respectively.

The data from Exp.1 (Cescut, 2009) were used to calculate the experimental yields and to select the active modes (AMs). The cutting into phases led to 175 EMs to describe phase *P1*, 69 for phase *P2*, and 105 EMs for phase *P3*. Flux and yield analysis were implemented taking into account that metabolites were at quasi-steady state in each phase, and that their principal products (Figure III.12). For phase *P1*, yield analysis was not feasible due to the existence of only two external metabolites that could be used (Biomass and Glucose); therefore, flux analysis was implemented. With regards on phases *P2* and *P3*, yield analysis was plausible, and thus carried out contemplating the lipid/glucose yield versus biomass/glucose yield for *P2* and the citrate/glucose yield versus lipid/glucose yield for *P3*. The results of these analyses are depicted in Table III.4, where two AMs were chosen to represent phase *P1*, and three AMs for each phase *P2* and *P3*; consequently n_r=8.

III.4.2. Parameter estimation

Two data sets were used to calibrate the three models (equations (III.18) – (III.28)): HCM (equations (III.29) – (III.34)), MBM (equation (III.36)(which replaces equation 17) and (III.36)), and Fuzzy MBM (equation (III.36)(which replaces equation 17) and (III.37)). The data obtained from the Fed-batch culture of *Yarrowia lipolytica* under nitrogen limitation (Exp.1) and data from the sequential batch (Exp.2) (Nitrogen deficiency t = 0 – 38 h) were used for parameter estimation. The limits for MBM in equation (III.36) were set as t_{12} = 3.6 h and t_{23} = 16.43 h for Exp.1 and t_{12} = t_{23} = 19.64 h for Exp.2.

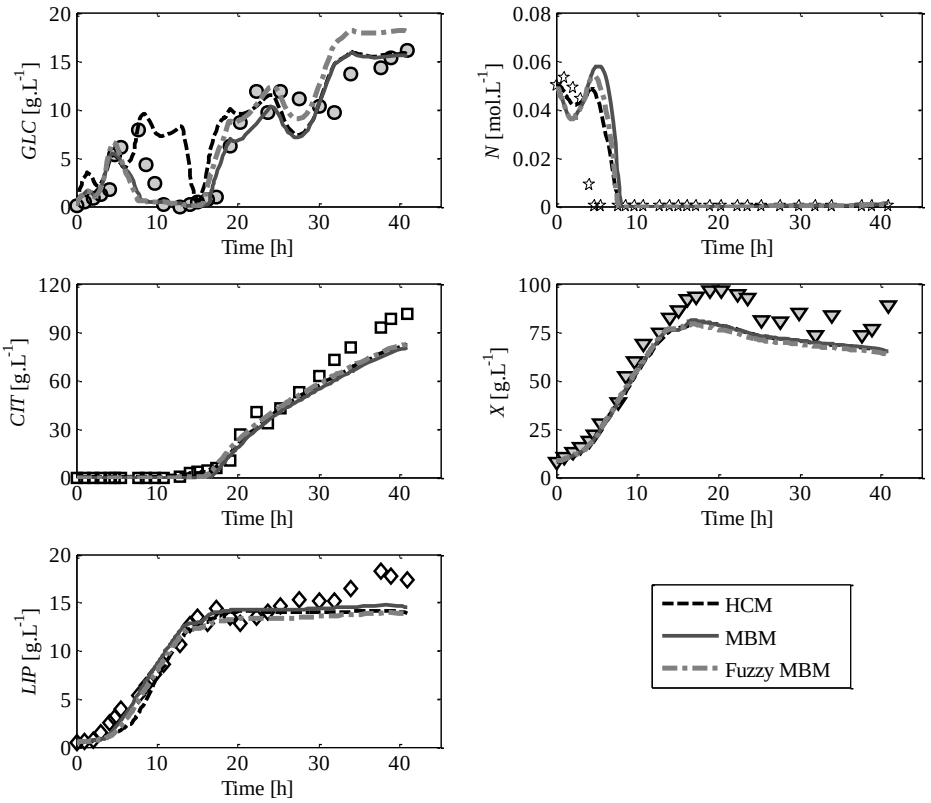


Figure III.13. Performance of the prediction of the three models with the data set from Exp.1 (Cescut, 2009)**. This set was used for calibration of the parameters. The symbols represent the concentrations of Residual Glucose (GLC), Residual Nitrogen (N), Citric acid (CIT), Total Biomass (X), and Total lipids (LIP).**

Parameter calibration was carried out by Particle Swarm Optimization (PSO) algorithm, which was further refined by the Pattern Search algorithm, both implemented in MATLAB®. The results of the calibration are displayed in Figure III.13 and Figure III.14. In a general overview, the three models were capable to properly represent Exp.1 and Exp.2. MBM and Fuzzy MBM exhibited similar behaviors with an overlap on all the state variables (Figure III.13 – Figure III.14). This fact proved that the fuzzy sets were well defined and identified to adapt the MBM model.

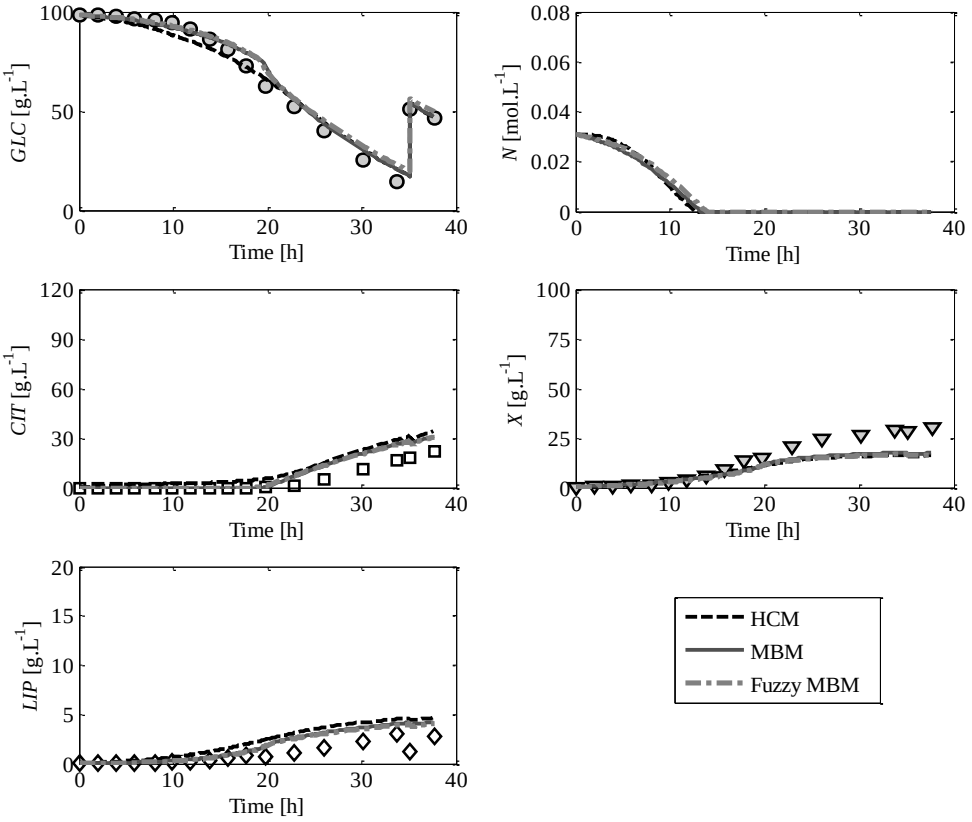


Figure III.14. Performance of the prediction of the three models with the data set from Exp.2 (*t* = 0 – 38*h*). The symbols represent the concentrations of Residual Glucose (GLC), Residual Nitrogen (N), Citric acid (CIT), Total Biomass (X), and Total lipids (LIP).

Table III.5. Estimated parameters for the models. The distribution of parameters is aligned to their correspondent phase of action.

Model	Parameters	Values							
		P1		*P2*			***P3***		
	Active Modes	***1***	***2***	*3*	*4*	*5*	***6***	***7***	***8***
All	$K_{GLC,i}$ [mmol*GLC*.L^{-1}]	3.56			74.05			29.02	
	$K_{N,1}$ [mmol*N*.L^{-1}]	463.2							
	$k_{1\text{-}2}$ [mmol.L^{-1}]	-			3.28			173.1	
HCM	k_i^{max} [mmol*GLC*.g*X*$^{-1}$.h^{-1}]	44.77	5.76	15.46	12.42	8.76	3.36	4.06	3.86
	$k_{E,i}$ [mmol*GLC*.g*X*$^{-1}$.h^{-1}]	4.26	10.91	89.17	44.68	61.99	0.45	11.79	4.26
	α_i* [h^{-1}]	0.1* $k_{E,i}$			0.1* $k_{E,i}$			0.1* $k_{E,i}$	
	β_i* [h^{-1}]	0.2			0.2			0.2	
MBM	k_i^{max} [mmol*GLC*.g*X*$^{-1}$.h^{-1}]	32.93	5.29	55.61	5.03	5.42	0.77	0.49	0.46
	k_3 [mmol.L^{-1}]	-			15.81			-	
	h_i	0.971			0.928			0.902	
	t_{12} [h]	3.6[a] 19.64[b] 18[c]							
	t_{23} [h]					16.43[a]			
Fuzzy MBM	k_i^{max} [mmol*GLC*.g*X*$^{-1}$.h^{-1}]	32.93	5.29	55.61	5.03	5.42	0.77	0.49	0.46
	k_3 [mmol.L^{-1}]	-			15.81			-	
	h_i	0.969			0.928			0.935	
	λ	0.1055							
	σ			0.0814					
	ψ					0.07701			
	ω							0.0747	

*Obtained from (Song *et al.*, 2009). [a]Exp.1, [b]Exp.2, [c]Exp.3

All holds for the parameters of the common kinetic structure of the three models (Equations (III.26) - (III.28)).

The values of the estimated parameters for HCM, MBM, and the Fuzzy MBM are presented in Table III.5. The values of $K_{GLC,i}$, $K_{N,i}$, $k_{1\text{-}2}$ were considered to be the same for the three models. The value of k_3, for nitrogen limitation, was estimated by MBM and also used for Fuzzy MBM. The specific rates ${k_i}^{max}$, and the weights h_i were the same for the Fuzzy MBM and MBM since the first was built on MBM. Notwithstanding, ${k_i}^{max}$ were estimated independently for HCM because this model assumes cellular regulation. This means that the shifting term ϕ_i is affected by the terms $v_i\ {e_i}^{rel}\ {k_i}^{max}$, which modulates the fluxes individually through each AM (Song *et al.*, 2009). Further discussion regarding those differences is presented in section III.4.4.

In Table III.5, it is observed that the weights h_i never achieved the maximum value ($h_i = 1$), which indicates that phase *P2* was never deactivated. Fuzzy MBM identified the critical value of the consumption ratio of nitrogen over carbon (*rN/rC*), at which citric acid production started, at (ψ) 0.07701 mol*N*.Cmol^{-1}. This value is three times higher than the previously reported of 0.019 – 0.022 mol*N*.Cmol^{-1} (Ochoa-Estopier, 2012; Ochoa-Estopier and Guillouet, 2014).

Some statistical analysis based on different information criteria were performed in order to determine the goodness of fit of the three dynamic metabolic models and to compare their overall performance. These results are shown in Table III.6 and Table III.7for the calibration experiments Exp.1, and Exp.2, respectively. These criteria were the determination coefficient R^2, the Root Mean Square Errors (*RMSE*), and the Akaike Information Criterion (*AIC*) (Akaike, 1974). The *AIC* criterion is the most common in model comparison because it discloses the probability of a model to reproduce the lack of information taking into account the number of parameters, the number of data, and the residuals. *AIC* is thus better to find a balance between model complexity, and data fitting, preventing over-fitting (Burnham and Anderson, 2004). For instance, the best model is the one explaining the experimental data with the minimum number of parameters. The *AIC* is thus calculated as,

$$AIC = 2\,p + n\,(\ln(2\pi) + \ln(SSE) - \ln(n) + 1) \tag{III.38}$$

where n is the number of data, p represents the number of parameters, and *SSE* is the Sum of Square Errors. In this case, the corrected *AIC* (*AICc*) (Hurvich and Tsai, 1995) was considered to manage the trade-off between the number of parameters and the number of ($n/p<30$).

$$AICc = AIC + \frac{2.p(p+1)}{n-p-1} \tag{III.39}$$

The model with the lowest value of *AICc* is chosen as the best for reproducing the data with an appropriate trade-off between fitness, complexity, and number of parameters.

Table III.6. Statistical evaluation of the dynamical metabolic models with Exp.1.

Model	**Criterion**	**Nitrogen Limitation**				
		Glucose	Nitrogen	Citric Acid	Biomass	Lipids
HCM	R^2	0.713	0.559	0.991	0.979	0.952
	RMSE	3.47	0.014	7.12	11.02	1.60
	AICc	395.57	86.19	435.79	460.26	352.30
MBM	R^2	0.831	0.381	0.991	0.978	0.954
	RMSE	2.32	0.018	8.18	10.77	1.31
	AICc	286.55	13.30	357.17	372.55	254.65
Fuzzy MBM	R^2	0.826	0.433	0.986	0.970	0.957
	RMSE	2.84	0.016	6.92	11.91	1.65
	AICc	384.30	94.78	434.18	464.57	353.94

Table III.7. Statistical evaluation of the dynamical metabolic models with Exp.2 (*t* = 0 – 38*h*).

Model	**Criterion**	**Nitrogen Deficiency**				
		Glucose	Nitrogen	Citric Acid	Biomass	Lipids
HCM	R^2	0.897		0.965	0.973	0.830
	RMSE	9.06		7.21	6.87	1.48
	AICc	-1.48		-9.25	-10.89	-63.15
MBM	R^2	0.895		0.968	0.989	0.844
	RMSE	9.05	*NA*	5.38	6.30	1.10
	AICc	-92.16		-109.89	-104.52	-163.81
Fuzzy *MBM*	R^2	0.899		0.963	0.986	0.840
	RMSE	8.73		5.28	6.78	1.03
	AICc	-2.70		-19.86	-11.34	-75.34

NA: Not Available

The obtained R^2 and *RSME* values for the three models were similar reflecting that all could predict the data with the same accuracy for Exp.1. However, as mentioned before, the most meaningful criterion is the *AICc*, for which MBM reported the lowest values. Nevertheless, MBM cannot be used for comparison because phase shifting is performed manually, and this model cannot be applied for optimization purposes. In this regard, the evaluation of *AICc* criterion for Exp.1 demonstrated comparable values for citric acid, biomass, and lipid

concentrations. Fuzzy MBM better estimated glucose concentration, whilst HCM had the lowest values for nitrogen concentration. This states that both models are able to reproduce the data at the same level of parsimony, where the similar *AICc* values were due to the equal number of parameters (22 for Fuzzy MBM, and 22 for HCM). Furthermore, these conclusions are in agreement with the R^2 and *RSME* values.

For experiment Exp.2, the same criteria (*AICc*, R^2, and *RSME*) were evaluated for each state variable with the exception of nitrogen concentration, where data were not available (Table III.7). In this case, Fuzzy MBM was better than HCM according to *AICc*, with comparable values for glucose, and biomass concentrations.

III.4.3. Model validation

The data of the sequential batch cultures of *Yarrowia lipolytica* under nitrogen deficiency (see Section III.3.1.2) were used in order to validate the models. These data comprised the second half of Exp.2 (t = 38–80 h) and Exp.3. In MBM, the limits for Exp.3 corresponded to $t_{12} = t_{23}$ =18 h.

The results for Exp.2 are displayed in Figure III.15, whereas the prediction results for Exp.3 are depicted in Figure III.16. The second half of Figure III.15 shows the validation of the three models, where all predicted similar concentrations for the plotted variables.

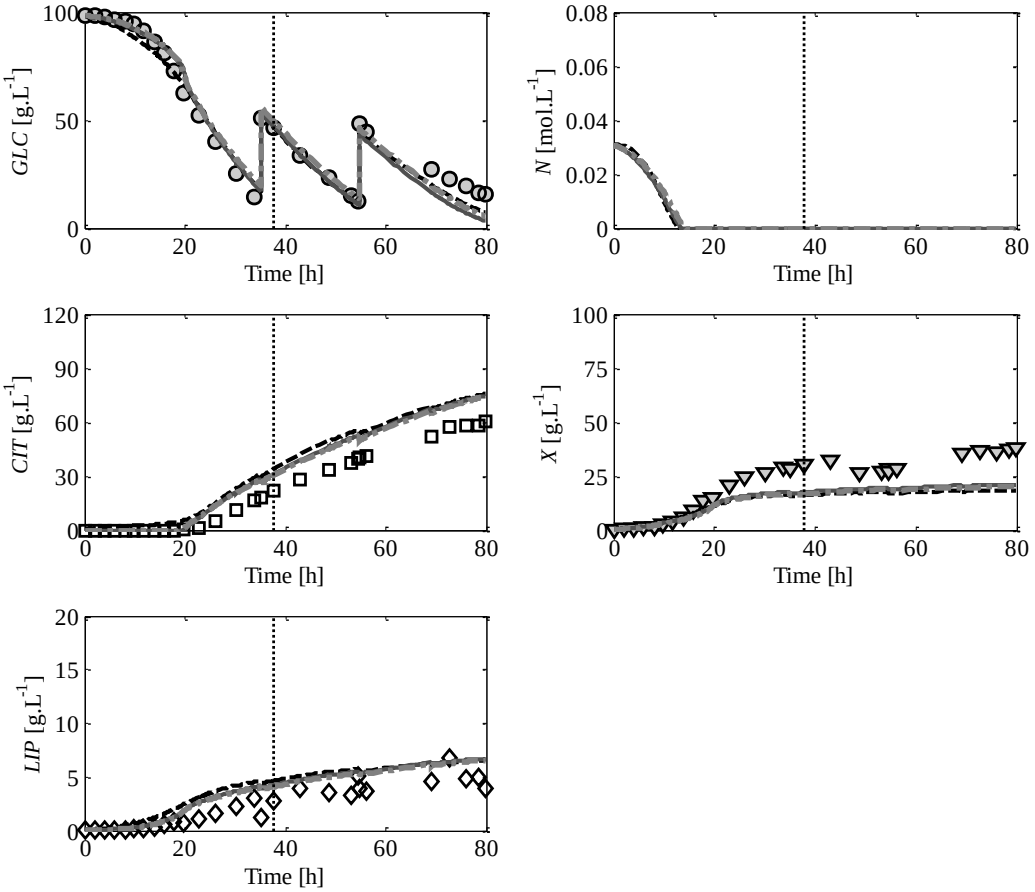


Figure III.15. Performance of the prediction of the three models with the data set from Exp.2. The line divides two subsets: the first (*t* = 0 – 38*h*) was used for model calibration; the second (*t* = 38 – 80*h*) for validation. The symbols represent the concentrations of Residual Glucose (GLC), Residual Nitrogen (N), Citric acid (CIT), Total Biomass (X), and Total lipids (LIP).

Figure III.16 shows an independent data set for validation, where the concentrations of lipids and glucose present, in general, an acceptable fitting. An overfitting on citric acid concentration was observed with the three models, which resulted in a lower estimation of biomass. This overfitting was also depicted on the results from Exp.2. This fact could be explained by the important trade-off between the Exp.1 and Exp.2 in calibration.

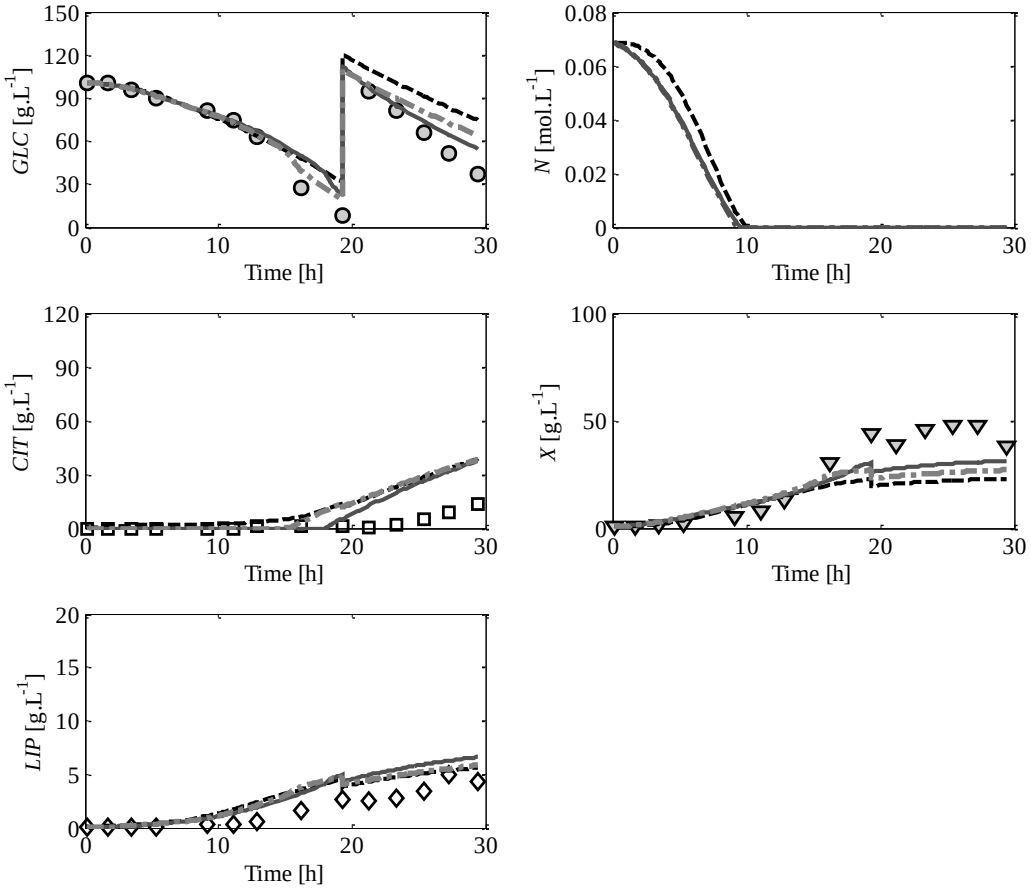


Figure III.16. Performance of the prediction of the three models with the data set from Exp.3. This set was used for model validation. The line divides the two subsets. The symbols represent the concentrations of Residual Glucose (GLC), Residual Nitrogen (N), Citric acid (CIT), Total Biomass (X), and Total lipids (LIP).

Statistical evaluation was performed in order to disclose the differences of fitting on the validation, and to compare the models. The same statistical criteria (*AICc*, R^2, and *RSME)* than in calibration were calculated for each state variable with the exception of nitrogen concentration where data were not available. The statistical evaluation of validation on Exp.2 and Exp.3 is reported on Table III.8 and Table III.9, respectively. As it is observed, the values of R^2, and *RMSE* for MBM and Fuzzy MBM were similar, which confirms the good definition of the Fuzzy rules for nitrogen deficiency and nitrogen limitation. However, it is worth to remember that MBM is excluded from comparison due to the manual setting of shifting times.

Table III.8. Statistical evaluation of the dynamical metabolic models with Exp.2 (*t* = 38 – 80*h*).

Model	**Criterion**	**Nitrogen Deficiency**				
		Glucose	Nitrogen	Citric Acid	Biomass	Lipids
HCM	R^2	0.503		0.989	0.415	0.397
	RMSE	11.80		15.33	14.90	1.76
	AICc	45.28		51.57	50.88	-0.34
MBM	R^2	0.506		0.992	0.467	0.404
	RMSE	13.26	*NA*	13.15	12.80	1.66
	AICc	39.09		38.88	38.24	-10.74
Fuzzy *MBM*	R^2	0.506		0.993	0.484	0.399
	RMSE	11.75		12.37	13.54	1.53
	AICc	45.19		46.42	48.59	-3.74

NA: Not Available

Table III.9. Statistical evaluation of the dynamical metabolic models with Exp.3.

Model	**Criterion**	**Nitrogen Deficiency**				
		Glucose	Nitrogen	Citric Acid	Biomass	Lipids
HCM	R^2	0.778		0.779	0.879	0.896
	RMSE	18.33		13.99	14.74	1.33
	AICc	52.72		45.15	46.61	-20.65
MBM	R^2	0.973		0.823	0.950	0.939
	RMSE	9.64	*NA*	12.20	9.84	1.64
	AICc	14.50		21.09	15.07	-35.08
Fuzzy *MBM*	R^2	0.921		0.741	0.905	0.905
	RMSE	11.30		14.06	12.14	1.39
	AICc	39.19		45.29	41.20	-19.30

NA: Not Available

The statistical analysis for Exp.2 reported similar *AICc* values for Fuzzy MBM and HCM, when comparing glucose concentrations. The *AICc* for biomass, lipids, and citric acid concentrations were smaller for Fuzzy MBM. Regarding Exp.3, Fuzzy MBM presented lower *AICc* values than HCM for glucose and biomass concentration, whereas comparable results between Fuzzy MBM and HCM were reported for lipids and citric acid concentration.

In general, the differences between *AICc* values remain under the allowed gap defined by Burhnham and Anderson (Burnham and Anderson, 2004). These authors stated that the difference between *AICc* values should be lower than 10. Therefore, supported by the statistical results of the three data sets, both Fuzzy MBM and HCM presented similar prediction capabilities, and both could be used for further applications, such as bioprocess control.

III.4.4. Comparison of Dynamic Metabolic Modeling approaches

Although the three dynamic metabolic models share a common structure in mass balances, they diverge on the manner of executing the shifting of phases, which has provided them with several advantages and disadvantages. In HCM, the uptake of the various elementary modes is combined using global cybernetic variables based on the maximization of substrate uptake kinetics. Additionally, HCM is the only one accounting with cellular regulation by the cybernetic "enzymes", which allows knowing the dynamic behavior of each active mode. However, these "enzymes" lack of biological sense and add to the model an important number of parameters to estimate. Otherwise, MBM is a straightforward approach to reproduce experimental data and to identify kinetics. It is also a useful model to identify biological parameters, such as the *rN/rC* critical ratio. Notwithstanding, MBM requires the inclusion of extra kinetic parameters, since the adaptation is based on kinetics. Furthermore, it is time dependent for the identification of metabolic shifts, and thus it is not a suitable candidate either for describing other experiments or for control objectives. On the other hand, Fuzzy MBM was constructed based on MBM, and therefore this approach can hold the same advantages with the addition of the independent switching capabilities.

A common advantage of these dynamic metabolic models is the possibility to estimate the intracellular fluxes of reactions through the metabolic network.

In this work, it is stated that the main difference in structure between these models rely on the shifting factor and the ϕ_i corresponding to equations (III.29) for HCM, (III.36) for MBM and (III.37) for Fuzzy MBM, and the addition of a kinetic term for the last two approaches. This distinction forced to estimate the maximum specific rates k_i^{max} independently for each model (Table 2). For the sake of illustration, the dynamic behavior of the shifting factor for the three models is depicted for the culture under nitrogen limitation (Exp.1), and one culture under

nitrogen deficiency (Exp.3) in Figure III.17 and Figure III.18, respectively. Since results from Exp.2 were similar to those from Exp.3, Exp.2 was not shown.

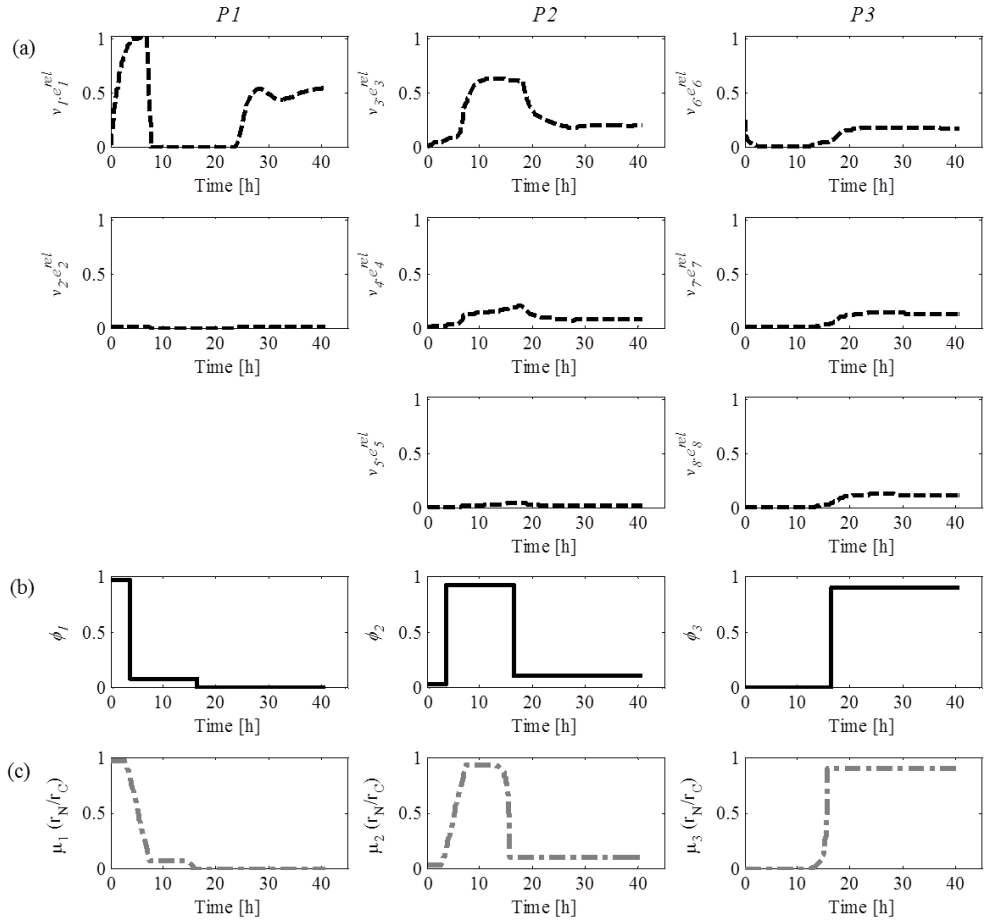


Figure III.17. Phase shifting ϕ allocation by the models with Exp.1. (a) and (--) HCM. (b) and (-) MBM. (c) and (-.-) Fuzzy MBM.

Figure III.17 and Figure III.18 depict that HCM assigns the ϕ_i value per AM, whereas MBM and Fuzzy MBM calculate ϕ_i by phase. Figure III.17 displays the culture under nitrogen limitation (Exp.1), in which the three phases are observed to be activated and deactivated. It is worth noting that for phase *P1* the active mode AM_1 has a more important activity, while phase *P2* is mostly regulated for the mode AM_3, and phase *P3* by the three modes describing this phase. Figure III.18 shows a culture under nitrogen deficiency (Exp.3). Although the differences on cultures of Exp.1 and Exp.3, HCM exhibited a similar behavior of the shifting term ϕ_i, whereas MBM and Fuzzy MBM considered a small value for phase *P2*. Figure III.17 (b-c) and Figure III.18 (b-c) show the shifting of MBM and Fuzzy MBM for experiments Exp.1

and Exp.3, respectively. In Figure III.18c, it is shown that Fuzzy MBM (equation (III.37)) is able to correctly identify the type of data according to the nitrogen inputs.

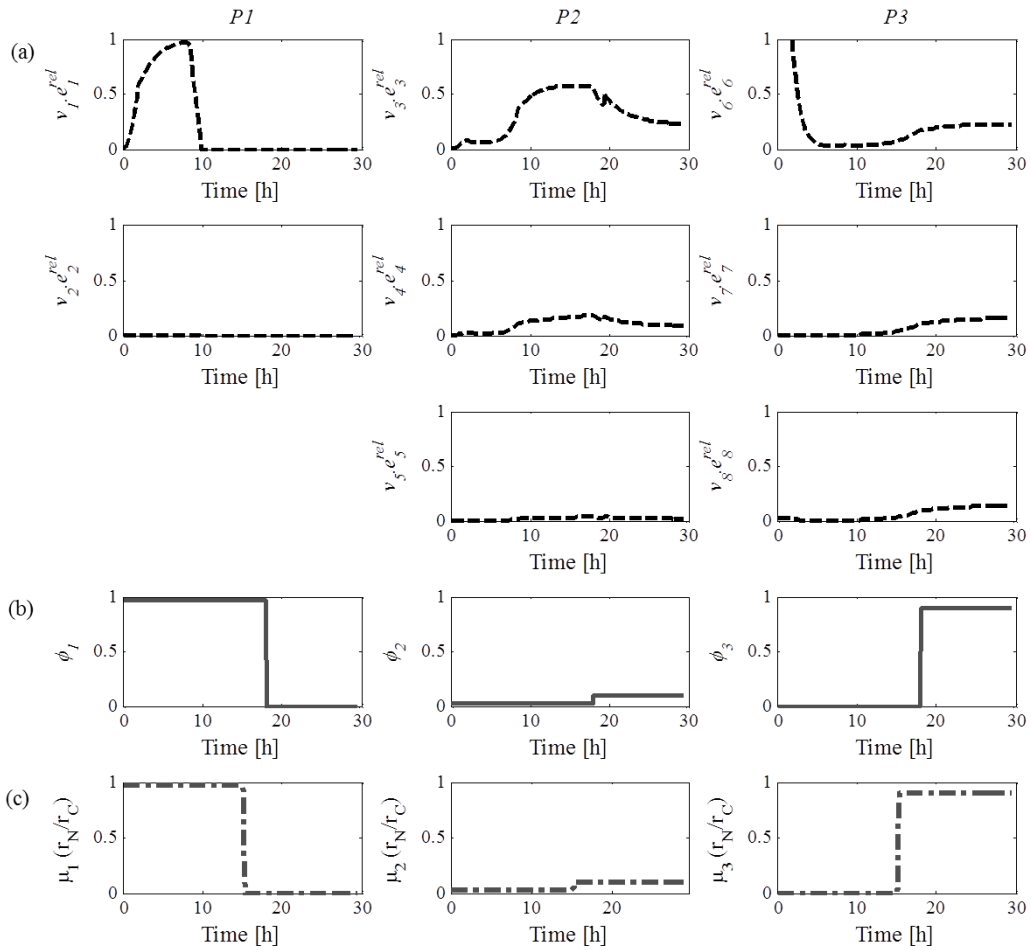


Figure III.18. Phase shifting ϕ allocation by the models with Exp.3. (a) and (--) HCM. (b) and (-) MBM. (c) and (-.-) Fuzzy MBM.

Summarizing, HCM and Fuzzy MBM were identified to have a good trade-off between complexity and knowledge. The main difference between HCM and Fuzzy MBM is that HCM manages the metabolic shifts with regulation provided by the cybernetic enzymes. These enzymes, however, lack of biological sense. Fuzzy MBM is based on kinetics and fuzzy rules defined on the *rN/rC* ratio, which can provide more information about the metabolic state of the culture. Nevertheless, HCM is stronger for applicability. This was observed when no change was needed to predict the different nature of the cultures (nitrogen limitation and nitrogen

deficiency). This change was defined as two separate cases for Fuzzy MBM. Although the applicability of HCM is better than Fuzzy MBM, special attention must be paid to the calculation of initial enzymes, which have a strong impact on the fitting of the models. It is therefore believed that both models HCM and Fuzzy MBM could be employed for further biotechnological applications targeting prediction, monitoring, control, and optimization of internal metabolites as lipids.

III.5. Conclusions

Three dynamic metabolic models were proposed in order to describe lipid accumulation by *Yarrowia lipolityca* growing on glucose. The three models HCM, MBM, and Fuzzy MBM were calibrated and validated with three independent experimental data sets concerning one fed-batch culture under nitrogen limitation, and two sequential batch cultures under nitrogen deficiency. One and a half data sets for each. Results of the three models showed acceptable fitting for the dynamics of the three data sets regardless of their intrinsic differences. Although MBM reported the best statistical values, this approach is time dependent for shifting phases. Neglecting MBM, both Fuzzy MBM and HCM were statistically similar. Automation of both models proved that they can be incorporated into control strategies to optimize lipid accumulation.

III.6. Supplementary Material

III.6.1. Reaction of the metabolic network of Yarrowia lipolytica.

Table III.10. Metabolites

AcACP	Acetyl Acyl Carrier Protein	*MAL*	Malate (cytosol)
AcCoA	Acetyl Coenzyme A (cytosol)	*MalACP*	Malonyl ACP
AcCoAm	Acetyl Coenzyme A (mitochondria)	*MALm*	Malate (mitochondria)
ADP	Adenosine Bisphosphate	*NAD*	Nicotinamide adenine dinucleotide oxidized (cytosol)
AKG	α - ketoglutarate	*NADH*	Nicotinamide adenine dinucleotide reduced (cytosol)
ATP	Adenosine Triphosphate	*NADHm*	Nicotinamide adenine dinucleotide reduced (mitochondria)
BIOM	Catalytic Biomass	*NADm*	Nicotinamide adenine dinucleotide oxidized (mitochondria)
CIT	Citric acid (cytosol)	*NADP*	Nicotinamide adenine dinucleotide phosphate oxidized (cytosol)
CITm	Citric acid (mitochondria)	*NADPH*	Nicotinamide adenine dinucleotide phosphate reduced (cytosol)
CITx	Citric acid (extracellular)	*NADPHm*	Nicotinamide adenine dinucleotide phosphate reduced (mitochondria)
CO2	Carbon dioxide (intracellular)	*NADPm*	Nicotinamide adenine dinucleotide phosphate oxidized (mitochondria)
CO2x	Carbon dioxide (extracellular)	*NH3*	Ammonia
DAG	Diacylglycerol	*NH4*	Ammonium
DHAP	Dihydroxyacetone phosphaste	*O2*	Oxygen
E4P	Erythrose 4 phosphate	*OAA*	Oxaloacetate
F6P	Fructose 6 phosphate	*oleic*	Oleic acid
FAD	Flavin adenine dinucleotide	*oleicACP*	Oleyl Acyl Carrier Protein
FADH2	Flavin adenine dinucleotide	*oleicCoA*	Oleyl Coenzyme A
FDP	Fructose diphosphate	*palm*	Palmitic acid
G6P	Glucose 6 phosphate	*palmACP*	Palmitoyl Acyl Carrier Protein
GAP	Glucose 3 phosphate	*palmCoA*	Palmitoyl Coenzyme A
GLC	Glucose (extracellular)	*PEP*	Phospho-enol pyruvate
GLUM	Glutamine	*PYR*	Pyruvate
GLUT	Glutamate	*R5P*	Ribose 5 phosphate
GOL	Glycerol 3 phosphate	*Ru5P*	Ribulose 5 phosphate
GOX	Glyoxilate	*S7P*	Sedoheptulose 7 phosphate
ICT	Isocitrate	*SUC*	Succinate
linoleic	Linoleic acid	*SUCCoA*	Succinate Coenzyme A
linoleicACP	Linoleyl Acyl Carrier Protein	*TAG*	Triacylglycerol
linoleicCoA	Linoleyl Coenzyme A	*X5P*	Xylose 5 phosphate
MAINT	Maintenance		

Table III.11. Reactions on the Reduced Metabolic Network of *Yarrowia lipolytica* holding different metabolic pathways. Irreversible reactions are described by '=>', whereas reversible reactions are denoted by '='. Biomass equation was modified from the one proposed by (Cakir *et al.*, 2007)

Glycolysis
R1 : GLC + ATP => G6P + ADP
R2 : G6P = F6P
R3 : F6P + ATP => FDP + ADP
R4 : FDP => F6P
R5 : FDP = DHAP + GAP
R6 : DHAP = GAP
R7 : GAP + NAD + ADP = PEP + NADH + ATP
R8 : PEP + ADP => PYR + ATP
R9 : DHAP + NADH => GOL + NAD
R10 : PYR + NADm => AcCoAm + NADHm + CO2
R11 : PYR + ATP + CO2 => OAA +ADP
Pentose Phosphate
R12 : G6P + 2 NADP => Ru5P + CO2 + 2 NADPH
R13 : Ru5P = X5P
R14 : Ru5P = R5P
R15 : R5P + X5P = S7P + GAP
R16 : X5P + E4P = F6P + GAP
R17 : S7P + GAP = F6P + E4P
Krebs cycle
R18 : OAA + AcCoAm => CITm
R19 : CITm = ICT
R20 : ICT + NADm => AKG + NADHm + CO2
R21 : ICT + NADPm => AKG + NADPHm + CO2
R22 : AKG + NADm => SUCCoA + NADHm + CO2
R23 : SUCCoA + ADP = SUC + ATP
R24 : SUC + FAD = MALm + FADH2
R25 : MALm + NADm = OAA + NADHm
R26 : NADH + NADm => NAD + NADHm
R27 : AcCoA => AcCoAm
Glyoxylate cycle
R28 : ICT => GOX + SUC
R29 : GOX + AcCoAm => MAL
Citrate – malate cycle
R30 : MAL + CITm = MALm + CIT
R31 : CIT + ATP => OAA + AcCoA
R32 : OAA + NADH = MAL + NAD
R33 : MAL + NADP => PYR + NADPH + CO2

Oxidative phosphorylation

R34 : 2 FADH2 + O2 + 2.4 ADP => 2.4 ATP + 2 FAD
R35 : 2 NADHm + O2 + 2.4 ADP => 2.4 ATP + 2 NADm

Maintenance

R36 : ATP => MAINT + ADP

Biomass formation

R37 : 0.07 DAG + 2.35 PYR + 5.64 GLUT + 1.40 OAA + 74 ATP + 1.15 GLUM
+ 1.11 PEP + 0.40 AcCoAm + 1.31 NADPHm + 0.40 R5P + 0.55 GOX
+ 0.07 GAP + 1.64 AcCoA + 0.36 E4P + 1.76 G6P + 5.43 NADPH + 1.75 NAD
+ 0.99 NADm =>
BIOM + 5.41 AKG + 1.75 NADH + 2.48 CO2 + 0.53 NH3 + 0.99 NADHm
+ 1.31 NADPm + 5.43 NADP + 74 ADP

Glutamine, glutamate metabolism

R38 : NH4 => NH3
R39 : NADPH + NH3 + AKG => GLUT + NADP
R40 : ATP + NH3 + GLUT => GLUM + ADP

Fatty acid synthesis

R41 : AcCoA => AcACP
R42 : AcCoA + ATP + CO2 = MalACP + ADP
R43 : AcACP + 14 NADPH + 7 MalACP => palmACP + 7 CO2 + 14 NADP
R44 : palmACP = palm
R45 : palm + ATP = palmCoA + ADP
R46 : palmACP + O2 + 3 NADPH + MalACP => oleicACP + CO2 + 3 NADP
R47 : oleicACP = oleic
R48 : oleic + ATP = oleicCoA + ADP
R49 : oleicACP + O2 + NADPH => linoleicACP + NADP
R50 : linoleicACP = linoleic
R51 : linoleic + ATP = linoleicCoA + ADP

Triacylglycerols

R52 : 0.58 oleicCoA + 0.38 palmCoA + 1.04 linoleicCoA + GOL => DAG
R53 : DAG + 0.29 oleicCoA + 0.19 palmCoA + 0.52 linoleicCoA => TAG
R54 : TAG => DAG + 0.29 oleic + 0.19 palm + 0.52 linoleic

β oxidation

R55 : palmCoA + 7 NAD + 7 FAD => 8 AcCoA + 7 NADH + 7 FADH2
R56 : linoleicCoA + 2 NADPH + 8 NAD + 8 FAD => 8 NADH + 9 AcCoA + 8 FADH2
+ 2 NADP
R57 : oleicCoA + 8 NAD + 8 FAD + NADPH => 9 AcCoA + 8 NADH + 8 FADH2
+ NADP

Excretion

R58 : CO2 => CO2x
R59 : CIT => CITx

III.7. References

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Trans. Automat. Contr. 19, 716–723.

Anastassiadis, S., Aivasidis, A., Wandrey, C., 2002. Citric acid production by Candida strains under intracellular nitrogen limitation. Appl. Microbiol. Biotechnol. 60, 81–7. doi:10.1007/s00253-002-1098-1

Arzumanov, T.T.E., Sidorov, I.I.A., Shishkanova, N.V.N., Finogenova, T.T. V., 2000. Mathematical modeling of citric acid production by repeated batch culture. Enzyme Microb. Technol. 26, 826–833. doi:http://dx.doi.org/10.1016/S0141-0229(00)00178-2

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol. Methods Res. 33, 261–304. doi:10.1177/0049124104268644

Cakir, T., Kirdar, B., Onsan, Z.I., Ulgen, K.O., Nielsen, J., 2007. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in Saccharomyces cerevisiae. BMC Syst. Biol. 1, 18. doi:10.1186/1752-0509-1-18

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Economou, C.N.C.N., Aggelis, G., Pavlou, S., Vayenas, D.V. V., 2011. Modeling of single-cell oil production under nitrogen-limited and substrate inhibition conditions. Biotechnol. Bioeng. 108, 1049–55. doi:10.1002/bit.23026

Franz, A., Song, H.-S., Ramkrishna, D., Kienle, A., 2011. Experimental and theoretical analysis of poly(β-hydroxybutyrate) formation and consumption in Ralstonia eutropha. Biochem. Eng. J. 55, 49–58. doi:10.1016/j.bej.2011.03.006

Hurvich, C.M., Tsai, C.L., 1995. Model selection for extended quasi-likelihood models in small samples. Biometrics 51, 1077–84.

Kamp, A. v., Schuster, S., von Kamp, A., Schuster, S., 2006. Metatool 5.0: Fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931. doi:10.1093/bioinformatics/btl267

Kim, J. Il, Varner, J.D., Ramkrishna, D., 2008. A Hybrid Model of Anaerobic E. coli GJT001:

Combination of Elementary Flux Modes and Cybernetic Variables. Biotechnol. Prog. 24, 993–1006. doi:10.1021/bp.73

Klamt, S., Saez-Rodriguez, J., Gilles, E.D., 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2. doi:10.1186/1752-0509-1-2

Meeuwse, P., Akbari, P., Tramper, J., Rinzema, A., 2012. Modeling growth, lipid accumulation and lipid turnover in submerged batch cultures of Umbelopsis isabellina. Bioprocess Biosyst. Eng. 35, 591–603. doi:10.1007/s00449-011-0632-x

Meng, X., Xian, M., Xu, X., Zhang, L., Nie, Q., Yang, J., Xu, X., Zhang, L., Nie, Q., Xian, M., 2009. Biodiesel production from oleaginous microorganisms. Renew. Energy 34, 1–5. doi:10.1016/j.renene.2008.04.014

Nicaud, J.M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.

Ochoa-Estopier, A., 2012. Analyses systématique des bascules métaboliques chez les levures d'intérêt industriel: application aux bascules du métabolisme lipidique chez Yarrowia lipolytica. Institut National des Sciences Appliquées de Toulouse.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Papanikolaou, S., Aggelis, G., 2010. Yarrowia lipolytica: A model microorganism used for the production of tailor-made lipids. Eur. J. Lipid Sci. Technol. 112, 639–654. doi:10.1002/ejlt.200900197

Papanikolaou, S., Aggelis, G., 2003a. Modeling lipid accumulation and degradation in Yarrowia lipolytica cultivated on industrial fats. Curr. Microbiol. 46, 398–402. doi:10.1007/s00284-002-3907-2

Papanikolaou, S., Aggelis, G., 2003b. Modelling aspects of the biotechnological valorization of raw glycerol: production of citric acid by Yarrowia lipolytica and 1,3-propanediol by Clostridium butyricum. J. Chem. Technol. Biotechnol. 78, 542–547. doi:10.1002/jctb.831

Papanikolaou, S., Aggelis, G., 2002. Lipid production by Yarrowia lipolytica growing on industrial glycerol in a single-stage continuous culture. Bioresour. Technol. 82, 43–49.

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Provost, A., Bastin, G., Agathos, S.N., Schneider, Y.-J.-J., 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. Bioprocess

Biosyst. Eng. 29, 349–66. doi:10.1007/s00449-006-0083-y

Ramkrishna, D., 2003. On modeling of bioreactors for control. J. Process Control 13, 581–589.

Ramkrishna, D., Song, H.-S.S., 2012. Dynamic Models of Metabolism : Review of the Cybernetic Approach. AIChE J. 58, 986–997. doi:10.1002/aic

Robles-Rodriguez, C.E., Bideaux, C., Gaucel, S., Larouche, B., Aceves-Lara, C.A., 2014. Reduction of metabolic models by polygons optimization method applied to Bioethanol production with co-substrates, in: IFAC 2014.

Schuster, S., Hilgetag, C., Woods, J.H.H., Fell, D.A.A., 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol. 45, 153–81. doi:10.1007/s002850200143

Song, H.-S., Morgan, J.A., Ramkrishna, D., 2009. Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol. Bioeng. 103, 984–1002. doi:10.1002/bit.22332

Song, H.-S., Ramkrishna, D., 2009a. When is the Quasi-Steady-State Approximation Admissible in Metabolic Modeling? When Admissible, What Models are Desirable? Ind. Eng. Chem. Res. 48, 7976–7985. doi:10.1021/ie900075f

Song, H.-S., Ramkrishna, D., 2009b. Reduction of a set of elementary modes using yield analysis. Biotechnol. Bioeng. 102, 554–568. doi:10.1002/bit.22062

Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24, 2229–35. doi:10.1093/bioinformatics/btn401

Ykema, A., Verbree, E.C., Van Verseveld, H.W., Smit, H., 1986. Mathematical modelling of lipid production by oleaginous yeasts in continuous cultures. Antonie Van Leeuwenhoek 52, 491–506.

Zadeh, L.A., 1970. Fuzzy algorithms. Inf. Control 17, 326–339. doi:10.1016/S0019-9958(70)80032-8

IV. MODELING AND OPTIMIZATION OF LIPID ACCUMULATION DURING FED-BARCH FERMENTATION BY *YARROWIA LIPOLYTICA* FROM GLUCOSE UNDER NITROGEN DEPLETION CONDITIONS

Carlos Eduardo Robles-Rodríguez[1+], Rafael Muñoz-Tamayo[1], Carine Bideaux[1], Nathalie Gorret[1], Stéphane E. Guillouet[1], Carole Molina-Jouve[1], Gilles Roux[2], César Arturo Aceves-Lara[1*]

[1]LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

[2] LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France

Submitted to Bioengineering and Biotechnology

Abstract

Oleaginous yeasts have been seen as a feasible alternative to produce the precursors of biodiesel due to their capacity to accumulate lipids as triacylglycerol having profiles with high content of unsaturated fatty acids. The yeast *Yarrowia lipolytica* is a promising microorganism that can produce lipids under nitrogen depletion conditions and excess of the carbon source. However, under these conditions, this yeast also produces citric acid (overflow metabolism) decreasing lipid productivity. This work presents two mathematical models for lipid production by *Yarrowia lipolytica* from glucose. The first model is based on Monod and inhibition kinetics, and the second one is based on the Droop quota model approach, which is extended to yeast. The two models showed good agreements with the experimental data used for calibration and validation. The quota based model presented a better description of the dynamics of nitrogen and glucose dynamics leading to a good managing of N/C ratio, which makes this model interesting for control purposes. Then, quota model was used to evaluate, by means of simulation, a scenario for optimizing lipid productivity and lipid content. For that, a control strategy was designed by approximating the flow rates of glucose and nitrogen with piecewise linear functions. Simulation results achieved a productivity of 0.95 $g.L^{-1}.h^{-1}$ and a lipid content fraction of 0.23 $g.g^{-1}$, which indicate a promising alternative for the optimization of lipids production.

Keywords: Biodiesel; Droop Model; Nitrogen quota; Unstructured modeling; *Yarrowia lipolytica*.

IV.1. Nomenclature

CIT	Citric acid concentration [g.L^{-1}]
$f_{N_{IN}}$	Input flow rate of nitrogen [L.h^{-1}]
$f_{N_{IN\max}}$	Maximum input flow rate of nitrogen [L.h^{-1}]
$f_{S_{IN}}$	Input flow rate of glucose [L.h^{-1}]
$f_{S_{IN\max}}$	Maximum input flow rate of glucose [L.h^{-1}]
F_{IN}	Input flow rate [L.h^{-1}]
F_{Sample}	Output flow rate[L.h^{-1}]
GUR_X	Glucose uptake rate [gS.gX_f^{-1}h^{-1}]
Ind_Q	Overflow indicator in Quota model [-]
Ind_U	Overflow indicator in Unstructured model [-]
$K_{S,i}$	Saturation constants of glucose [gS.L^{-1}]
K_N	Saturation constant of N [gN.L^{-1}]
k_i	Inhibition constants [g.L^{-1}]
n	Number of data [-]
n_f	Number of Piecewise linear functions
N	Nitrogen concentration [g.L^{-1}]
$N_{\max}$	Maximum residual nitrogen concentration [g.L^{-1}]
N_{IN}	Nitrogen input concentration [g.L^{-1}]
Of_U	Overflow saturation constant [gS.gX_f^{-1}h^{-1}]
P_L	Lipid productivity [g.(L-h)$^{-1}$]
P_W	Normalized lipid fraction [g.g^{-1}]
Q_0	Minimal quota for growth [gN.gX_T^{-1}]
q_N	Internal nitrogen quota [gN.gX_T^{-1}]
q_N*	Critical quota for citric acid formation [mol$N$.Cmol^{-1}]
S	Glucose concentration [g.L^{-1}]
$S_{\max}$	Maximum residual glucose concentration [g.L^{-1}]
t_f	Final optimal time [h]
$t_{\min}$	Minimum allowed time for the optimization [h]

t_{max}	Maximum allowed time for the optimization [h]
S_{IN}	Glucose input concentration [g.L^{-1}]
V	Volume [L]
V_0	Initial volume [L]
V_{max}	Maximal volume [L]
w_L	Lipid content fraction [g.g^{-1}]
$w_{L,max}$	Maximum theoretical value of lipid content fraction [g.g^{-1}]
w_P	weight coefficient to favor a high lipid fraction [-]
X_T	Total biomass concentration [g.L^{-1}]
X_f	Functional biomass concentration [g.L^{-1}]
X_L	Total Lipid biomass concentration [g.L^{-1}]
X_{LA}	Accumulated lipid biomass concentration (TAGs) [g.L^{-1}]
X_{LS}	Structural lipid biomass concentration [g.L^{-1}]
$Y_{X/S}$	Yield coefficient for functional biomass production with respect to glucose [gX_f.gS^{-1}]
$Y_{LIP/S}$	Yield coefficient for lipid production with respect to glucose [g*LIP*.gS^{-1}]
$Y_{CIT/S}$	Yield coefficient for citric acid production with respect to glucose [g*CIT*.gS^{-1}]
$Y_{X/N}$	Yield coefficient for functional biomass production with respect to nitrogen [gX_f.gN^{-1}]

Abbreviations

PSO	Particle Swarm Optimization
PWL	Piecewise linear functions
R^2	Correlation Coefficient
RMSE	Root Mean Square Error
TAG	Triacylglycerol

Greek symbols

α	maximum lipid content [g*LIP*.gX_T^{-1}]
γ	constant fraction of the functional biomass [gX_L.gX_f^{-1}]
μ	Growth rate [h^{-1}]
μ^{max}	Maximal growth rate on carbon and nitrogen [h^{-1}]
μ^*	Theoretical maximal growth rate on carbon and nitrogen [h^{-1}]

π_{LIP}	Specific rate of lipid accumulation [g*LIP*. gX_f^{-1}h^{-1}]
π_{LIP}^{max}	Maximal specific rate of lipid accumulation [g*LIP*.gX_f^{-1}h^{-1}]
π_{CIT}	Specific rate of citric acid production [g*CIT*.(gX_f-h)$^{-1}$]
π_{CIT}^{max}	Maximal specific rate of citric acid production [g*CIT*. gX_f^{-1}h^{-1}]
ρ_N	Specific uptake rate of nitrogen [g*N*. gX_f^{-1}h^{-1}]
ρ_N^{max}	Maximal specific uptake rate of nitrogen [g*N*. gX_f^{-1}h^{-1}]

Superscripts

max	Maximum

Subscripts

i	Identification of number of constants
max	Maximum

Operators

$\frac{d}{dt}$	Derivative of a variable with respect to time
max	Maximization

IV.2. Introduction

The use of microbial lipids as precursors of biodiesel production has outraised in the last years as an attractive and sustainable alternative (Papanikolaou and Aggelis, 2009). These lipids are produced by oleaginous microorganisms up to 20% of their dry weight (Ratledge, 2004). An interesting oleaginous microorganism is the yeast *Yarrowia lipolytica*. This yeast is able to use several substrates such as: glucose, fructose, lactose, sucrose, whey, glycerol, fatty acids, molasses, glucose/glycerol, glucose/stearin and glycerol/stearin (Nicaud et al., 2002; Papanikolaou and Aggelis, 2011a, 2003a, 2002; Rakicka et al., 2015). *Yarrowia lipolytica* has been identified as a potential candidate for the synthesis of biodiesel due to its production of high quantity of triacylglycerols (TAG) (>90% with respect to its total accumulation) with a rich fatty acid profile in linoleic acid (C18:2) (56%), oleic acid (C18:1) (28%), and palmitic acid (C16:0) (11%) (Beopoulos et al., 2009a; Papanikolaou and Aggelis, 2011b). Furthermore, this yeast is the only oleaginous yeast with a complete sequenced genome (Barth and Gaillardin, 1997), making it a suitable candidate for genetic and metabolic engineering aiming the production of large amounts of lipids with specific fatty acids composition (Nicaud 2012).

When *Yarrowia lipolytica* grows in nitrogen depletion conditions with substrates as glucose, ethanol, and starch hydrolysates; citric acid is secreted in significant amounts (Papanikolaou *et al.*, 2006). This issue hampers the optimization of the carbon flux into lipids. The outbreak of lipid accumulation follows the exhaustion of a nutrient in the medium, mainly nitrogen (Papanikolaou and Aggelis, 2011a). This exhaustion reduces cell growth, but increases lipid content. The main challenge of this trade-off relationship is thus focused on increasing lipid content while maintaining an adequate cell growth through the manipulation of nitrogen fluxes (Yoo *et al.*, 2014).

Few mathematical models have been proposed to study lipid accumulation by oleaginous yeasts and fungus in different conditions and substrates (C. N. C. N. Economou et al., 2011; Fakas et al., 2009; Meeuwse et al., 2012, 2011; Ykema et al., 1986). For example, the model developed by (C. N. C. N. Economou et al., 2011) for the fungus *Mortierella isabellina*, assumes that high substrate concentration inhibits microbial growth and that high nitrogen concentration decreases lipid concentration. Meeuwse *et al.*, (2012) reported a model for the fungus *Umbelopsis isabellina* in three phases: exponential growth phase, lipid accumulation phase, and lipid turnover phase. This was accomplished by the definition of certain fixed times to switch phases. Models with a particular interest on *Yarrowia lipolytica* have considered the mechanisms of lipid accumulation (Papanikolaou and Aggelis, 2003b, 2002), and citric acid production (Arzumanov *et al.*, 2000; Papanikolaou and Aggelis, 2003a), separately. All these models represented acceptably the dynamics of growth and accumulated lipids concentration from fat substrates. However, there is only one existing model for the concomitant production of lipids and citric acid (Papanikolaou *et al.*, 2006), which fails to predict lipid accumulation since it was applied for the production of structural lipids (membrane lipids).

Other modeling approaches have considered variant yields, such as the model proposed by Droop (Droop, 1968), where biomass production depended on a variable cell quota defined as the fraction of internal nutrient per unit of biomass (Lemesle and Mailleret, 2008). These models have only been presented for microalgae on: autotrophic (Mairet *et al.*, 2011), heterotrophic (Abdollahi and Dubljevic, 2012; De la Hoz Siegler et al., 2011; Surisetty et al., 2010a), and mixotrophic (Adesanya *et al.*, 2014; Yoo *et al.*, 2014) conditions. Nevertheless, lipid accumulation in those models for microalgae was also triggered by the depletion of nitrogen, and thus this similarity between microalgae and yeast is encouraging to propose new

models based on quota definitions. In this context, this paper proposes a quota based model for yeast in order to expose its efficiency on modeling and optimization problems.

The present paper is organized as follows: Section IV.3 describes the experimental data used in this study, and it details the developed mathematical models. Section IV.4 presents the calibration and the validation of the models, and it proposes the optimization of lipid accumulation by finding the optimal input flow rates. Finally, conclusions are drawn in section IV.5.

IV.3. Materials and Methods

IV.3.1. Strain and medium

Three independent experiments (culture A, B and C) were performed with the strain *Yarrowia lipolytica* W29 obtained from the Microbiology and Molecular genetics laboratory (Paris-Grignon, France). The optimal growth conditions of this yeast are: 28 °C and pH of 5.6 (Faure, 2002).

The strain was maintained at −80°C in yeast extract peptone dextrose (YPD) medium (10 $g.L^{-1}$ yeast extract, 10 $g.L^{-1}$ peptone and 10 $g.L^{-1}$ glucose) supplemented with 30% (v/v) glycerol. For each culture, one glycerol stock was streaked on a Petri dish (YPD agar medium containing 10 $g.L^{-1}$ agar) and incubated at 30°C for 48 h. Pre-cultivations were carried out first in a 5 mL tube of YPD medium at 30°C for 16 h inoculated with one colony from Petri dish. Then, two successive steps of 16 h pre-cultures were carried out in baffled-shake flasks inoculated with a 10 % (*v*/*v*) ratio with increasing culture volumes and containing mineral medium with the following composition: K_2HPO_4: 3 $g.L^{-1}$; $(NH_4)_2SO_4$: 3 $g.L^{-1}$; NaH_2PO_4,H_2O: 3 $g.L^{-1}$; $MgSO_4,7H_2O$: 1 $g.L^{-1}$; $ZnSO_4,7H_2O$: 0.04 $g.L^{-1}$; $FeSO_4,7H_2O$: 0.0163 $g.L^{-1}$; $MnSO_4,H_2O$: 0.0038 $g.L^{-1}$; $CoCl_2,6H_2O$: 0.0005 $g.L^{-1}$; $CuSO_4,5H_2O$: 0.0009 $g.L^{-1}$; $Na_2MoSO_4,2H_2O$: 0.00006 $g.L^{-1}$; $CaCl_2,2H_2O$: 0.023 $g.L^{-1}$; H_3BO_3: 0.003 $g.L^{-1}$; and 10 mL of vitamins solution (Ochoa-Estopier, 2012). The reactor, which contained the same composition of mineral medium without $(NH_4)_2SO_4$, was inoculated with a 10% (v/v) ratio.

IV.3.2. Analytical methods

The data from the three independent experiments were determined with the same protocols. Yeast concentration was quantified by spectrophotometry at 600 nm in a spectrophotometer HITACHI U-1100. For dry cell weight, culture samples (5–10 mL) were harvested by filtration on a 0.45 μm membrane (Sartorius) and dried at 200 mm Hg and 60 °C during 48 h until a constant weight was achieved. Determination of alcohols, organic acids, and sugar concentrations from supernatants was performed by HPLC using a column Aminex HPX-87H (300 mm*7.8 mm) with the following conditions: Temperature of 50 °C, with 5 mM H_2SO_4 as eluent (flow rate of 0.5 mL.min^{-1}) and dual detection (refractometer and UV at 210 nm). Compounds were identified and quantified with respect to standards. An ammonium ion electrode (PH/ISE meter model 710A + Ammonia Gas Sensing Electrode Model 95-12, Orion Research Inc. Boston USA) was used to quantify the residual ammonia concentration in the culture medium. The quantification of lipids was carried out as described by Browse et al. (1986) 20 mg of lyophilized cells were added to 1 mL of methanol solution containing 25 ml.L-1 of sulfuric acid (95%) and internal standard (C9:0). The mix in sealed tubes was placed in a 80°C bath for 90 minutes. After cooling, 450µL hexane and 1.5 mL water were added and vortexed. The lipidic phase was then sampled and the composition was analyzed by Gas Chromatography as mentioned in Cescut *et al.*, (2011).

IV.3.3. Experimental data

Culture A (Figure III.19a) is a fed-batch reported by (Cescut, 2009) where *Yarrowia lipolytica* was grown in a 20 L bioreactor (Biostat E. Braun, Melsungen, Germany). The dissolved oxygen was maintained above 20% of saturation by modulating the stirring rate and/or the air flow rate in order to avoid oxygen limitation. Inoculation volume was of 800 mL. Operating conditions were fixed and controlled to a temperature of 28°C and a pH value of 5.5 regulated by addition of a 10 M NH_3 solution (growth phase) or 5 M KOH solution (after nitrogen limitation). Input air flow rates and composition of input and output air, as well as the bioreactors parameters (*e.g.* Temperature, pO2, pH, agitation) were measured online and acquired by custom made software. The flow rates of nitrogen (N_{IN} , ammonia 5 M) and carbon (S_{IN} , glucose 730 g.L^{-1}) sources varied continuously (Figure 1 a) to induce different levels of nitrogen limitation.

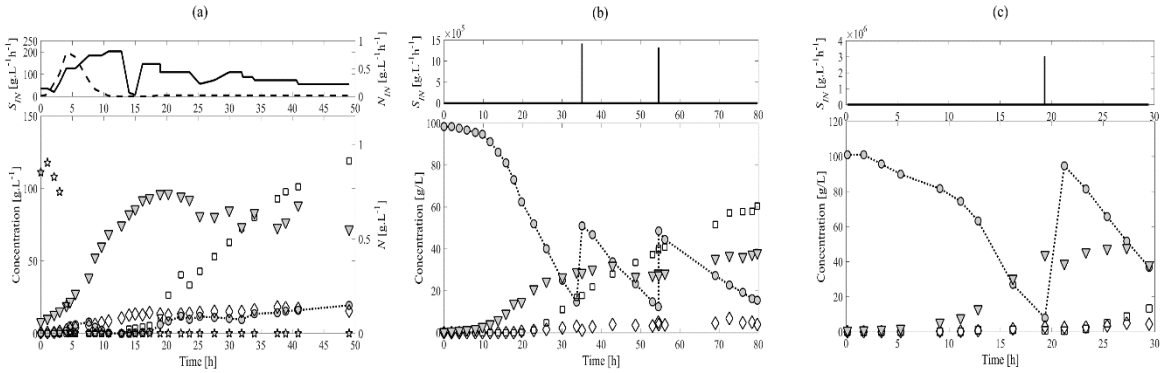


Figure III.19. Experimental data (a) <u>Culture A</u>: Fed-Batch culture (Cescut, 2009). (b) <u>Culture B</u>: Sequential-batch culture. (c) <u>Culture C</u>: Sequential-batch culture for model validation. The solid line (-) represents the input flow rate of glucose and (--) the input flow rate of ammonia. (-●-) Glucose, (□) Citrate, (◊) Lipids, (▼) Biomass, (☆) Nitrogen.

Culture B (Figure III.19) corresponds to a sequential batch culture under nitrogen starvation. This was obtained from a 5 L stirred tank bioreactor (Biostat B. Braun Biotech International, Sartorius AG, Germany) with the acquisition software MFCS/win 2.1®. The inoculation volume was 300 mL. The dissolved oxygen was maintained above 20% of saturation. The working volume of the reactor was 3 L, where the operating conditions were set to a controlled temperature of 28°C and a pH value of 5.6 regulated by a 5 M KOH solution. The initial medium contained 100 $g.L^{-1}$ of glucose and 4.14 $g.L^{-1}$ of $(NH_4)_2SO_4$. The carbon source (glucose) was fed by two pulses at 35 h and 55 h (Figure 1 b). The first half of this culture was used for model calibration and the second half for model validation.

Culture C (Figure III.19c) is a sequential batch culture under nitrogen starvation, where the carbon source (glucose) was fed by one pulse at 19.3 h (Figure III.19c). The initial medium contained 6.6 $g.L^{-1}$ of $(NH_4)_2SO_4$ and 100 $g.L^{-1}$ of glucose. The reactor and operating conditions were the same as the reported for culture B.

IV.3.4. Models assumptions

Our models assume that the metabolism and cell growth of *Yarrowia lipolytica* are driven by the uptake of two extracellular nutrients: carbon (*i.e.* glucose) and nitrogen (*i.e.* ammonia).

Figure III.20 displays the different physiological states occurring in the cell. These phenomena have as first priority the production of functional biomass (catalytically active biomass), which has an elementary composition of $CH_{1.744}O_{0.451}N_{0.132}$ (Ochoa-Estopier, 2012). Under intracellular nitrogen deficient or limited conditions, the yeast accumulates lipids (mainly TAGs) and storage compounds (*e.g.* carbohydrates). However, under these conditions and carbon excess citric acid is produced and excreted (Papanikolaou *et al.*, 2006 ; Ochoa-Estopier and Guillouet, 2014). This production is the result of overflow metabolism leading to by-product formation (Amribt *et al.*, 2013; Santos *et al.*, 2012). These by-products are often inhibitory or competing products (*i.e.* citric acid) that could decrease the production of the target metabolite (TAGs in our case).

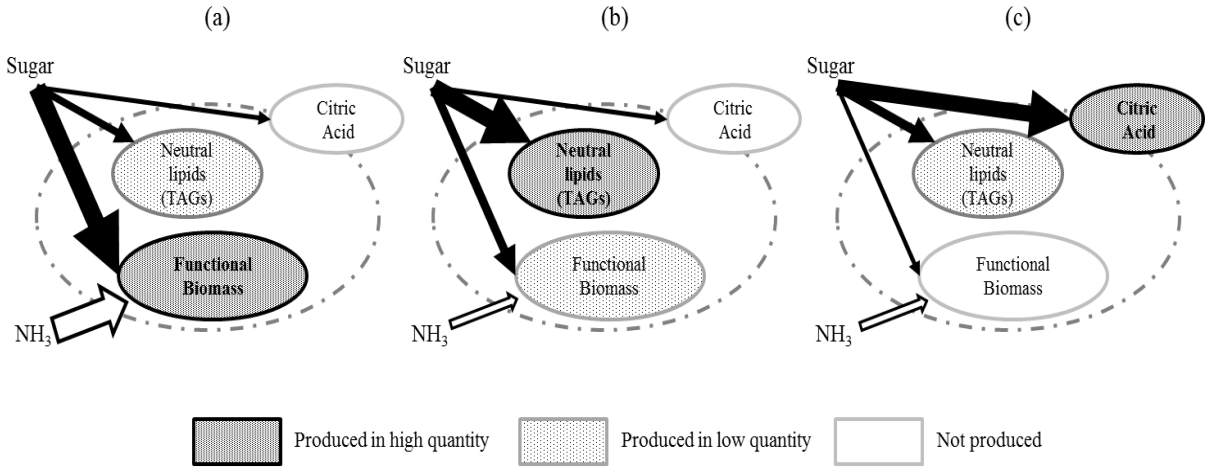


Figure III.20. Distribution of carbon flux in the metabolism of *Yarrowia lipolytica* for different physiological states: (a) Production of functional biomass. (b) Lipid accumulation and growth. (c) Excretion of citric acid.

Sugars can also have an inhibitory effect and it was thus considered as suggested by Economou *et al*. (2011). The dynamics of *Yarrowia lipolytica* metabolism under discontinued operation can be described by balance equations considering the concentrations ($g.L^{-1}$) of Glucose (*S*), Nitrogen (*N*), Functional biomass (X_f), Lipid biomass (X_L), Citric Acid (*CIT*), and volume (*V*, expressed in L) undertaking homogenous well-mixed conditions with non-limited oxygen requirements. It is noteworthy that the functional biomass contains a part of lipids, which are components of the membrane. These lipids are called structural lipids X_{LS} and are intrinsic to the functional biomass in a constant fraction γ equivalent to 0.06 $gX_L.gX_f^{-1}$, (Robles-Rodriguez *et al.*, 2017). This definition allows to accounting for a separate term of accumulated lipids X_{LA}.

For simplicity and consistency with experimental data, only Total lipids ($X_L = X_{LS} + X_{LA}$ with $X_{LS} = \gamma . X_f$) are described in the mass balances.

$$\frac{d}{dt} S = S_{IN} . \frac{f_{S_{IN}}}{V} - \left(\mu \frac{1}{Y_{X/S}} + \pi_{LIP} \frac{1}{Y_{LIP/S}} + \pi_{CIT} \frac{1}{Y_{CIT/S}} \right) X_f - \frac{S}{V} F_{IN} \quad \text{(III.40)}$$

$$\frac{d}{dt} N = N_{IN} . \frac{f_{N_{IN}}}{V} - \rho_N X_f - \frac{N}{V} F_{IN} \quad \text{(III.41)}$$

$$\frac{d}{dt} X_f = \mu . X_f - \frac{X_f}{V} F_{IN} \quad \text{(III.42)}$$

$$\frac{d}{dt} X_L = (\pi_{LIP} + \gamma . \mu) X_f - \frac{X_L}{V} F_{IN} \quad \text{(III.43)}$$

$$\frac{d}{dt} CIT = \pi_{CIT} . X_f - \frac{CIT}{V} F_{IN} \quad \text{(III.44)}$$

$$\frac{dV}{dt} = F_{IN} - F_{Sample} \quad \text{(III.45)}$$

hold for the input concentration of glucose and nitrogen, respectively. The term F_{IN} (L.h^{-1}) considers the inputs to the reactor which is mathematically defined as the sum of the input flow rate of glucose $f_{S_{IN}}$ and nitrogen $f_{N_{IN}}$ as $F_{IN} = f_{S_{IN}} + f_{N_{IN}}$. The volume equation (III.45) includes a term F_{Sample} to take into account sampling volumes. The terms $Y_{X/S}$, $Y_{LIP/S}$, and $Y_{CIT/S}$ are the yield coefficients (g.g^{-1}) for functional biomass, lipids, and citric acid production with respect to glucose, respectively. Growth rate, lipid production rate, and the citric acid production rate, are represented by μ, π_{LIP}, and π_{CIT} respectively, whereas the nitrogen uptake rate is ρ_N. For consistency with experimental data, total Biomass concentration (X_T) is accounted as the sum of the functional biomass (without the structural lipid part) and lipid biomass as,

$$X_T = X_f (1 - \gamma) + X_L \quad \text{(III.46)}$$

Growth rate μ was defined to take into account the dependency of yeast growth on carbon and nitrogen sources. The impact of nitrogen on growth and lipid accumulation is further detailed. This nitrogen effect has been studied on yeast and fungus by kinetic models (C. N. C. N. Economou et al., 2011; Papanikolaou et al., 2006; Ykema et al., 1986), and in microalgae by

the definition of intracellular quotas as in the Droop model (Droop, 1968; Mairet *et al.*, 2011). In this context, two macroscopic models are here developed based on equations (III.40) – (III.45).

IV.3.5. Unstructured Model

The unstructured model assumes that external nitrogen is taken up directly for the production of biomass. Therefore, in the growth rate (μ), the uptake of carbon and nitrogen is described by Monod-type equations denoting a double substrate limitation on nitrogen and glucose as,

$$\mu = \mu^{\max}\left(\frac{S}{K_{S,1} + S + S^2/k_{I,1}}\right)\left(\frac{N}{K_N + N}\right) \tag{III.47}$$

where $K_{S,1}$, and K_N are the saturation constants for glucose and nitrogen, respectively, $\mu^{\max}$ is the specific maximum growth rate, and $k_{I,1}$ is the glucose inhibition constant. In addition, the nitrogen uptake rate (ρ_N) is considered to be proportional to growth rate,

$$\rho_N = \mu\frac{1}{Y_{X/N}} \tag{III.48}$$

where $Y_{X/N}$ is the yield coefficient for functional biomass production with respect to nitrogen.

The specific lipid accumulation rate π_{LIP} was assumed to follow Andrew's equation for glucose. Nitrogen depletion was considered by an inhibition constant (k_1) ensuring that lipid accumulation is low at high concentrations of nitrogen as developed by (C. N. C. N. Economou et al., 2011). Moreover, an inhibition by citric acid was added to include a competition with the production of citric acid.

$$\pi_{LIP} = \pi_{LIP}^{\max}\left(\frac{S}{K_{S,2} + S + S^2/k_{I,2}}\right)\left(\frac{k_1}{k_1 + N}\right)\left(\frac{k_2}{k_2 + CIT}\right) \tag{III.49}$$

The term π_{LIP} is the maximum lipid production rate. $K_{S,2}$, $k_{I,2}$, and k_2 represent the saturation and inhibition constants for glucose, and the inhibition constant due to citric acid concentration, respectively.

Citric acid production occurs as overflow metabolism under nitrogen starvation or limitation conditions and excess of carbon substrate (*i.e.* glucose). This implies a decrease on the production rate of functional biomass (μ) and a stabilization of lipid production rate (π_{LIP}).

Therefore, it can be assumed that when the Glucose Uptake Rate for lipids and biomass (GUR_X) decreases to a minimum point Of_U , cells switch to overflow metabolism.

$$GUR_X = \mu \frac{1}{Y_{X/S}} + \pi_{LIP} \frac{1}{Y_{LIP/S}} \tag{III.50}$$

Hence, glucose uptake rate and the saturation point Of_U can be used to deduce an indicator (Ind_U) of overflow metabolism (Amribt *et al.*, 2013).

$$Ind_U = \begin{cases} 0, & GUR_X > Of_U \\ 1, & GUR_X \leq Of_U \end{cases} \tag{III.51}$$

The indicator Ind_U is defined as Boolean for which $Ind_U = 1$ assumes overflow. Citric acid production rate (π_{CIT}) is therefore described by this indicator as,

$$\pi_{CIT} = \pi_{CIT}^{\max}\left(Ind_U\right)\left(\frac{S}{K_{S,3} + S + S^2/k_{I,3}}\right)\left(\frac{k_3}{k_3 + CIT}\right) \tag{III.52}$$

Additionally, the citric acid production rate, π_{CIT} , also considers an inhibition due to citric acid. The maximum citric acid production rate is indicated by $\pi_{CIT}^{\max}$, inhibition constants for glucose, and the inhibition constant for citric acid are $K_{S,3}$, $k_{I,3}$, and k_3 , respectively.

IV.3.6. Intracellular Quota Model

In this work, it is assumed that the Droop model (Droop, 1968) used for microalgae on heterotrophic conditions can also be applied for yeasts. The main characteristic of this model is the inclusion of structured variables to represent variable yield coefficients called quotas. In this way, the Droop model provides regulation, particularly during substrate limitation or starvation, where the internal quota of a limiting substrate drives the production of functional biomass (Mairet *et al.*, 2011).

Microbial growth is therefore modulated by the internal nitrogen quota q_N , defined by

$$\frac{d}{dt} q_N = \rho_N - \mu\, q_N \tag{III.53}$$

where ρ_N is the nitrogen uptake rate, and μ is the growth rate. In this model, the uptake of carbon for growth is described by the Andrew's equation, while the nitrogen effect is described by the Droop model. Hence, growth rate is denoted as,

$$\mu = \mu^{*}\left(1-\frac{Q_0}{q_N}\right)\left(\frac{S}{K_{S,1}+S+S^2/k_{I,1}}\right) \quad \text{(III.54)}$$

where Q_0 is the minimum nitrogen quota at which cell growth is possible, and μ^{*} is the theoretical maximum growth rate. The nitrogen uptake rate ρ_N is given by

$$\rho_N = \rho_N^{\max}.\left(\frac{N}{K_N+N}\right) \quad \text{(III.55)}$$

where K_N is the nitrogen saturation constant, and $\rho_N^{\max}$ is the maximum nitrogen uptake rate.

The lipid accumulation rate, π_{LIP}, was also described by Andrew's equation for glucose. An inhibitory effect on citric acid was also included as in (III.49). Moreover, it is assumed that lipid production decreases as the lipid content reaches its theoretical maximum (De la Hoz Siegler et al., 2011; Surisetty et al., 2010b; Yoo et al., 2014).

$$\pi_{LIP} = \pi_{LIP}^{\max}.\left(\frac{S}{K_{S,2}+S+S^2/k_{I,2}}\right)\left(1-\frac{1}{\alpha}\frac{X_L}{X_T}\right)\left(\frac{k_2}{k_2+CIT}\right) \quad \text{(III.56)}$$

The term α represents the maximum lipid content, which is 0.36 g*L*.g$X_T$$^{-1}$ for *Yarrowia lipolytica* W29 (Beopoulos et al., 2009a; Ratledge and Wynn, 2002). The overflow to produce citric acid is the result of an excessive substrate supply. In this context, the carbon uptake for lipids and functional biomass must encounter a saturation point. It is known that lipid accumulation in yeast depends on the N/C ratio of the inflow rate of nitrogen and carbon (Beopoulos et al., 2009a; Cescut, 2009; Ochoa-Estopier and Guillouet, 2014; I. R. Sitepu et al., 2014; Ykema et al., 1986). The critical values for the N/C ratio at which citric acid is produced concomitantly with lipids have been reported at 0.021 molN.Cmol^{-1} (Ochoa-Estopier and Guillouet, 2014) for D-stat experiments, and identified at 0.077 molN.Cmol^{-1} (Robles-Rodriguez *et al.*, 2017) after model identification. In addition to the input flow rates, some studies have focused on the changes of the consumption rate of the N/C ratio, which is related to the intracellular change of N/C. Based on these ideas, it is straightforward to think that there exists a critical quota (q_N*), where the process shifts towards citric acid overflow. Therefore, it is possible to assume that citric acid is produced when q_N is lower than a critical value q_N*. In order to keep consistency with the formulation of (III.52), a Boolean indicator (Ind_Q) is also proposed as,

$$Ind_Q = \begin{cases} 0, & q_N > q_N{}^* \\ 1, & q_N \le q_N{}^* \end{cases} \quad \text{(III.57)}$$

Therefore, citric acid production occurs when $q_N \le q_N{}^*$. Citric acid production rate, π_{CIT}, was then taken as in (III.52) in which the Boolean indicator Ind_U was replaced by Ind_Q,

$$\pi_{CIT} = \pi_{CIT}^{\max} . (Ind_Q) . \left(\frac{S}{K_{S,3} + S + S^2 / k_{I,3}} \right) \left(\frac{k_3}{k_3 + CIT} \right) \quad \text{(III.58)}$$

The difference to the unstructured model relies on the manner of computing the indicator. Summarizing, the unstructured model describes 6 state variables whose dynamics are represented by equations (III.40) – (III.45), and their kinetics by equations (III.47) – (III.52). The resulting model has 18 parameters. The quota based model has also 18 parameters with 7 mass balance equations: 6 state variables (equations (III.40) – (III.45)), and the intracellular nitrogen quota (equation (III.53)). Kinetics are disclosed on equations (III.54) – (III.58). The total biomass is evaluated by equation (III.46) for both models. The calibration and validation of the models are presented in the next section.

IV.4. Results and Discussions

IV.4.1. Model calibration

Parameter calibration was carried out by a Particle Swarm Optimization (PSO) algorithm (Eberhart and Kennedy, 1995), and further refined by Pattern Search (Lewis and Torczon, 1996) implemented in MATLAB®. The data of Culture A and Culture B (t=0 – 38h) were used to calibrate the models, whose results are displayed in Figure III.21.

The unstructured model and the quota model depict similar fit in both cultures. However, it is remarkable that nitrogen concentration is overestimated by 2 $g.L^{-1}$ before 10 h with the unstructured model for culture A, whereas the estimation with the quota model is in better agreement with the experimental data. Both models display small underestimation in lipids and citric acid concentration at the end of culture A. This is due to the estimation trade-off between the data from both cultures where lipids and citric acid concentrations are smaller for culture B than for culture A.

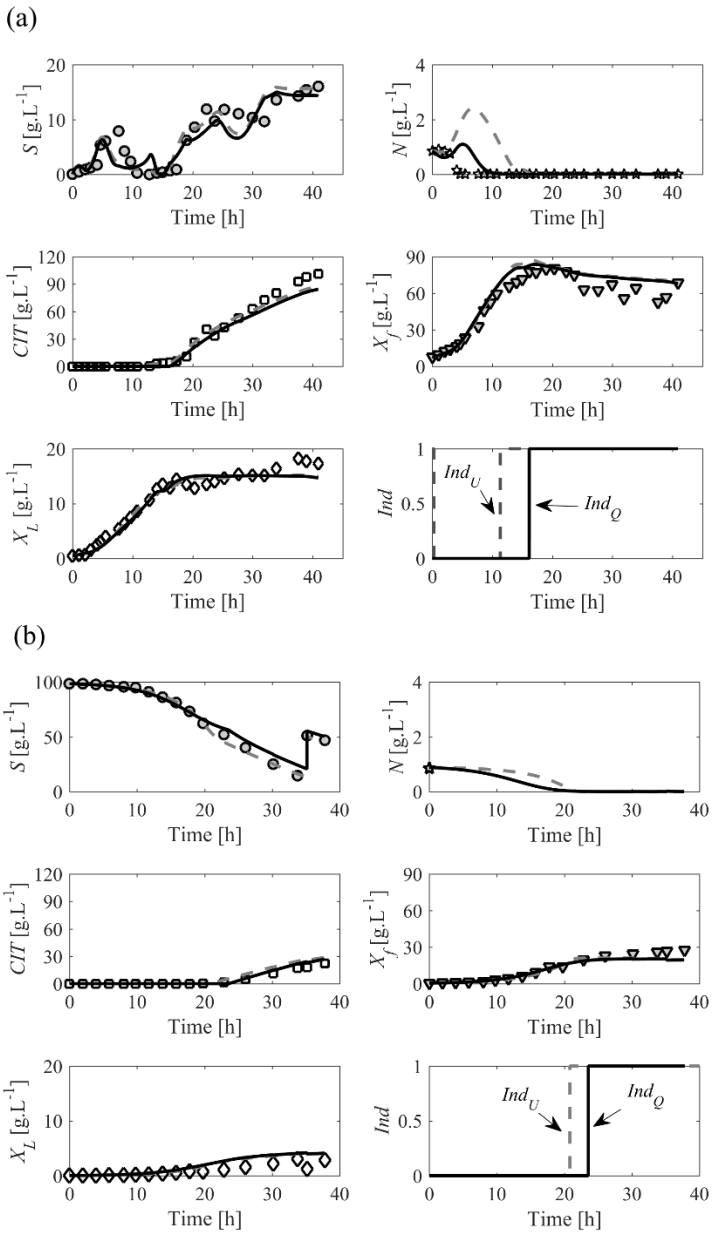


Figure III.21. Model calibration. Performance of the two models with the experimental data of (a) Culture A. and (b) Culture B (*t* = 0 – 38h). (— –) Unstructured model. (——) Quota model

In Figure III.21b, it is observed that when nitrogen was depleted (21 h for the quota and 23 h the unstructured model) functional biomass was no longer produced. This could be explained regarding some of the assumptions of the models. The unstructured model assumed that growth was based on glucose and nitrogen consumption; thus when nitrogen was depleted the total biomass would have increased due to lipids accumulation. On the other hand, the quota model assumed that there is a minimum nitrogen quota Q_0 to assure growth, which was reached

subsequently to the extracellular nitrogen depletion causing a stabilization of functional biomass. Nevertheless, the results indicate that functional biomass concentration continued to rise probably due to the storage of other compounds that were not measured (*i.e.* carbohydrates). The overflow indicators (Ind_U or Ind_Q) are displayed for both cultures (Figure III.21), where it is shown that the unstructured model started before due to the differences on the depleted nitrogen for both cultures.

The estimated parameters for both models are presented in Table III.12 and compared with previous values reported in the review of Papanikolaou and Aggelis, (2011a). The values of $\mu^{\max}$, $\pi_{LIP}^{\max}$, and $Y_{LIP/S}$ were also experimentally identified for culture A. The theoretical maximum growth rate $\mu*$ for the quota model is higher than the one obtained experimentally, as expected since it depends on the maximum and minimum values of the quota. In the unstructured model, $\mu^{\max}$ is close but also higher than the experimental value due to the glucose inhibition term. With respect to the values of the specific production rates $\pi_{LIP}^{\max}$, they were consistent with the parameters estimated by (Fakas *et al.*, 2009), and the experimental values. Moreover, lipid yields $Y_{LIP/S}$ were also consistent with the experimental values. With respect to the parameters affecting citric acid, the estimated maximum specific production rate $\pi_{CIT}^{\max}$ and the yields $Y_{CIT/S}$ were higher than the reported in literature (Papanikolaou et al., 2008), which could be explained by the high production of citric acid when glucose was added continuously and the differences on the kinetic equations. Regardless their differences in structure, both models reported similar results in the non-growth related parameters. The main distinction was observed on the saturation constant for nitrogen K_N, where the higher value on the quota model is compensated by the nitrogen quota and it is almost equivalent to $Y_{X/N}$. This was expected since the equations of nitrogen uptakes for both approaches (Unstructured (III.48) and Quota model (III.55)) are different. The values of the inhibition constants (Table III.12) show that the glucose inhibition effect differs in magnitude between citric acid production, lipid accumulation, and growth. The estimated value of the critical quota (q_N*) was interestingly found to be similar to the critical value of the N/C ratio reported by (Robles-Rodriguez *et al.*, 2017).

Table III.12. Estimated kinetic parameters for the fed-batch cultures of *Yarrowia lipolytica*.

Parameter	This work		References	
	Unstruc	Quota		
μ^{max} [h^{-1}]	0.415	-	0.26	Culture A[a]
			0.18-0.36	(Papanikolaou and Aggelis, 2003b)[b]
μ^{*} [h^{-1}]	-	0.755	0.566	(C. N. C. N. Economou et al., 2011)[d]
π_{LIP}^{max} [g*LIP*.(gX_f-h)$^{-1}$]	0.030	0.039	0.034	Culture A[a]
			0.028	(Fakas *et al.*, 2009)[c]
			0.0027	(Papanikolaou *et al.*, 2006)[b]
π_{CIT}^{max} [g*CIT*.(gX_f-h)$^{-1}$]	0.301	0.314	0.035-0.13	(Papanikolaou and Aggelis, 2003a)[b]
			0.013	(Papanikolaou *et al.*, 2008)[b]
$Y_{X/S}$ [gX_f.gS^{-1}]	0.603	0.519	0.78 – 0.86	(Papanikolaou and Aggelis, 2003b)[b]
$Y_{LIP/S}$ [g*LIP*.gS^{-1}]	0.276	0.272	0.23	Culture A[a]
			0.63-0.86	(Papanikolaou and Aggelis, 2003b)[b]
$Y_{CIT/S}$ [g*CIT*.gS^{-1}]	0.88	0.89	0.63	(Papanikolaou and Aggelis, 2003a)[b]
			0.45-0.82	(Papanikolaou *et al.*, 2008)[b]
$Y_{X/N}$ [gX_f.gN^{-1}]	23.23	-	31.9-33.6	(Papanikolaou and Aggelis, 2003a)[b]
			39.3	(Fakas *et al.*, 2009)[c]
			18.21	(C. N. C. N. Economou et al., 2011)[d]
$K_{S,1}$ [gS.L^{-1}]	1.517	1.116	20	(Fakas *et al.*, 2009)[c]
			1.256	(C. N. C. N. Economou et al., 2011)[d]
$K_{S,2}$ [gS.L^{-1}]	0.978	1.482	69.27	(C. N. C. N. Economou et al., 2011)[d]
$K_{S,3}$ [gS.L^{-1}]	11.26	12.22		-
K_N [gN.L^{-1}]	0.074	23.08	0.196	(Papanikolaou and Aggelis, 2003a)[b]
			0.179	(Arzumanov *et al.*, 2000)[b]
			0.085	(C. N. C. N. Economou et al., 2011)[d]
k_1 [gN.L^{-1}]	924	-	835	(C. N. C. N. Economou et al., 2011)[d]
k_2 [g*CIT*.L^{-1}]	8.3	9.75		-
k_3 [g*CIT*.L^{-1}]	40.44	37.63		-
$k_{I,1}$ [gS.L^{-1}]	83.14	83.64	20	(C. N. C. N. Economou et al., 2011)[d]
$k_{I,2}$ [gS.L^{-1}]	35.2	44.17	0.399	(C. N. C. N. Economou et al., 2011)[d]
$k_{I,3}$ [gS.L^{-1}]	28.8	40.11		
Overflow:				
Of_U [gS.(gX_f-h)$^{-1}$]	0.268	-		-
Quotas:				
ρ_N^{max} [gN.(gX_f-h)$^{-1}$]	-	0.716		-
Q_0 [gN.(gX_T)$^{-1}$]	-	0.045		-
q_N* [mol$N$.Cmol^{-1}]	-	0.084		-

[a]Experimental values of Culture A with *Yarrowia lipolytica*. [b]*Yarrowia lipolytica*. [c]*Thamnidium elegans.*

[d]*Mortierella isabellina*

Table III.13. Statistical evaluation of the models on the calibration data sets.

Model	Criterion	Culture A				
		S	N	CIT	X_f	X_L
Unstructured Model	R^2	0.845	0.027	0.987	0.963	0.966
	RMSE	2.195	0.944	5.236	8.722	1.142
Quota Model	R^2	0.811	0.307	0.992	0.963	0.958
	RMSE	2.359	0.341	6.754	7.845	1.245
		Culture B (*t*=0 – 38 *h*)				
Unstructured Model	R^2	0.916	*NA*	0.984	0.957	0.835
	RMSE	9.433		4.091	3.044	1.063
Quota Model	R^2	0.904		0.995	0.966	0.836
	RMSE	8.398		2.472	3.182	1.144

NA: Not Available

The determination coefficient R^2, and the Root Mean Squared Errors (*RMSE*) were calculated to prove the quality and the enforcement of each model (Table III.13). The results for nitrogen in culture B (nitrogen starvation culture) are not reported since nitrogen data was not available. The results of the *RMSE* and R^2 demonstrate similar conclusions for the different variables, which are also consistent with Figure III.21. With respect to the *RMSE* for culture A, it is seen that the unstructured model presented smaller values for citric acid and glucose, whereas the quota model was better for functional biomass, nitrogen and lipid concentrations. For culture B, the *RMSE* values for glucose and citric acid concentrations were smaller for the quota model with similar values for functional biomass and lipid concentrations.

IV.4.1. Model validation

Culture B (t = 38 – 80h) and culture C were used to validate the model, considering sequential carbon feeding cultures under nitrogen starvation. The validation results of the two models are displayed in Figure 4. Both models presented good agreement with the experimental data where a similar performance is observed for both validation cultures. It is worth noting that experimental data for nitrogen concentration was not available for culture B, and only the initial point of culture C was known from the initial medium composition.

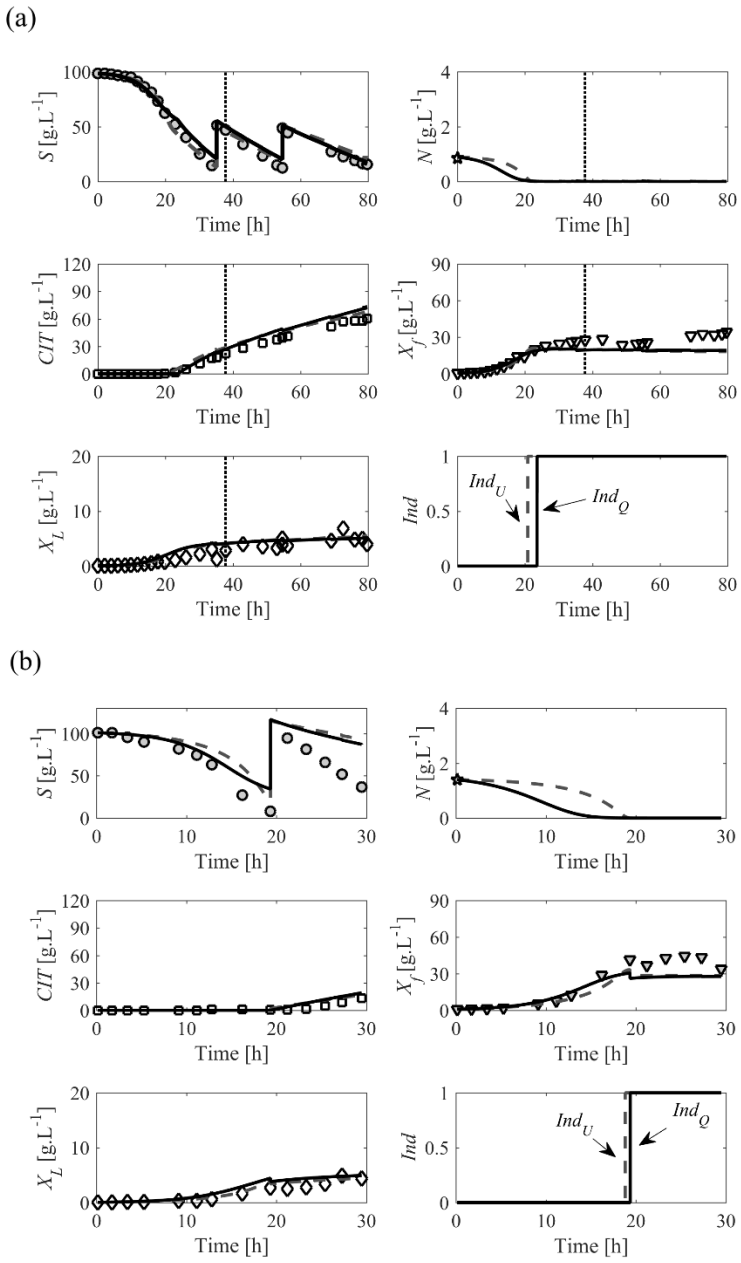


Figure III.22. Model validation. Prediction of the models with the experimental data of sequential batch cultures under nitrogen deficiency. (a) Culture B (*t* = 38 – 80*h*). (b) Culture C. (— –) Unstructured model. (——) Quota model

In Figure III.22, functional biomass concentration was stable after nitrogen depletion due to the growth dependency on the nitrogen source availability. The main differences between models were observed on nitrogen concentrations, for which no comparison could be made due to the lack of data. However, this difference was reflected on the shift to overflow metabolism identified by the indicators in equations (III.51) and (III.57) for the unstructured and the quota

model, respectively. The switching times for overflow in culture B (Figure III.22a) were at 21 h for the quota model and at 23 h for the unstructured model. The difference between switching times were also observed for culture C (Figure III.22b) and culture A (Figure III.21a). However, this difference is bigger on culture A. This can be explained by the two different phenomena acting on the cell metabolism: nitrogen limitation or nitrogen starvation.

Statistical analysis was also performed for both models with the validation cultures, whose results are depicted in Table III.14. The values of the *RMSE* showed that the quota model performed better estimation for glucose and functional biomass concentrations, whilst citric acid was lower with the unstructured model, and similar results were reported for lipid concentrations in culture B. These results for *RMSE* were maintained for culture C.

Table III.14. Statistical evaluation of the models on the validation data sets.

Model	**Criterion**	**Culture B (*t*=38 – 80 *h*)**				
		S	N	CIT	X_f	X_L
Unstructured Model	R^2	0.392	*NA*	0.989	0.622	0.363
	RMSE	10.293		7.731	10.368	1.041
Quota Model	R^2	0.479		0.991	0.476	0.361
	RMSE	9.617		9.439	9.836	0.985
		Culture C				
Unstructured Model	R^2	0.645	*NA*	0.854	0.973	0.918
	RMSE	23.285		2.855	7.303	0.575
Quota Model	R^2	0.719		0.894	0.949	0.902
	RMSE	20.545		3.174	7.695	0.884

NA: Not Available

To strengthen the validity of our model, *in silico* simulations were challenged against published experimental data of different strains of *Yarrowia lipolytica* growing on glucose under batch conditions (Kavšček et al., 2015; Papanikolaou et al., 2009, 2008, 2006). The comparison of results is displayed in Figure III.23. Our model was able to reproduce accurately the experiments reported in the literature. However, the model predicted higher lipid concentrations when initial glucose concentration was high (150 $g.L^{-1}$). Additionally, the model was tested to represent the dynamics of a mutant strain of *Y. lipolytica* where results demonstrated the strong predictive capabilities of the model (See Supplementary Material D).

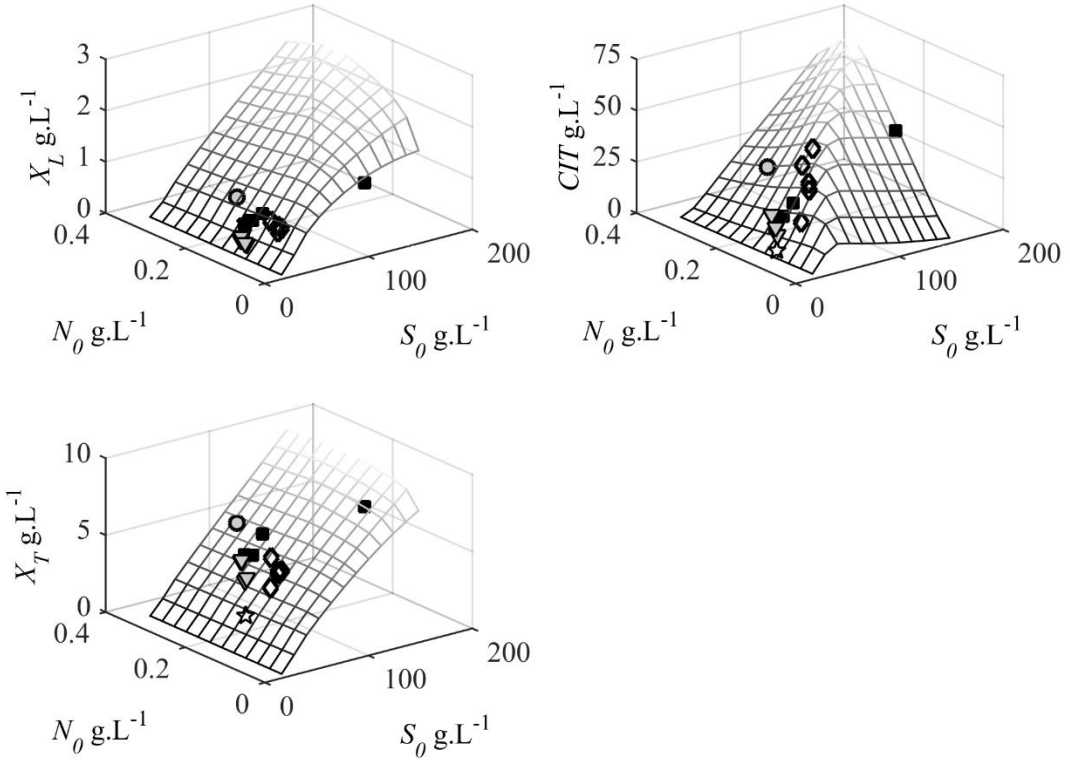


Figure III.23. Model validation with different experimental data from the literature in Batch conditions with glucose as substrate. (■) (Papanikolaou et al., 2006). (●) (Papanikolaou et al., 2008) (☆) (Kavšček et al., 2015). (▼) C/N=100 (Papanikolaou et al., 2009). (◊) C/N=200 (Papanikolaou et al., 2009).

The proposed models captured the underlying dynamics of *Yarrowia lipolytica* keeping track of the variations in concentrations of two different types of cultures and two phenomena on the cell metabolism. The models were capable of identifying overflow metabolism, which depends on both the carbon and nitrogen source. Overflow is not occurring if extracellular nitrogen is not depleted or if the critical intracellular nitrogen quota is not reached. However, the level of nitrogen limitation directly affects lipid accumulation. Special care should be payed to the feeding profiles of carbon and nitrogen to have a proper value of N/C ratio allowing maximization of lipids. Therefore, a control strategy to define the optimal feeding rates is needed. In this context, the quota model was selected to evaluate a control strategy aiming at maximizing lipid accumulation by manipulating the feeding profiles of nitrogen and glucose. This choice was based on the lower *RMSE* values for glucose concentrations with the quota

model for all the cultures, and the better description of extracellular nitrogen for Culture A, which are essential for regulating the N/C ratio.

IV.4.2. Optimization of Lipid accumulation

In fed-batch conditions, the optimization problem focusses on determining the feed rate policies for the substrates during the entire period of operation maximizing an objective function defined in terms of the status of the bioreactor. Until now, there has been a considerable amount of research about yeast cell culture optimization to avoid overflow metabolism (*e.g.* ethanol formation) (Renard and Vande Wouwer, 2008; Valentinotti *et al.*, 2003) while maximizing biomass. However, it has never been applied for other yeasts experiencing overflow, such as *Yarrowia lipolytica*, where it is necessary to avoid citric acid formation to favor lipid accumulation. Hereof, maximizing lipid production could be achieved by a careful determination of the feeding profiles of glucose and nitrogen to maintain the quota q_N near to the identified value of the critical quota q_N*. Optimizing lipid accumulation in yeast fed-batch cultures can be seen as a two-fold objective at a final fermentation time (t_f). The first objective is the maximization of the lipid productivity at a final time of operation,

$$P_L = \frac{X_L(t_f)}{t_f} \tag{III.59}$$

The second objective is the production of high lipid content biomass. In order to do it, let us consider the lipid content fraction of biomass as $w_L(t) = \frac{X_L}{X_T}$, and a target lipid content $w_{L,\max}$, which are put together into a new normalized variable P_W as,

$$P_W = \frac{w_L(t_f)}{w_{L,\max}} \tag{III.60}$$

where the target lipid fraction was set to $w_{L,\max}$ = 0.36 (g.g^{-1}), which is the maximum theoretical value (Beopoulos et al., 2009a).

The multi-objective problem is defined as,

$$\max_{f_{S_{IN}}, f_{N_{IN}}, t_f} \quad w_P . P_L + P_W \tag{III.61}$$

$$\text{Subject to: } \begin{cases} Eq\,(III.40 - III.45) \\ Eq\,(III.52) \\ S(t_f) < S_{\max},\ N(t_f) < N_{\max} \\ V_0 < V < V_{\max},\ t_{\min} \le t_f \le t_{\max}, \\ 0 \le f_{S_{IN}} \le f_{S_{IN}\max},\ 0 \le f_{N_{IN}} \le f_{N_{IN}\max},\ n_{PWL} = n_f \end{cases} \tag{III.62}$$

The parameter w_P is a weight coefficient to favor a high lipid content and it was identified from constant flow rates and fixed to 0.25. The quota model was used to determine the optimal inputs and to predict the response of the state variables. Even if the model allows to expressing the accumulated lipids X_{LA}, the term of total lipids $X_L = X_{LS} + X_{LA}$ was preferred to keep consistency with literature values. Additionally, the results of using X_{LA} or X_L were the same since the structural lipids $X_{LS} = \gamma . X_f$ are proportional to the functional biomass. Initial conditions of the state variables were taken from Culture A (Supplementary Material C). The final fermentation time was set between 20 and 90 h. The maximum allowed residual concentrations of glucose $S_{\max}$ and nitrogen $N_{\max}$ at the end of the culture were set to 2 g.L^{-1}, and 0.1 g.L^{-1}, respectively. The volume was bounded the initial value V_0 = 8 L to the maximum volume of the bioreactor V_{max} = 15 L. The upper limits for flow rates of carbon $f_{S_{IN}\max}$ and nitrogen $f_{N_{IN}\max}$ sources were fixed to 0.3 L.h^{-1}. A number n_{PWL} of piecewise linear (PWL) functions were used to parameterize the control $f_{S_{IN}}$ and $f_{N_{IN}}$. The optimization was performed with the Pattern Search algorithm of MATLAB®. For the sake of comparison, the optimization problem was solved for two cases: (i) constant optimal flow rates along the complete culture (n_f = 1); and (ii) five piecewise linear functions (n_f = 5). The PWL functions were defined with open time intervals to optimize the time length t_f of each linear function between $t_{\min}$ and $t_{\max}$ for a given n_f.

Figure III.24 shows the optimization results concerning the dynamics of the input variables to maximize the objective functions and the prediction of the state variables from the quota model. The constant input flow rates were: 0.162 L.h^{-1} for glucose, and 0.021 L.h^{-1} for nitrogen. The use of these constant optimal flow rates (dashed lines) led to a constant feeding with N/C ratio of 0.027 mol*N*.Cmol^{-1}, which allowed the production of citric acid. Although the value was

constant, nitrogen limitation occurred around 11.5 h, and was maintained until the end of the culture (30.3 h) where citric acid concentration achieved 1.28 $g.L^{-1}$. Concerning the optimization with 5 PWL functions (dark line), glucose feeding was increasing for 4 h, with a further small decrease of 13 h to assure growth and to moderate the concentration of residual glucose. After 17 h, the feeding profile increased again for 8 h to provide the necessary carbon for lipid accumulation. Finally, the glucose feeding decreased with two different slopes to fulfill the condition of residual glucose. Nitrogen feeding had a similar trend to glucose, where the second piecewise decreases the nitrogen feeding to induce nitrogen limitation, which is necessary for enhancing lipid accumulation. After 10 h the feeding rate was increased again at two different slopes for 19h to regulate the N/C ratio and to avoid citric acid production. The nitrogen flow dropped at the end of the culture to assure a low value of residual nitrogen. Concerning the quota, it is observed that for both cases the final values were around the critical quota q_N*, which is reflected on the low citric acid concentrations.

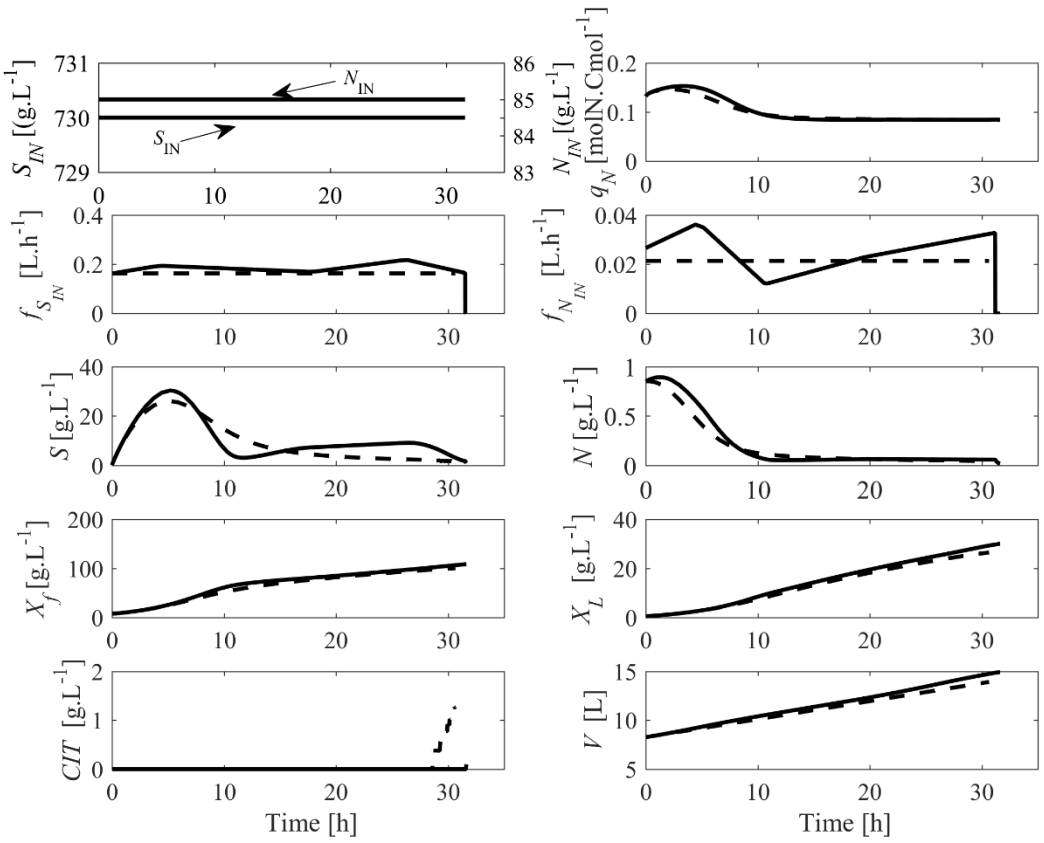


Figure III.24. Results of the control strategy. (▯▯▯) Constant flow rates (——) 2 PWL (——) 5 PWL.

The summary of the results of the control strategy are displayed on Table III.15 to clarify the enforcement of this strategy. Another case was also studied with two piecewise linear functions ($n_f = 2$) to analyze the impact of the quantity of PWL functions (See supplementary material B). Table IV shows that the three compared cases reported high biomass content, which is important for harvesting the cells at the end of the culture. The use of piecewise linear functions increased 1.1 times productivity and 1.04 times lipid content fraction with respect to an operation at constant flow rates. These results suggested that a constant flow rate could be used, but should be avoided if citric acid production is to be maintained at low values. The performance of 2 PWL functions was slightly better to the constant flow rates, which supports the use of a higher number of piecewise linear function for optimizing both objectives.

Table III.15. Performance of the control strategy

	t_f	$V(t_f)$	$CIT(t_f)$	$X_L(t_f)$	$X_f(t_f)$	$P_L(t_f)$	$w_L(t_f)$	$w_P.P_L + P_W$
	(h)	**(L)**	**(g.L^{-1})**	**(g.L^{-1})**	**(g.L^{-1})**	**(g.L^{-1}.h^{-1})**	**(g.g^{-1})**	
Constant flow rate	30.64	13.91	1.266	26.41	100.22	0.861	0.218	0.821
2 PWL functions	30.15	14.46	0	27.78	105.59	0.921	0.218	0.836
5 PWL functions	31.62	14.91	0.098	29.98	108.31	0.948	0.227	0.867

Literature values of lipid content fraction and volumetric productivity were reported from 0.17 to 0.25 g.g^{-1} and from 0.08 to 0.27 g.L^{-1}.h^{-1}, respectively (Aggelis and Komaitis, 1999; Beopoulos et al., 2009a; Cescut, 2009; Ochoa-Estopier, 2012). Even if our results are encouraging for being competitive with the literature results, the real implementation of the developed control strategy is needed to have a fair comparison. For that, the development of software sensors, which are challenging and time demanding, is required to reconstruct the state variables of the model. This will allow the proper monitoring of system and the correction of model uncertainties.

IV.5. Conclusions

An unstructured model based on Monod and inhibition kinetics and a model based on Droop quotas were developed to describe the accumulation of lipids and the production of citric acid on *Yarrowia lipolytica* on fed-batch and sequential batch cultures. The results of both models showed good agreement with the calibration and validation experimental data sets. Parameters from both models were consistent with literature values. Even if both models presented similar goodness-of-fit, the quota model was selected to develop a model based optimization of lipid accumulation due to a better description of residual glucose and extracellular nitrogen concentrations, which was important for the accurate manipulation of the N/C ratio. Two objectives were maximized: lipid productivity and lipid content fraction. Optimization results indicated that it would be possible to produce higher values than the literature proving the effectiveness of the control strategy. Further work need to deal with the experimental implementation to validate the simulation results of the presented control strategy.

IV.6. Supplementary Material A

Figures of the Internal Quota for the different cultures.

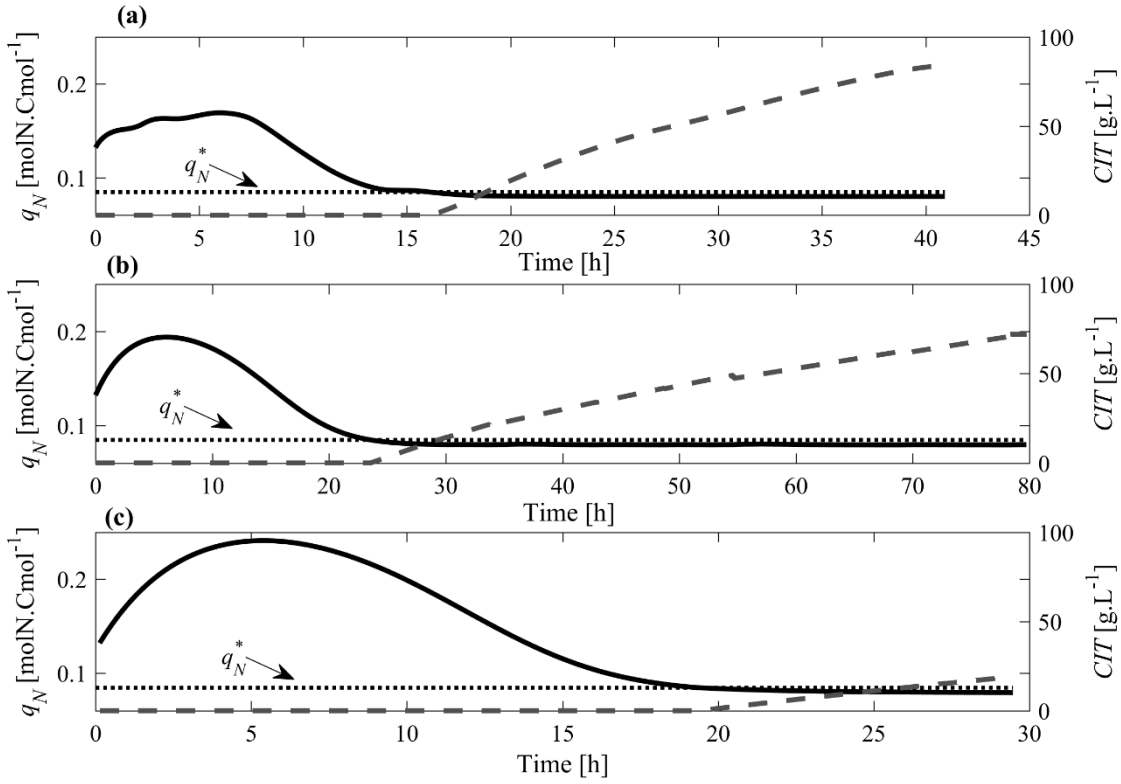


Figure III.25. Impact of the dynamics of the nitrogen quota along the cultures on the time evolution of citric acid concentration. (a) Culture A. (b) Culture B. (c) Culture C. The solid line represents the nitrogen quota. The dashed line indicates citric acid concentration. The arrows and the dotted line show the critical value of the nitrogen quota (q_N*) corresponding to the start of the overflow to citric acid production.

IV.7. Supplementary Material B

Information on optimization results.

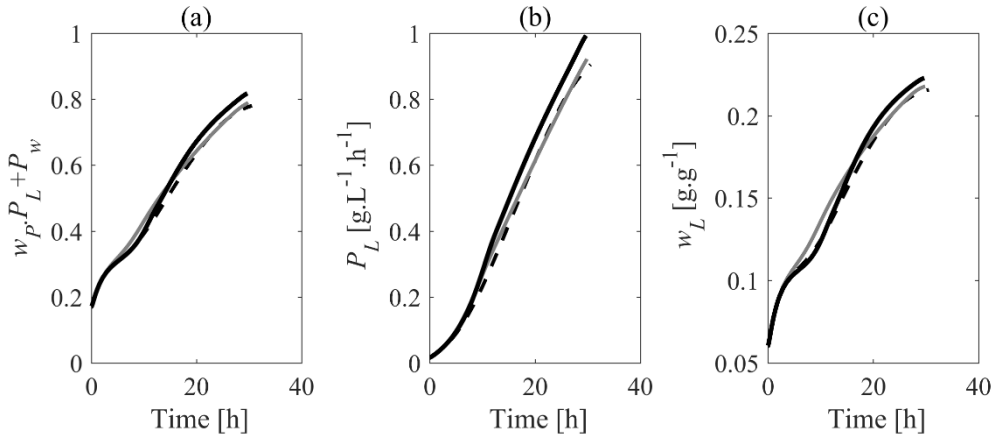


Figure III.26. Evolution of the objective functions along the fermentation time. (a) General cost function. (b) Lipid productivity. (c) Lipid content fraction. (— –) Constant flow rates (——) 2 PWL (——) 5 PWL

Figure III.26 shows the cost objective function for the three cases, which was defined as,

$$\max_{f_{S_{IN}}, f_{N_{IN}}, t_f} \quad w_P.P_L + P_W$$

$$\text{Subject to: } \begin{cases} Eq\,(1-7) \\ Eq\,(14) \\ q_N(t_f) \geq q_N{}^*,\ S(t_f) < S_{\max},\ N(t_f) < N_{\max} \\ V_0 < V < V_{\max},\ t_{\min} \leq t_f \leq t_{\max}, \\ 0 \leq f_{S_{IN}} \leq f_{S_{IN}\max},\ 0 \leq f_{N_{IN}} \leq f_{N_{IN}\max},\ n_{PWL} = n_f \end{cases}$$

For this type of two-fold objective problems, it is useful to see the objectives separately. Figure III.26 b-c depicts the dynamics of the two objectives: lipid productivity (P_L), and lipid content fraction ($w_L(t)$). It is observed that for all cases lipid content is maximized with a higher impact until 8 h, when the optimization switches to favor lipid productivity. This was expected as the penalization parameter was selected to favor lipid content fraction.

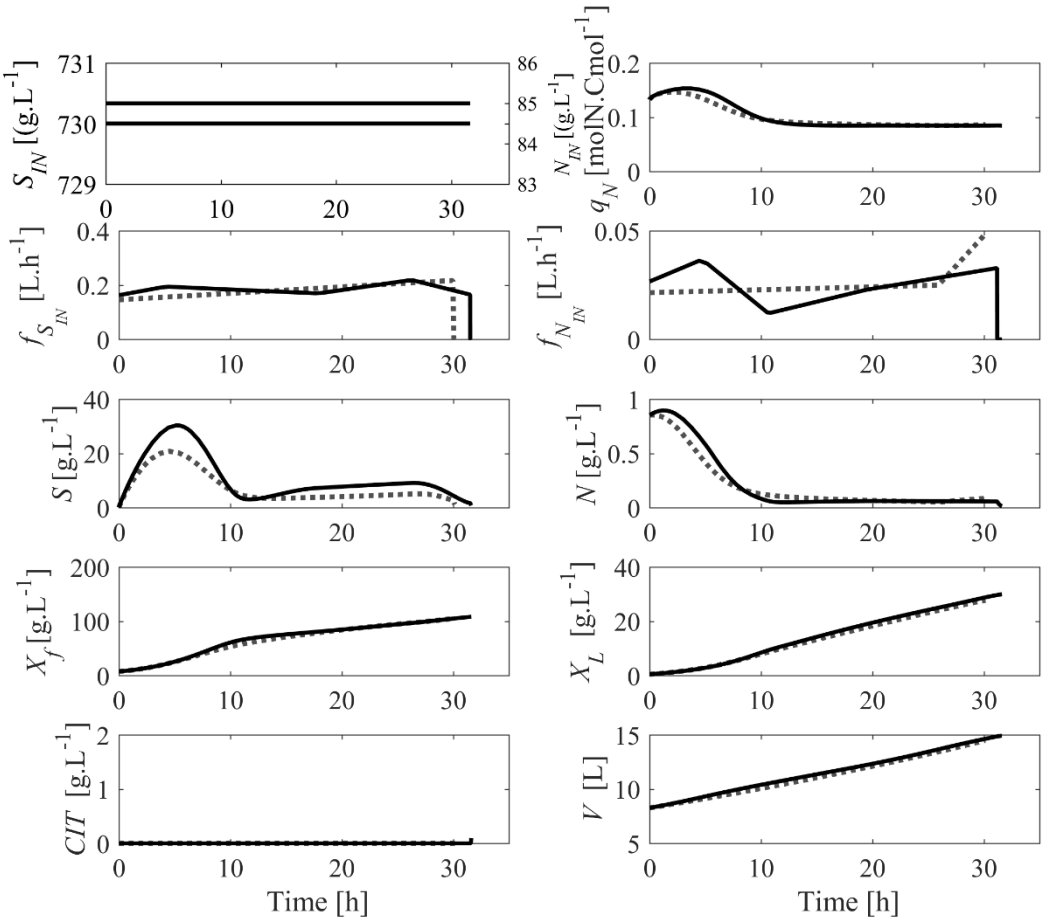


Figure III.27. Results of the control strategy. (▯▯▯) 2 PWL (——) 5 PWL.

The optimization with 2 PWL functions (gray dotted lines in Figure III.27) suggested a constant increase of glucose flow rate for more than 30 h, with a further decrease at the end of the fermentation time to attain the final glucose concentration constraint. Contrary, the nitrogen profile was maintained almost constant for 27 hours with an increase at the end to modulate the N/C ratio.

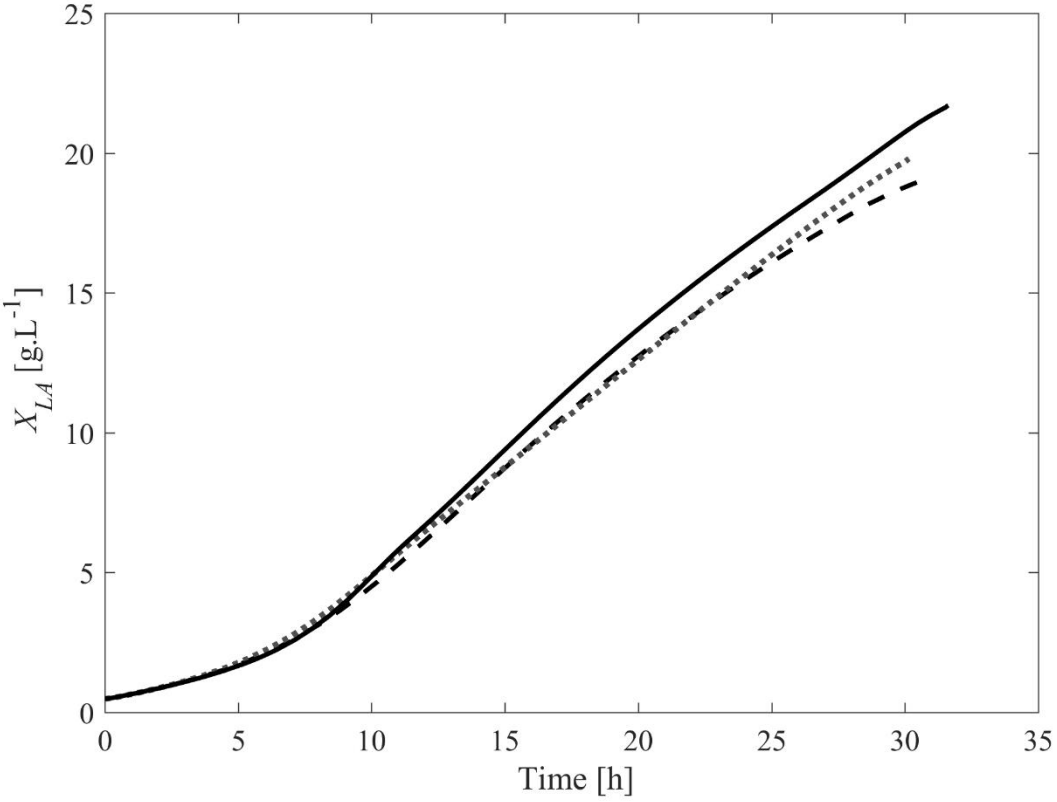


Figure III.28. Evolution accumulated lipids along the optimization. (— –) Constant flow rates (□□□)2 PWL (——) 5 PWL

IV.8. Supplementary Material C

Initial conditions of Experiments for calibration and validation

Table III.16. Initial conditions for the cultures A, B, and C.

Culture	S_0 (g.L^{-1})	N_0 (g.L^{-1})	X_L (t_0) (g.L^{-1})	CIT (t_0) (g.L^{-1})	X_f (t_0) (g.L^{-1})	X_T (t_0) (g.L^{-1})	q_N (t_0) (gN.gX_T^{-1})	V(t_0) (g.L^{-1})
A	0.02	0.853	0.48	0	8.10	7.64	0.075	8.26
B	98.34	0.885	0.028	0	0.47	0.47	0.075	4.02
C	100.95	1.4	0.047	0	0.792	0.792	0.075	3.03

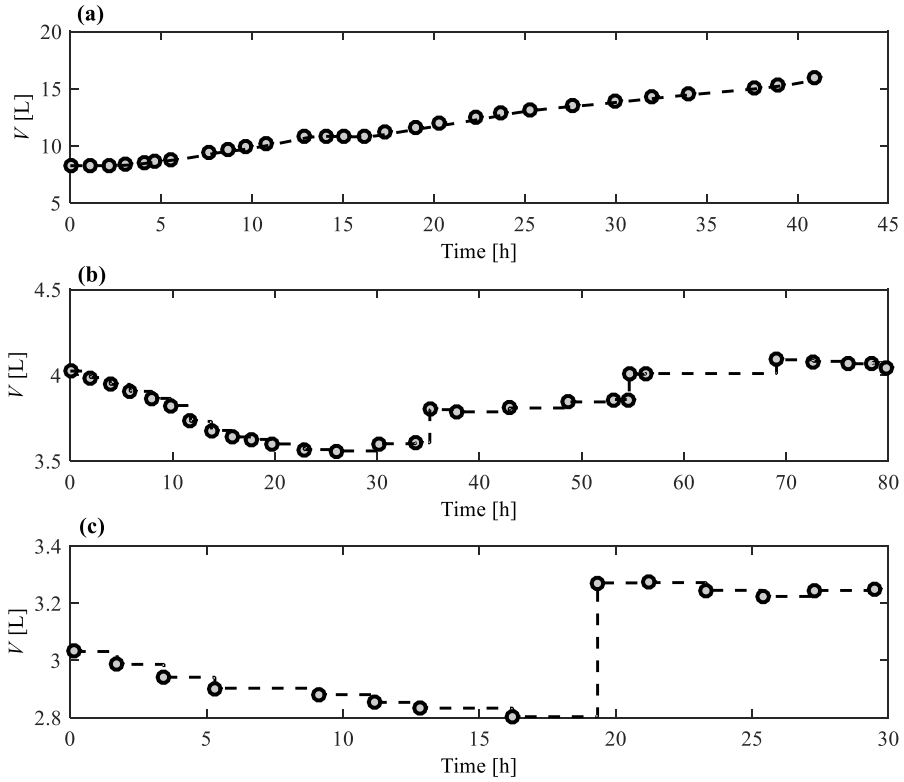


Figure III.29. Volume evolution along cultures (a) Culture A. (b) Culture B. (c) Culture C.

Table III.17. Experimental data from the literature in Batch conditions with glucose as substrate.

S_0 ($g.L^{-1}$)	N_0 ($g.L^{-1}$)	X_L (t_f) ($g.L^{-1}$)	CIT (t_f) ($g.L^{-1}$)	X_T (t_f) ($g.L^{-1}$)	Strain	Reference
34	0.14	0.5	10.5	6.1	ACA-DC-50109	(Papanikolaou et al., 2006)
42	0.141	0.52	15	5.9	ACA-DC-50109	(Papanikolaou et al., 2006)
52	0.142	0.6	20.1	7.1	ACA-DC-50109	(Papanikolaou et al., 2006)
149.5	0.139	0.7	42.9	7.2	ACA-DC-50109	(Papanikolaou et al., 2006)
65	0.25	0.52	27.6	6.5	ACA-DC-50109	(Papanikolaou et al., 2008)
20	0.1	0.73	4.8	2.85	H222	(Kavšček et al., 2015)
24.8	0.11	0.26	11.1	5.1	ACA-YC-5033	(Papanikolaou et al., 2009)
23.9	0.112	0.28	14.5	4.9	ACA-YC-5033	(Papanikolaou et al., 2009)
29.2	0.136	0.23	18	5.8	W29	(Papanikolaou et al., 2009)
29.7	0.135	0.3	17.2	5.9	JMY798	(Papanikolaou et al., 2009)
29.5	0.135	0.3	17.1	5.9	JMY794	(Papanikolaou et al., 2009)
58.8	0.137	0.44	38.2	5.5	ACA-YC-5033	(Papanikolaou et al., 2009)
50.8	0.118	0.53	13.2	3.9	ACA-YC-5033	(Papanikolaou et al., 2009)
57.7	0.108	0.4	49	5	W29	(Papanikolaou et al., 2009)
43.6	0.081	0.5	36.2	5.5	JMY798	(Papanikolaou et al., 2009)
45.5	0.085	0.59	32.3	5.3	JMY794	(Papanikolaou et al., 2009)

IV.9. Supplementary Material D

Assessment of the predictive capabilities of the quota model to represent the dynamics of the mutant strain *Yarrowia lipolytica* JMY3501

Two independent experiments performed in our lab were used to strengthen the predictive capabilities of our model for representing the dynamics of a mutant strain of *Y. lipolytica*.

The mutant strain identified as JMY3501 was obtained from the Laboratoire Microbiologie et Génétique Moléculaire, INRA, (Paris-Grignon, France). The construction of this strain was derived from the JMY1233 strain (Beopoulos *et al*., 2008). Lazar *et al*. (2014) give specific details of the construction which was performed in three steps. (i) TGL4 was inactivated by

introducing the disruption cassette tgl4::URA3ex from JMP1364 (Dulermo *et al*., 2013), which generated JMY2179. (ii) An excisable auxotrophic marker, URA3ex, was then excised from JMY2179 using JMP547 (Fickers *et al*., 2003), which generated JMY3122. (iii) Successive introductions of pTEF-DGA2-LEU2ex from JMP1822, and pTEF-GPD1-URA3ex from JMP1128 (Dulermo and Nicaud, 2011), into JMY3122 generated JMY3501. JMP1822 was obtained by replacing the URA3ex marker from JMP1132 (Beopoulos *et al*., 2008). The JMY3501 strain showed a strong improvement in production performances on glucose in terms of lipid content (increase from 18 to 55%), lipid yield and by-product formation (decrease in citric acid yield) (Sagnak *et al*., 2017).

Culture D corresponds to a sequential batch culture under nitrogen starvation. This culture was obtained from a 5 L stirred tank bioreactor (Biostat B. Braun Biotech International, Sartorius AG, Germany) with the acquisition software MFCS/win 2.1®. The inoculation volume was 300 mL. The dissolved oxygen was maintained above 20% of saturation. The working volume of the reactor was 4 L, where the operating conditions were set to a controlled temperature of 28°C and a pH value of 5.6 regulated by a 5 M KOH solution. The carbon source (glucose) was fed by two pulses at 32 h and 62 h. In this case, 5.2 $g.L^{-1}$ of $(NH_4)_2SO_4$ and 100 $g.L^{-1}$ of glucose were added to the initial medium.

Culture E is also a sequential batch culture under nitrogen starvation, where the carbon source (glucose) was fed by one pulse at 53 h. The initial medium contained 2 $g.L^{-1}$ of $(NH_4)_2SO_4$ and 88 $g.L^{-1}$ of glucose. The working volume and the temperature were the same as for culture D. The pH for this culture was maintain around 4.5.

Both cultures were used for calibration. The obtained parameters are presented on Table III.18, and the calibration results are displayed on Figure III.30. The term α in equation (17) that represents the maximum lipid content, was here considered at 0.55 $gL.gX_T^{-1}$ for *Yarrowia lipolytica* JMY3501. Although the model does not account for pH effects, the simulations show good agreement for both experimental culture studies. Glucose is consumed in a lower rate when pH is lower. It is interesting to note the usefulness of mathematical modelling for assessing quantitatively the metabolic differences between the wild type and the mutant strains *via* the analysis of the specific model parameters. For instance, the maximum specific lipid production rate was higher as lipid contents increased for this strain, whereas the maximum specific citric acid production rate was lower as less citric acid was produced.

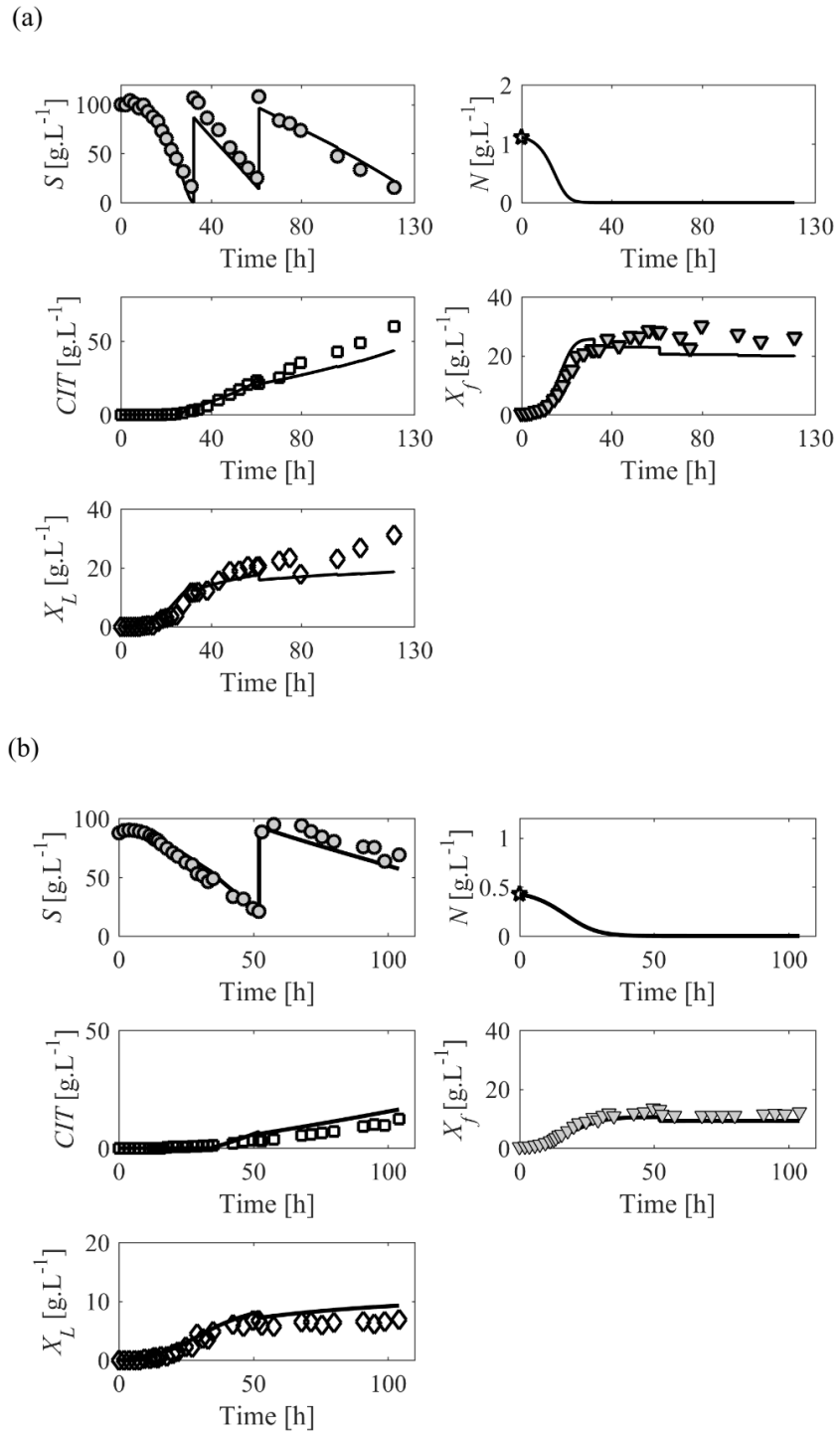


Figure III.30. Model calibration: Performance of the quota model with the experimental data of (a) Culture D and (b) Culture E.

Table III.18. Estimated parameters for different strains of *Yarrowia lipolytica*

Parameter	This work	
	Quota W29	Quota JMY3501
μ^* [h^{-1}]	0.755	0.855
π_{LIP}^{max} [g*LIP*.(gX_f-h)$^{-1}$]	0.039	0.109
π_{CIT}^{max} [g*CIT*.(gX_f-h)$^{-1}$]	0.314	0.118
$Y_{X/S}$ [gX_f.gS^{-1}]	0.519	0.32
$Y_{LIP/S}$ [g*LIP*.gS^{-1}]	0.272	0.17
$Y_{CIT/S}$ [g*CIT*.gS^{-1}]	0.89	0.44
$K_{S,1}$ [g*S*.L^{-1}]	1.116	1.204
$K_{S,2}$ [g*S*.L^{-1}]	1.482	2.531
$K_{S,3}$ [g*S*.L^{-1}]	12.22	11.24
K_N [g*N*.L^{-1}]	23.08	51.59
k_2 [g*CIT*.L^{-1}]	9.75	9.454
k_3 [g*CIT*.L^{-1}]	37.63	56.06
$k_{I,1}$ [g*S*.L^{-1}]	83.64	108.90
$k_{I,2}$ [g*S*.L^{-1}]	44.17	65.80
$k_{I,3}$ [g*S*.L^{-1}]	40.11	21.79
Quotas:		
ρ_N^{max} [g*N*.(gX_f-h)$^{-1}$]	0.716	0.94
Q_0 [g*N*.(gX_T)$^{-1}$]	0.045	0.044
q_N* [mol*N*.Cmol^{-1}]	0.084	0.080

IV.10. References

Abdollahi, J., Dubljevic, S., 2012. Lipid production optimization and optimal control of heterotrophic microalgae fed-batch bioreactor. Chem. Eng. Sci. 84, 619–627. doi:10.1016/j.ces.2012.09.005

Adesanya, V.O., Davey, M.P., Scott, S.A., Smith, A.G., 2014. Kinetic modelling of growth and storage molecule production in microalgae under mixotrophic and autotrophic conditions. Bioresour. Technol. 157, 293–304. doi:10.1016/j.biortech.2014.01.032

Aggelis, G., Komaitis, M., 1999. Enhancement of single cell oil production by Yarrowia lipolytica growing in the presence of Teucrium polium L. aqueous extract. Biotechnol. Lett. 21, 747–749. doi:10.1023/A:1005591127592

Amribt, Z., Niu, H., Bogaerts, P., 2013. Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. Biochem. Eng. J. 70, 196–

209. doi:10.1016/j.bej.2012.11.005

Arzumanov, T.T.E., Sidorov, I.I.A., Shishkanova, N.V.N., Finogenova, T.T. V., 2000. Mathematical modeling of citric acid production by repeated batch culture. Enzyme Microb. Technol. 26, 826–833. doi:http://dx.doi.org/10.1016/S0141-0229(00)00178-2

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol. Rev. 19, 219–37.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M.-T., Hapala, I., Papanikolaou, S., Chardot, T., Nicaud, J.-M., 2008. Control of lipid accumulation in the yeast Yarrowia lipolytica. Appl. Environ. Microbiol. 74, 7779–89. doi:10.1128/AEM.01412-08

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Cescut, J., Severac, E., Molina-Jouve, C., Uribelarrea, J.L., 2011. Optimizing pressurized liquid extraction of microbial lipids using the response surface method. J. Chromatogr. A 1218, 373–379. doi:10.1016/j.chroma.2010.12.003

De la Hoz Siegler, H., Ben-Zvi, A., Burrell, R.E., McCaffrey, W.C., 2011. The dynamics of heterotrophic algal cultures. Bioresour. Technol. 102, 5764–74. doi:10.1016/j.biortech.2011.01.081

Droop, M.R., 1968. Vitamin B12 and marine ecology. The kinetics of uptake, growth and inhibition in monochrysis Lutheri. J. Mar. Assoc. 689–733.

Dulermo, T., Nicaud, J.-M., 2011. Involvement of the G3P shuttle and β-oxidation pathway in the control of TAG synthesis and lipid accumulation in Yarrowia lipolytica. Metab. Eng. 13, 482–491. doi:10.1016/j.ymben.2011.05.002

Dulermo, T., Tréton, B., Beopoulos, A., Kabran Gnankon, A.P., Haddouche, R., Nicaud, J.-M., 2013. Characterization of the two intracellular lipases of Y. lipolytica encoded by TGL3 and TGL4 genes: New insights into the role of intracellular lipases and lipid body organisation. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1831, 1486–1495. doi:10.1016/j.bbalip.2013.07.001

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. MHS'95. Proc. Sixth Int. Symp. Micro Mach. Hum. Sci. 39–43. doi:10.1109/MHS.1995.494215

Economou, C.N.C.N., Aggelis, G., Pavlou, S., Vayenas, D.V. V., 2011. Modeling of single-cell

oil production under nitrogen-limited and substrate inhibition conditions. Biotechnol. Bioeng. 108, 1049–55. doi:10.1002/bit.23026

Fakas, S., Stamatina, M., Aggelis, G., 2009. Single cell oil and gamma-linolenic acid production by Thamnidium elegans on raw glycerol, in: NY: Nova Science Publishers (Ed.), Aggelis G, Editors. Microbial Conversions of Raw Glycerol. pp. 85–100.

Fickers, P., Le Dall, M.-T., Gaillardin, C., Thonart, P., Nicaud, J.-M., 2003. New disruption cassettes for rapid gene disruption and marker rescue in the yeast Yarrowia lipolytica. J. Microbiol. Methods 55, 727–737. doi:10.1016/J.MIMET.2003.07.003

Kavšček, M., Bhutada, G., Madl, T., Natter, K., 2015. Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. BMC Syst. Biol. 9, 72. doi:10.1186/s12918-015-0217-4

Lazar, Z., Dulermo, T., Neuveglise, C., Crutz-Le Coq, A.-M., Nicaud, J.-M., 2014. Hexokinase—A limiting factor in lipid production from fructose in Yarrowia lipolytica. Metab. Eng. 26, 89–99. doi:10.1016/J.YMBEN.2014.09.008

Lemesle, V., Mailleret, L., 2008. A Mechanistic Investigation of the Algae Growth "Droop" Model. Acta Biotheor. 56, 87–102. doi:10.1007/s10441-008-9031-3

Lewis, R.M., Torczon, V., 1996. Pattern Search algorithms for bound constrained minimization. SIAM J. Optim.

Mairet, F., Bernard, O., Masci, P., Lacour, T., Sciandra, A., 2011. Modelling neutral lipid production by the microalga Isochrysis aff. galbana under nitrogen limitation. Bioresour. Technol. 102, 142–9. doi:10.1016/j.biortech.2010.06.138

Meeuwse, P., Akbari, P., Tramper, J., Rinzema, A., 2012. Modeling growth, lipid accumulation and lipid turnover in submerged batch cultures of Umbelopsis isabellina. Bioprocess Biosyst. Eng. 35, 591–603. doi:10.1007/s00449-011-0632-x

Meeuwse, P., Tramper, J., Rinzema, A., 2011. Modeling lipid accumulation in oleaginous fungi in chemostat cultures: I. Development and validation of a chemostat model for Umbelopsis isabellina. Bioprocess Biosyst. Eng. 34, 939–49. doi:10.1007/s00449-011-0545-8

Nicaud, J.-M., Madzak, C., Broek, P., Gysler, C., Duboc, P., Niederberger, P., Gaillardin, C., 2002. Protein expression and secretion in the yeast Yarrowia lipolytica. FEMS Yeast Res. 2, 371–379. doi:10.1111/j.1567-1364.2002.tb00106.x

Nicaud, J.M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.

Ochoa-Estopier, A., 2012. Analyses systématique des bascules métaboliques chez les levures d'intérêt industriel: application aux bascules du métabolisme lipidique chez Yarrowia lipolytica. Institut National des Sciences Appliquées de Toulouse.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Papanikolaou, S., Aggelis, G., 2011a. Lipids of oleaginous yeasts. Part II: Technology and potential applications. Eur. J. Lipid Sci. Technol. 113, 1052–1073. doi:10.1002/ejlt.201100015

Papanikolaou, S., Aggelis, G., 2011b. Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. Eur. J. Lipid Sci. Technol. 113, 1031–1051. doi:10.1002/ejlt.201100014

Papanikolaou, S., Aggelis, G., 2009. Biotechnological valorization of biodiesel derived glycerol waste through production of single cell oil and citric acid by Yarrowia lipolytica. Lipid Technol. 21, 83–87. doi:10.1002/lite.200900017

Papanikolaou, S., Aggelis, G., 2003a. Modelling aspects of the biotechnological valorization of raw glycerol: production of citric acid by Yarrowia lipolytica and 1,3-propanediol by Clostridium butyricum. J. Chem. Technol. Biotechnol. 78, 542–547. doi:10.1002/jctb.831

Papanikolaou, S., Aggelis, G., 2003b. Modeling lipid accumulation and degradation in Yarrowia lipolytica cultivated on industrial fats. Curr. Microbiol. 46, 398–402. doi:10.1007/s00284-002-3907-2

Papanikolaou, S., Aggelis, G., 2002. Lipid production by Yarrowia lipolytica growing on industrial glycerol in a single-stage continuous culture. Bioresour. Technol. 82, 43–49.

Papanikolaou, S., Chatzifragkou, A., Fakas, S., Galiotou-Panayotou, M., Komaitis, M., Nicaud, J.-M.M., Aggelis, G., 2009. Biosynthesis of lipids and organic acids by Yarrowia lipolytica strains cultivated on glucose. Eur. J. Lipid Sci. Technol. 111, 1221–1232. doi:10.1002/ejlt.200900055

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Papanikolaou, S., Galiotou-Panayotou, M., Fakas, S., Komaitis, M., Aggelis, G., 2008. Citric acid production by Yarrowia lipolytica cultivated on olive-mill wastewater-based media. Bioresour. Technol. 99, 2419–2428. doi:10.1016/j.biortech.2007.05.005

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., Nicaud, J.M., 2015. Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. Biotechnol. Biofuels 8, 104. doi:10.1186/s13068-015-0286-z

Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. Biochimie 86, 807–15. doi:10.1016/j.biochi.2004.09.017

Ratledge, C., Wynn, J.P., 2002. The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. Adv. Appl. Microbiol. 51, 1–51.

Renard, F., Vande Wouwer, A., 2008. Robust adaptive control of yeast fed-batch cultures. Comput. Chem. Eng. 32, 1238–1248. doi:10.1016/j.compchemeng.2007.05.008

Robles-Rodriguez, C.E., Bideaux, C., Guillouet, S.E., Gorret, N., Cescut, J., Uribelarrea, J.-. L., Roux, G., Molina-Jouve, C., Aceves-Lara, C.., 2017. Dynamic metabolic modeling of lipid accumulation and citric acid prodiction by Yarrowia lipolytica. Comput. Chem. Eng. 100, 139–152. doi:10.1016/j.compchemeng.2017.02.013

Sagnak, R., Cochot, S., Molina-Jouve, C., Nicaud, J.-M., Guillouet, S.E., 2017. Modulation of the Glycerol Phosphate availability led to concomitant reduction in the citric acid excretion and increase in lipid content and yield in Yarrowia lipolytica. J. Biotechnol. doi:10.1016/j.jbiotec.2017.11.001

Santos, L.O., Dewasme, L., Coutinho, D., Wouwer, A. Vande, 2012. Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. Comput. Chem. Eng. 39, 143–151. doi:10.1016/j.compchemeng.2011.12.010

Sitepu, I.R., Garay, L.A., Sestric, R., Levin, D., Block, D.E., German, J.B., Boundy-Mills, K.L., 2014. Oleaginous yeasts for biodiesel: current and future trends in biology and production. Biotechnol. Adv. 32, 1336–60. doi:10.1016/j.biotechadv.2014.08.003

Surisetty, K., De la Hoz Siegler, H., McCaffrey, W.C., Ben-Zvi, A., 2010a. Robust modeling of a microalgal heterotrophic fed-batch bioreactor. Chem. Eng. Sci. 65, 5402–5410. doi:10.1016/j.ces.2010.06.008

Surisetty, K., Hoz Siegler, H.D. la, McCaffrey, W.C., Ben-Zvi, A., 2010b. Model re-parameterization and output prediction for a bioreactor system. Chem. Eng. Sci. 65, 4535–4547. doi:10.1016/j.ces.2010.04.024

Valentinotti, S., Srinivasan, B., Holmberg, U., Bonvin, D., Cannizzaro, C., Rhiel, M., von Stockar, U., 2003. Optimal operation of fed-batch fermentations via adaptive control of overflow metabolite. Control Eng. Pract. 11, 665–674. doi:10.1016/S0967-0661(02)00172-7

Ykema, A., Verbree, E.C., Van Verseveld, H.W., Smit, H., 1986. Mathematical modelling of lipid production by oleaginous yeasts in continuous cultures. Antonie Van Leeuwenhoek 52, 491–506.

Yoo, S.J., Kim, J.H., Lee, J.M., 2014. Dynamic modelling of mixotrophic microalgal photobioreactor systems with time-varying yield coefficient for the lipid consumption. Bioresour. Technol. 162, 228–35. doi:10.1016/j.biortech.2014.03.128

V. MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION (MOPSO) OF LIPID ACCUMULATION IN FED-BATCH CULTURES

Carlos Eduardo Robles-Rodriguez[1], Carine Bideaux[1], Stéphane E. Guillouet[1], Nathalie Gorret[1], Carole Molina-Jouve[1], Gilles Roux[2], César Arturo Aceves-Lara[1*]

[1] LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

[2] LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France

Accepted article for Oral presentation on the 24th Mediterranean Conference on Control and Automation. (MED 2016). June 21 – 24, 2016. Athens, Greece.

Abstract

Dynamic optimization of fermentation processes could demand the use of multiple criteria to attain certain objectives, which in most cases are conflicting to each other. The use of Pareto optimal sets supplies the necessary information to take decisions about the trade-offs between objectives. In this work, a multi-objective optimization algorithm based on particle swarm optimization (MOPSO) is used to optimize lipid contents in fermentations with *Yarrowia lipolytica*. A reduced model was developed to shorten the computation time of MOPSO. A pattern search algorithm was sequentially coupled to MOPSO to execute a dynamic optimization handling physical constraints. Three cases are analyzed to emphasize the response of our control strategy. Simulation results showed that MOPSO – pattern search algorithm achieved high lipid fraction and productivity.

Keywords: Multi-objective optimization, Particle swarm optimization, Pattern search, Fed-batch fermentation, Lipid accumulation.

V.1.Introduction

Fed-batch cultures are of special interest for the biotechnological production of value-added goods. The manipulation of feeding profiles provides information about the microbial kinetics and metabolites production. However, the optimization and control of fed-batch processes is a dynamic problem where the feeding rates have strong impact on yields, liquid volume, productivity, and fermentation times. For example, low feeding rates can cause starvation, whereas an overfeeding can lead to substrate inhibition and/or formation of undesirable products (Jin *et al*., 2014).

The design of optimal feeding strategies can be determined by solving model-based optimization problems. Several control strategies have been described and applied for single objective optimization including optimal control (Riascos and Pinto, 2004), adaptive control (Smets *et al*., 2004), and model predictive control (Ashoori *et al*., 2009). Nevertheless, most of the engineering problems have multiple conflicting objectives, *e.g.* minimize cost, maximize performance, maximize reliability, minimize fermentation time, *etc*.

In the case of multi-objective optimization (MO), two general approaches have been presented (Marler and Arora, 2004). The first approach relies on the combination of the objective function into a single function by arithmetic operations. The most common approach is the weighted sum method. This method assigns a weight to each objective function. Its principal issue lies on the proper selection of the weights which is complicated in practice. Moreover, each combination of weights will generate a single arbitrary solution with a constant trade-off between objectives. The second approach considers that finding a unique solution is neither possible nor desirable. The solution of a MO is a set of points that represent the best trade-offs between the objective functions. This optimal solution set is considered to be a Pareto optimal set, where the solutions are non-dominated to each other (Coello Coello, 1999). These optimal sets are characterized by trade-offs and, thus, there is a multitude of Pareto optimal solutions (Coello *et al*., 2004). Therefore, finding the largest allowed number of Pareto optimal solutions is highly desirable. The algorithms to resolve these problems have to be able to provide more than one solution. The most common algorithms are the ones based on population, such as Evolutionary Algorithms (EA) (Zitzler *et al*., 2000), where the Non-dominated sorting genetic algorithm (NSGA-II), and the Pareto Archived Evolutionary Strategy (PAES) have been

successfully applied (Konak *et al*., 2006). However, both present some disadvantages. The first one is computationally expensive, and the crowding distance works in objective space only (Deb *et al*., 2002). Otherwise, the second one is a local method which performance depends on population sizes (Knowles and Corne, 2000).

Other alternative to treat multi-objective optimization is through the implementation of Particle Swarm Optimization (PSO) (Coello-Coello et al., 2002). This algorithm is simple in application and quick in convergence. A compliance of application in different fields can be found in (Lalwani *et al*., 2013). MOPSO differs from EAs in the use of leaders to guide the search instead of individuals within the search space. Furthermore, MOPSO sets the velocity of a particle by a unidirectional mutation operation, whereas EA have an omnidirectional mutation. Notwithstanding, MOPSO is not able to handle constraints.

Therefore, this work proposes a static multi-objective optimization by PSO (MOPSO) coupled to a dynamic optimization by pattern search. The method is applied to optimize lipid accumulation of *Yarrowia lipolytica* from glucose, where citric acid production hampers lipid production. Section V.2 describes the model for lipid accumulation, which is further reduced to lower the time of optimization in MOPSO. The algorithms for optimizing single and multi-objective problems are explained in section V.3. Section V.4expresses the dynamic model-based multi-objective control strategy. The simulation results and the performance evaluation are shown in Section V.5. Finally, conclusions and perspectives end this paper.

V.2.Modeling Lipid Accumulation

V.2.1. Dynamic metabolic model for lipid accumulation

A detailed dynamic metabolic model for lipid accumulation in fed-batch culture by *Yarrowia lipolytica* from glucose was presented by (Robles-Rodriguez *et al*., 2016) and rewritten as,

$$\frac{dC}{dt} = S_C.Z.r_M.X - \frac{C}{V}\frac{dV}{dt} + \frac{F_{IN}}{V} \quad \text{(III.63)}$$

$$\frac{dLIP}{dt} = (S_{LIP} + \gamma.S_X).Z.r_M.X - \frac{LIP}{V}\frac{dV}{dt} - \gamma.\frac{X_{Cat}}{V}\frac{dV}{dt} \quad \text{(III.64)}$$

$$\frac{dX}{dt} = (S_X + S_{LIP}).Z.r_M.X - \frac{X}{V}\frac{dV}{dt} \quad \text{(III.65)}$$

$$\frac{dX_{Cat}}{dt} = S_X Z.r_M X - \frac{X_{Cat}}{V}\frac{dV}{dt} \quad \text{(III.66)}$$

$$\frac{dV}{dt} = F_{IN} - F_{OUT} \quad \text{(III.67)}$$

where S_C and S_{LIP} are the n_C *x* n_r and n_{LIP} *x* n_r stoichiometric matrices for external species concentration $C = [G, N, CIT]^T$, and internal metabolites $[LIP]$, respectively. Variables C and LIP are expressed in $(g.L^{-1})$, with exception of N in $(mol.L^{-1})$. The inlet mass vector $F_{IN} = [qG_{IN}.G_{IN}, qN_{IN}.N_{IN}, 0]^T$ considers the input flow rates qG_{IN} and qN_{IN}; and their respective concentrations G_{IN} and N_{IN}. The vector r_M represents the n_r intracellular and exchange rates expressed per gram of total biomass X. Equation (III.64)assumes that total lipids LIP are the sum of accumulated lipids and the structural lipids that are intrinsic to the catalytic biomass X_{Cat} in a constant fraction γ. The volume V in (III.67) holds the output flow rate, F_{OUT}, to deal with sampling volume.

V.2.2. Model Reduction

The calculation of a Pareto optimal set is computationally expensive when the integration of the model must be performed for every particle. Therefore, a reduced model was proposed to diminish this computation time by using the methodology proposed by Cassagranda *et al,* (2015) (Casagranda *et al*., 2015).

Bioprocess models can be decomposed in linear combinations of processes (kinetics) that evolve dynamically over time. Consequently and following (Casagranda *et al*., 2015), the model (III.64) - (III.67) can be rewritten as follows,

$$\frac{dx_i}{dt} = \sum_j f_{i,j}(x(t)) \quad \text{(III.68)}$$

where $f_{i,j}$ represents the j^{th} process involved in the evolution of the i^{th} state variable. In the sake simplicity, only one state variable will be used here to illustrate the methodology. Let us consider the total biomass (X) concentration, which from model correspond to x_5. The first term in (III.65) contains eight processes regarding to the description of the model (Robles-Rodriguez *et al*., 2016):

$$\frac{dX}{dt} = (S_X + S_{LIP}).Z.r_{M,1}.X + (S_X + S_{LIP}).Z.r_{M,2}.X + (S_X + S_{LIP}).Z.r_{M,3}.X + (S_X + S_{LIP}).Z.r_{M,4}.X + (S_X + S_{LIP}).Z.r_{M,5}.X + (S_X + S_{LIP}).Z.r_{M,6}.X + (S_X + S_{LIP}).Z.r_{M,7}.X + (S_X + S_{LIP}).Z.r_{M,8}.X - \frac{X}{V}\frac{dV}{dt} \quad \text{(III.69)}$$

According to (III.68), equation (III.69) can be represented in an equivalent form as,

$$\frac{dx_5}{dt} = f_{5,1} + f_{5,2} + f_{5,3} + f_{5,4} + f_{5,5} + f_{5,6} + f_{5,7} + f_{5,8} - f_{5,9} \quad \text{(III.70)}$$

where $x_5 = X$, $f_{5,1} = (S_X + S_{LIP}).Z.r_{M,1}.X, \ldots, f_{5,9} = \frac{X}{V}\frac{dV}{dt}$

A weighing factor was assigned to each $f_{i,j}$ to compare the influence of the different processes $f_{i,j}$, along the culture time for each state variable x_i. The weighing factor involves the absolute value of each $f_{i,j}$ and scales it with the sum of all the process functions.

$$W_{i,j}(t) = \frac{\left| f_{i,j}(x(t)) \right|}{\sum_j \left| f_{i,j}(x(t)) \right|} \quad \text{(III.71)}$$

where $0 \le W_{i,j}(t) \le 1$ and $\sum_j W_{i,j}(t) = 1$.

Afterwards, $W_{i,j}$ was evaluated to define if a process was active or inactive throughout the time. A threshold value of $\theta = 0.1$ was selected to determine whether a process $f_{i,j}$ was active or inactive (Casagranda *et al.*, 2015). This value was chosen to ensure that only the important processes remained. The processes with $W_{i,j}$ below θ for the complete time set were considered inactive, and they were thus erased from the model.

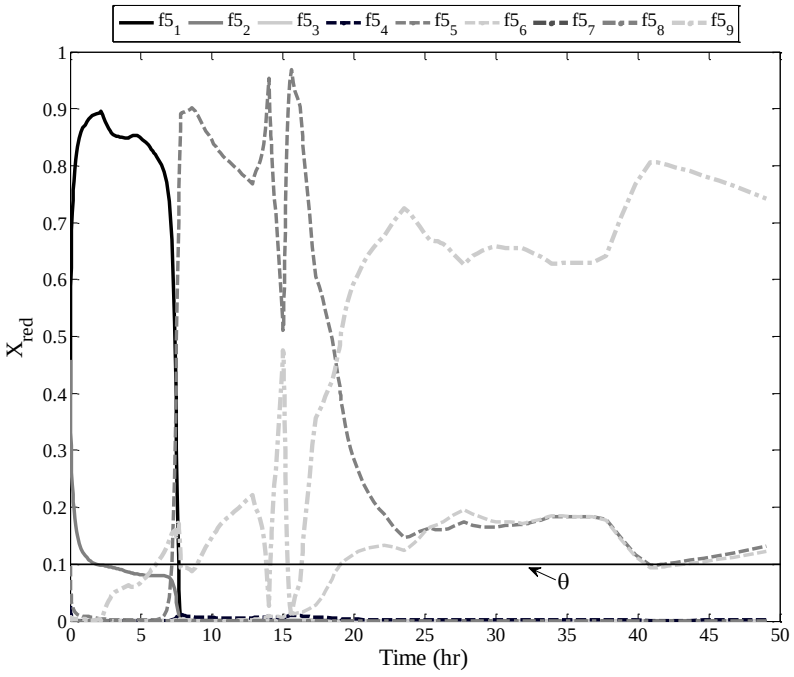


Figure III.31. Example of the methodology for model reduction

Figure III.31 displays the different $f_{i,j}$ processes for total biomass (X) and the threshold limit to perform this reduction. The 9 initial processes were reduced to 5 ($f_{5,1} = (S_X + S_{LIP}).Z.r_{M,1}.X$, $f_{5,2} = (S_X + S_{LIP}).Z.r_{M,2}.X$, $f_{5,5} = (S_X + S_{LIP}).Z.r_{M,5}.X$, $f_{5,6} = (S_X + S_{LIP}).Z.r_{M,6}.X$ and $f_{5,9} = \frac{X}{V}\frac{dV}{dt}$).

Once this procedure was applied to the entire state variables, the resulting reduced model can be expressed as,

$$\frac{dx_{i,red}}{dt} = \sum_j f_{i,j}\left(x(t)\right) \tag{III.72}$$

From the 74 processes identified in the complete model (III.68) only 48 processes remained in (III.72).

Table III.19. Comparison of Reduced and Complete model

Variable	WSSE	**Variable**	WSSE
Glucose	0.227	*Lipids*	0.008
Nitrogen	0.000	*Total Biomass*	0.057
Citrate	0.137	*Catalytic Biomass*	0.054

The reduced model (III.72) was compared with the results of the model (III.68) in a simulation for 45h and the same initial conditions. The Weighted Sum of Square Errors (WSSE) is displayed in Table I which prove the good performance of the reduction.

V.3.Optimization Methods

V.3.1. Multi-objective Optimization problem

A minimization multi-objective problem is defined as:

$$\min_x J(x) = [J_1(x), J_2(x), \ldots, J_k(x)]^T$$

$$\text{Subject to: } \begin{cases} g_i(x) \le 0 & i = 1,2,\ldots,m \\ h_i(x) = 0 & i = 1,2,\ldots,p \\ x_i^L \le x_i \le x_i^U \end{cases} \tag{III.73}$$

The solution of (11) is a vector $x = [x_1,x_2,\ldots,x_n]^T$ of k decision variables. $J(x)$ is the vector of k objective functions to be minimized with respect to the inequality $g_i(x)$, $i = 1,\ldots,m$, and the

equality $h_i(x)$, $i = 1,\ldots p$, constraints respectively. The values of x_i^U and x_i^L describe the lower and upper bounds of the solution vector x.

V.3.2. Pareto optimality and dominance

Given two feasible solution vectors $x, y \in R^k$, it is said that x dominates another solution y if and only if their respective objective functions are $J_i(x) \leq J_i(y)$ for $i = 1,\ldots,k$ and $J_i(x) < J_i(y)$ for at least one i^{th} objective function.

A solution is Pareto optimal if it is not dominated by any other solution in the solution space. The set of all non-dominated solutions is referred as to the Pareto optimal set, and for a given Pareto optimal set, the corresponding objective function values in the objective space are called the Pareto front.

V.3.3. Multi-objective optimization algorithm based on Particle Swarm Optimization (MOPSO)

MOPSO was first used by (Coello-Coello et al., 2002) to deal with multi-objective optimization problems. It focus on maximizing the number of elements of the Pareto optimal set; the minimization of the distance of the Pareto front, and the maximization the spread of the solutions (smooth and uniform distribution) (Reyes-Sierra and Coello Coello, 2006).

MOPSO (Konak *et al*., 2006) begins with the initialization of the swarm and the set of leaders in an external archive with the non-dominated particles from the swarm. Once these are set, the algorithm follow the next steps:

(i) Sort solutions to define a leader for each particle of the swarm;
(ii) Update position (flight);
(iii) Mutation to define the new directions of search;
(iv) Evaluation of the particle to find the corresponding *pbest;*
(v) Update *pbest* (best non-dominated positions for the particles). After all the particles are being updated,
(vi) The set of the leaders is also updated.

The process is repeated for a fixed number of iterations.

The sparseness of the solutions and the updating of the leaders are key components when designing MOPSO (Coello *et al.*, 2004). The spread of the solution is promoted by the inertia weight of PSO algorithm. This inertia weight includes a mutation operator to allow particles to perform a global search. In the other hand, the selection of the leaders assumes that every non-dominated solution is a new leader, and a new leader is chosen by the nearest neighbor density estimator to have an idea of the closest neighbors of a particle. The particle with the largest value is selected as the new leader (Reyes-Sierra and Coello Coello, 2006).

Nonetheless, the method requires the storage of the non-dominated solutions. In this context, the proposition of (Coello-Coello et al., 2002) included a repository or archive. The archive allows the entrance of a solution only if it is non-dominated with respect to the contents of the archive or if it dominates any of the solutions within the archive. Moreover, the repository has a limited size, and if it is full, new solutions are inserted based on the retention criterion, that is giving priority to solutions located in less crowded areas of the objective space. Some of the disadvantages of MOPSO are the static analysis of trade-offs and the handling of constraints.

V.3.4. Dynamic Optimization

The results of MOPSO are introduced in a pattern search algorithm to dynamically solve the optimization problem. The sequential coupling is explained in next section.

Pattern search algorithms are feasible point methods with global convergence (Lewis and Torczon, 1996). The search is based on a series of exploratory moves around the current iterate x_j to choose a new iterate $x_{j+1} = x_j + s_j$ for some feasible steps s_j. The selection of the steps $s_j = \Delta_j . P_j$ is determined by exploratory moves comprising a pattern P_j, and a step length Δ_j.

Given an initial guess $x_0 \epsilon R^n$, and an initial choice of step length Δ_0, the method is performed as follows:

(i) Compute $f(x_j)$;

(ii) Determine a step $s_j = \Delta_j . P_j$;

(iii) Calculate $\rho_j = f(x_j) - f(x_j + s_j)$; (iv) if $\rho_j > 0$, then $x_{j+1} = x_j + s_j$, otherwise $x_{j+1} = x_j$;

(iv) Update (Δ_j, P_j).

The updates of Δ_j aim to force $\rho_j > 0$. Hence, if $\rho_j \leq 0$, then $\Delta_{j+1} = \sigma.\Delta_j$; otherwise, $\Delta_{j+1} = \lambda.\Delta_j$ where $\sigma \epsilon (0,1)$ and $\lambda \geq 1$. The search for an optimum is successive until $f(x_j+sj) < f(x_j)$.

The optimization with this method was implemented in MATLAB® with the *patternsearch* function. The differential equations of the model were solved by ode113. It is worth noting that pattern search does not need the calculation of derivatives at each iterate x_j.

V.4.Definition of the Control Strategy

The proposed control strategy for optimizing lipid accumulation is divided in two main parts to attain several objectives. In a first time, the optimization problem is focused on maximizing the concentration of catalytic biomass (growth phase). In a second time, the optimization problem maximizes the storage lipids (production phase) with two criteria: productivity and fraction. In the control framework, the reduced model (III.72) is used to compute the Pareto sets, whereas the model (III.68) is employed to choose the control action. The manipulated variables are the input flow rates qG_{IN}, and qN_{IN} to manage the N/C ratios that trigger lipid accumulation and citric acid production (Ochoa-Estopier and Guillouet, 2014). The glucose input concentration G_{IN} is considered as a disturbed variable, whereas nitrogen concentration N_{IN} is maintained constant.

V.4.1. Growth optimization

The first step of the optimization aims to produce catalytic biomass (cells without storage lipids). A single objective is considered involving the maximization of the concentration of catalytic biomass by minimizing the time for growth ($t_{f,growth}$).

Equation (III.74) presents the formulation of the optimal problem, where (III.75) depicts the constraints of this problem. The first constraint is the complete model, whereas the others are boundary constraints. These involve a maximum quantity of citric acid concentration CIT_{max}; an initial volume V_0 and maximum volume V_{max}; and an upper boundary for the input flow rates. The solution of (III.74) – (III.75) is done by pattern search until the optimal values are obtained. This optimization is performed every time ($t_{Control}$) until $t_{Control}=t_{f,growth}$.

$$\max_{qG_{IN},qN_{IN},t_{f,growth}} \frac{X_{Cat}\left(t_{f,growth}\right)}{t_{f,growth}} \tag{III.74}$$

$$\text{Subject to: } \begin{cases} Eq.(6) \\ V_0 < V < V_{\max} \\ CIT < CIT_{\max} \\ 0 < t_{f,growth} < t_{\max,growth} \\ 0 \le qG_{IN} \le qG_{IN\max} \\ 0 \le qN_{IN} \le qN_{IN\max} \end{cases} \tag{III.75}$$

The results of growth optimization are used to compute the initial state variables of the second step of the optimization.

V.4.2. Optimization of lipid accumulation

In this second step, the optimization is described by a multi-objective problem. The two-fold objective problem considers the maximization of the lipid productivity (P_{LIP}), and the lipid fraction (F_{LIP}) with respect to the total biomass at the final production time ($t_{f,prod}$) calculated as $P_{LIP} = \frac{V\left(t_{f,prod}\right)LIP}{t_{f,prod}}$ and $F_{LIP} = \frac{LIP}{X}\left(t_{f,prod}\right)$.

The proposition of a single objective weighted function (III.76) is here regarded to handle the dynamic maximization of the conflicting objectives with constraints (III.77).

$$\max_{qG_{IN},qN_{IN},t_{f,prod}} \left(\frac{P_{LIP}^{*} - P_{LIP}}{P_{LIP}^{*}}\right) + \left(\frac{F_{LIP}^{*} - F_{LIP}}{F_{LIP}^{*}}\right) \tag{III.76}$$

$$\text{Subject to: } \begin{cases} Eq.\ (6) \\ V_0 < V < V_{\max} \\ CIT < CIT_{\max} \\ t_{f,growth} < t_{f,prod} < t_{\max,prod} \\ 0 \le qG_{IN} \le qG_{IN\max} \\ 0 \le qN_{IN} \le qN_{IN\max} \end{cases} \tag{III.77}$$

Equation (III.76) holds the tracking values for lipid productivity and lipid fraction denoted as P_{LIP}^{*}, and F_{LIP}^{*}. The optimal problem with constraints is resolved by Pattern search to determine the optimal and feasible values of the input flow rates qG_{IN}, qN_{IN}, and the production time ($t_{f,prod}$).

The tracking values (P_{LIP}^{*} and F_{LIP}^{*}) of the conflicting objectives are calculated by MOPSO solving (III.78) – (III.79).

$$\max_{qG_{IN},qN_{IN}} \left[P_{LIP}^{*},\ F_{LIP}^{*} \right] \tag{III.78}$$

$$\text{Subject to: } \begin{cases} Eq.\ (11) \\ 0 \le qG_{IN} \le qG_{IN\max} \\ 0 \le qN_{IN} \le qN_{IN\max} \end{cases} \tag{III.79}$$

The reduce model (III.72) is used to compute the Pareto optimal sets for a prediction horizon fixed to t_h to determine the trade-offs of both objectives and the respective input flow rates to attain them. Subsequently, a unique trade-off (P_{LIP}^{*}, F_{LIP}^{*}) is selected as a tracking goal for both conflicting objectives. The selection is performed by finding the Pareto optimal point with the highest distance from a reference point ($P_{LIP,0}$, $F_{LIP,0}$), which values were placed at the origin.

The optimal values of the input flow rates qG_{IN}, qN_{IN} and the tracking values of both objective functions P_{LIP}^{*}, and F_{LIP}^{*} are stored to be introduced into (III.76) – (III.77) for the dynamic optimization with constraints. The results of (III.76) – (III.77) are then use to simulate, with the model (III.68), the new values of the state variables after a control action time $t_{Control}$. All those values are further used by MOPSO to determine new tracking objectives. Note that a new trade-off was searched at each $t_{Control}$. The sequential MOPSO Pattern search continues until the optimal conditions are obtained, the final fermentation time is equal to the optimal time or the constraints are violated.

V.5.Results and Discussion

V.5.1. Cases of study

In order to study the performance of the control strategy, three different cases are considered. Figure III.32 displays the execution of the control strategy. Growth optimization is equally accomplished for the three cases with Pattern Search until $t=t_{f,growth}$. Subsequently, lipid optimization is implemented from $t_{f,growth}$ to $t_{f,prod}$ in three different manners, denoted by the algorithm used, as follows:

1. Growth Control. Input flow rates remain constant ($qG_{IN} = qG_{IN}(t_{f,growth})$, and $qN_{IN} = qN_{IN}(t_{f,growth})$) until $t_{f,prod}$.
2. Pattern search. A single objective function (III.80) is used assuming equal weights for both conflicting objectives. The dynamic optimization takes into account the constraints defined in (III.77).

$$\max_{qG_{IN},qN_{IN},t_{f,prod}} P_{LIP} + F_{LIP} \tag{III.80}$$

3. MOPSO - Pattern Search. Sequential dynamic optimization with constraints (III.76) – (III.79).

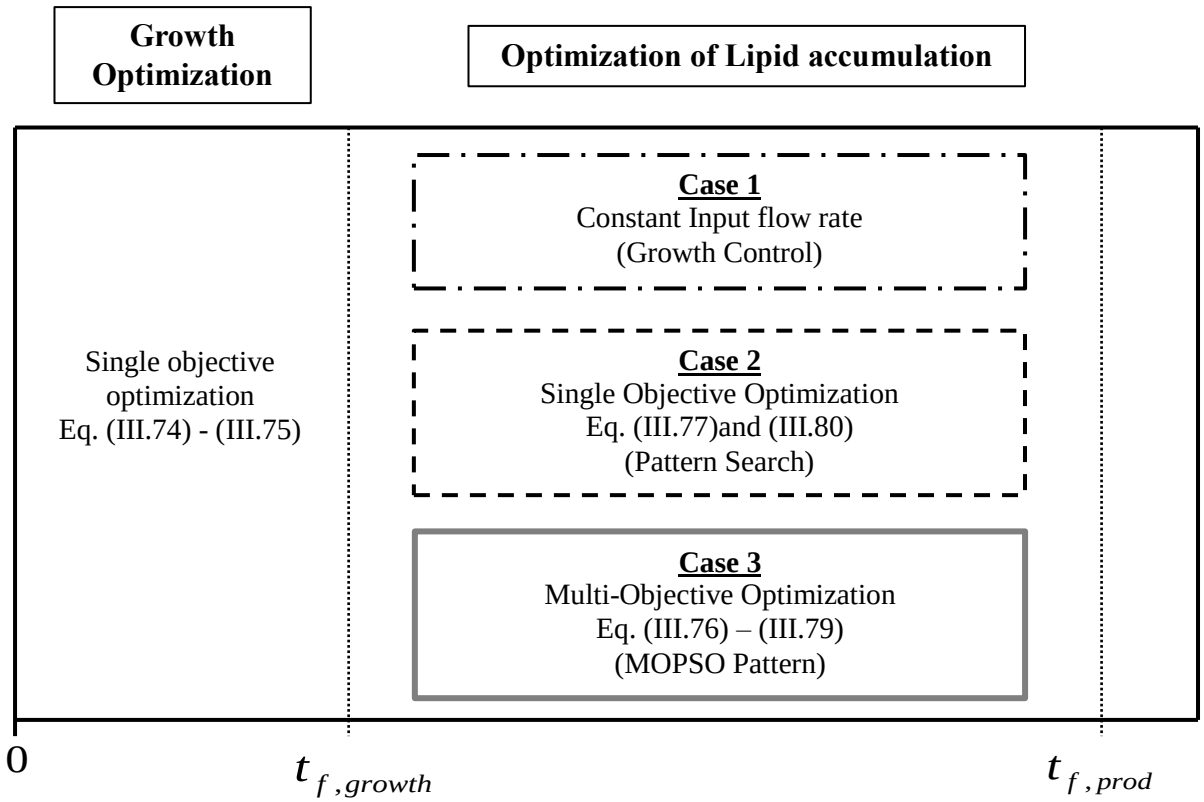


Figure III.32. Control implementation with the three cases of study.

Initial conditions for the state variables, volume limits (8 - 15*L*), manipulated variables, and disturbed variables are the same as the ones defined by (Cescut, 2009). The values of the inequality constraints are defined for the two steps of the optimization as follows,

Growth Optimizati on	Lipid accumulati on
$CIT < 1g/L$	$CIT < 5g/L$
$2 < t_{f,growth} < 15h$	$t_{f,growth} < t_{f,prod} < 60h$
$0 \le qG_{IN} \le 0.15L/h$	$0 \le qG_{IN} \le 0.3L/h$
$0 \le qN_{IN} \le 0.05L/h$	$0 \le qN_{IN} \le 0.1L/h$

The predictive horizon for MOPSO was $t_h = 3h$. The control action was set to $t_{Control} = 0.5h$. The results were simulated by (III.68) to generate the new initial values to be reinserted into the cycle for a new calculation until $t_{f,prod}$. Pareto fronts were calculated after 30 iterations, with a repository of 30 *pbest*.

V.5.2. Performance of the control strategy

The simulation results of the three cases are displayed in Figure III.33. The optimal time for growth phase was 4.5*h*.

The first 4 plots in Figure III.33 represent the considered disturbances and the input flow rates resulted from the optimization. It is seen that the optimization of lipids depicts an increasing profile for glucose feeding until the upper limit is achieved. Otherwise, the profiles for nitrogen are almost constant and low to induce a limitation that favors lipid accumulation. The impact of these feeding rates is thus illustrated on the next 5 plots of Figure III.33. Glucose and nitrogen input rates where increased reflecting an enlargement of residual glucose and nitrogen concentrations in the time for growth optimization. Those are further transformed into catalytic biomass, which value became also higher with a null impact on citric concentration and a small repercussion on lipid concentration.

After 4.5*h*, the multi-objective lipid optimization starts. Note that the obtained feeding profiles resulted from growth phase were maintained constant for case 1. This fact limited nitrogen consumption, reducing the production of lipids. For cases 2 and 3, the regulation of feeding profiles depicted an inverse behavior where glucose was increased, whereas nitrogen was lowered to manage an appropriate N/C ratio (Ochoa-Estopier and Guillouet, 2014). In case 2, a constant increase rate is observed for lipids and catalytic biomass concentrations causing its slower production with respect to MOPSO. For case 3, it is noted that glucose concentration

increases fast after 5h which is further consumed to boost catalytic biomass and lipid concentrations. The disturbance on glucose input concentration started at 10h. The response of case 2 was to increase the input flow rate close to its maximal value, whereas for case 3, this was forced to remain at its maximum value. A decrease was observed at 18h for cases 2 and 3 where the input concentration went to its original value. Citric acid started to be produced after 12h achieving a maximal value of 3g.L^{-1} of case 2, and a value near to 5 g.L^{-1} at 21h for case 3.

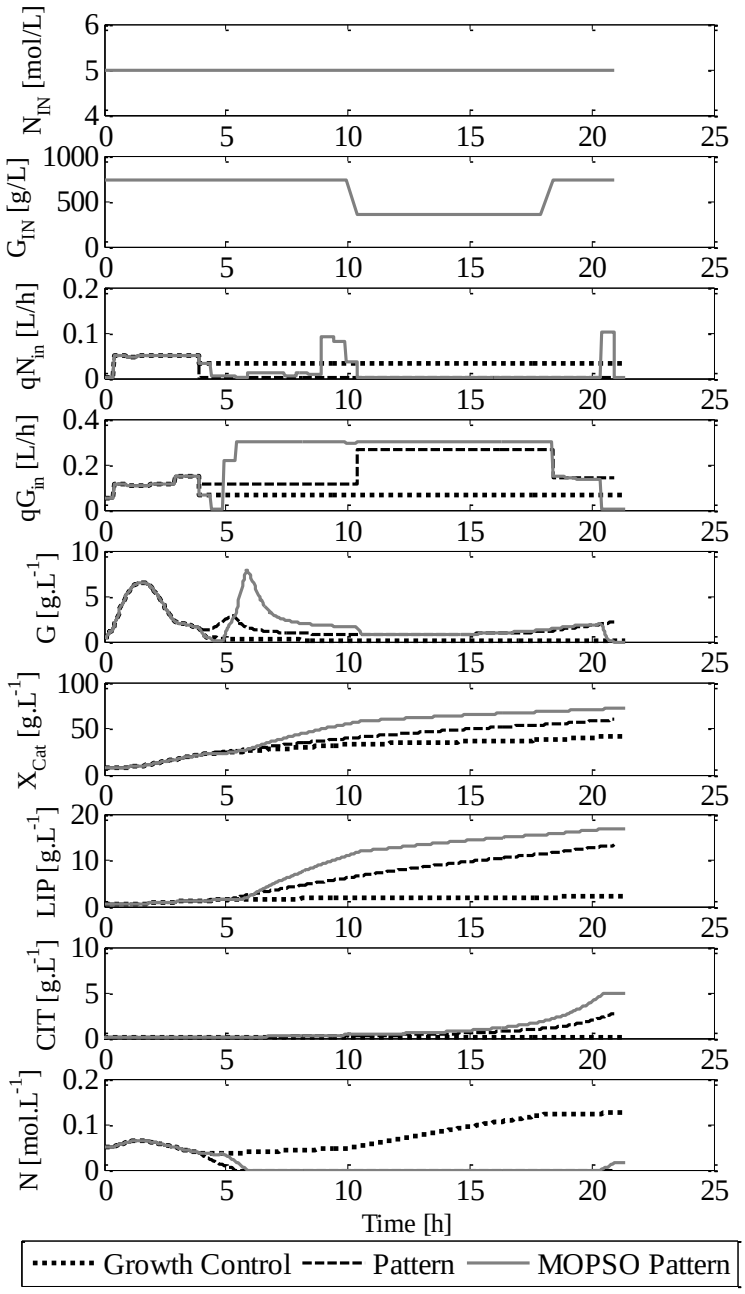


Figure III.33. Results of the Control strategy with the three cases: (- -) Growth control, (·-·) Pattern search, and (-) MOPSO Pattern Control

The weighted optimization by pattern search, and the dynamic MOPSO Pattern search demonstrated similar behaviors to manage feeding profiles and similar response against disturbances. However, the differences are remarkable when the two objectives are analyzed.

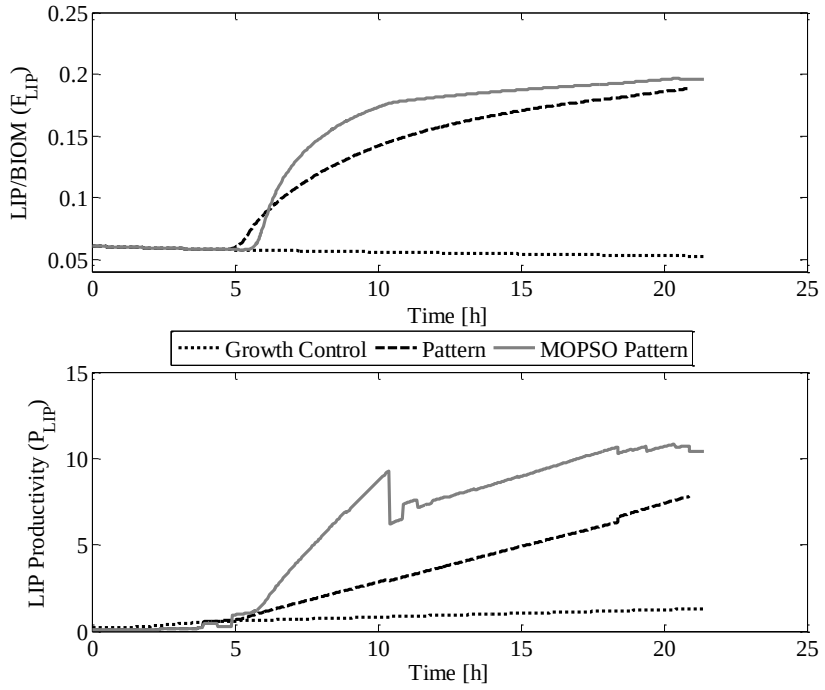


Figure III.34. Comparison of objectives throughout the fermentation time.

The optimization of P_{LIP} and F_{LIP} is displayed in Figure III.34. It is observed that when MOPSO actions started, the trade-off led to the maximization of both objectives. Since 10h, it is easily seen that the optimization chose to maximized lipid fraction where the productivity decreased by 2 units. Even if case 2 does not consider the trade-offs of the objectives, F_{LIP} is maximized. Notwithstanding, case 1 reflects a negative effect on both objectives.

Table III.20. Comparison of Objectives in the control strategy

Variable	P_{LIP} ($g.L^{-1}/h$)	F_{LIP} (g/g)	V (L)	LIP (g)	$t_{f,growth}$ (h)	$t_{f,prod}$ (h)
Growth	1.1	0.05	10.55	23.4	4.5	21.5
Pattern Search	7.68	0.188	12.14	160.6	4.5	21
MOPSO Pattern	10.43	0.196	13.40	223.3	4.5	21.5

Table III.20 shows the summary of the performance of the three cases. The obtained values demonstrate that a control is needed to optimize the storage of lipids. It is observed that the optimization time in case 2 is smaller than 3, which is reflected in the final volume. The lipid fraction with MOPSO provides a small improvement of 4.3% with respect to the equally weighted optimization. Notwithstanding, the mass of lipids and thus the productivity rose 40% highlighting the assets of considering trade-offs to deal with conflicting objectives.

V.6.Conclusions

This work presented a straightforward methodology towards the optimization of bioprocesses applied to fed-batch culture for lipid accumulation with *Yarrowia lipolytica*. The proposed MOPSO – pattern search dynamic method was applied to optimize lipid productivity and lipid fraction. Three cases were compared to analyze the performance of the method. The results demonstrated an important improvement of the simulation results even under the effect of a disturbance. The use of a reduced model lessened the computing time of Pareto optimal sets for the visualization and selection of trade-offs for the conflicting objectives. Further simulations will be performed under different conditions. An experimental validation will be performed to prove the multi-objective optimization of lipid accumulation.

V.7.References

Ashoori, A., Moshiri, B., Khaki-Sedigh, A., Bakhtiari, M.R., 2009. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. J. Process Control 19, 1162–1173. doi:10.1016/j.jprocont.2009.03.006

Casagranda, S., Ropers, D., Gouz, J., 2015. Model reduction and process analysis of biological models, in: Mediterranean Conference on Control and Automation (MED). pp. 1132–1139. doi:10.1109/MED.2015.7158908

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Coello Coello, C.A., 1999. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. Knowl. Inf. Syst. 1, 269–308. doi:10.1007/BF03325101

Coello, C.A.C. a C., Pulido, G.T.T., Lechuga, M.S.S., 2004. Handling multiple objectives with

particle swarm optimization. Evol. Comput. IEEE Trans. 8, 256–279. doi:10.1109/TEVC.2004.826067

Coello-Coello, C.A., Salazar-Lechuga, M.S., 2002. MOPSO: a proposal for multiple objective particle swarm optimization, in: Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600). IEEE, pp. 1051–1056. doi:10.1109/CEC.2002.1004388

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197. doi:10.1109/4235.996017

Jin, H., Chen, X., Yang, J., Wu, L., Wang, L., 2014. Hybrid intelligent control of substrate feeding for industrial fed-batch chlortetracycline fermentation process. ISA Trans. 53, 1822–1837. doi:10.1016/j.isatra.2014.08.015

Knowles, J.D., Corne, D.W., 2000. Approximating the nondominated front using the Pareto Archived Evolution Strategy. Evol. Comput. 8, 149–172. doi:10.1162/106365600568167

Konak, A., Coit, D.W., Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: A tutorial. Reliab. Eng. Syst. Saf. 91, 992–1007. doi:10.1016/j.ress.2005.11.018

Lalwani, S., Singhal, S., Kumar, R., Gupta, N., 2013. A comprehensive survey: Applications of multi-objective particle swarm optimization (MOPSO) algorithm. Trans. Comb. 2, 39–101.

Lewis, R.M., Torczon, V., 1996. Pattern Search algorithms for bound constrained minimization. SIAM J. Optim.

Marler, R.T., Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. 26, 369–395. doi:10.1007/s00158-003-0368-6

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Reyes-Sierra, M., Coello Coello, C.A., 2006. Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art. Int. J. Comput. Intell. Res. 2, 287–308. doi:10.5019/j.ijcir.2006.68

Riascos, C., Pinto, J.M., 2004. Optimal control of bioreactors: a simultaneous approach for complex systems. Chem. Eng. J. 99, 23–34. doi:10.1016/j.cej.2003.09.002

Robles-Rodriguez, C.E., Bideaux, C., Guillouet, S.E., Gorret, N., Cescut, J., Uribelarrea, J.-L., Roux, G., Molina-Jouve, C., Aceves-Lara, C.., 2016. Dynamic metabolic modeling of lipid accumulation and citric acid prodiction by Yarrowia lipolytica.

Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P., Van Impe, J.F., 2004. Optimal adaptive control of (bio)chemical reactors: past, present and future. J. Process Control 14, 795–805. doi:10.1016/j.jprocont.2003.12.005

Zitzler, E.E., Deb, K., Thiele, L., 2000. Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. 8, 173–95. doi:10.1162/106365600568202

VI. SOFT-SENSORS FOR LIPID FERMENTATION VARIABLES BASED ON PSO SUPPORT VECTOR MACHINES (PSO-SVM)

Carlos Eduardo Robles-Rodriguez[1,2,3*], Carine Bideaux[1,2,3], Gilles Roux[4], Carole Molina-Jouve[1,2,3], César Arturo Aceves-Lara[1,2,3*]

[1]Université de Toulouse ; UPS, INSA, INP, LISBP; F-31077 Toulouse, France
[2]INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France
[3]CNRS, UMR5504, Toulouse, France 135 Avenue de Rangueil, Toulouse Cedex, F-31077
[4]LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France

Accepted article for Oral presentation on the 13th International Conference on Distributed Computing and Artificial Inelligence (DCAI 2016). June 1 – 3, 2016. Seville, Spain.

Abstract

On-line monitoring fermentation variables (*e.g.* biomass) can improve the performance of bio-processes, as well as the quality of the targeted products. However, on-line estimation could be a challenging task when an accurate model is not available. Over the existing methods for state estimation, the support vector machine (SVM) is an attractive method for its fast convergence and generalization of the approximated function. In this paper, a soft-sensor based on SVM and coupled to Particle Swarm Optimization (PSO) algorithm is presented and applied to estimate the concentrations of lipid fermentation variables: lipids, biomass, and citric acid. The soft-sensor was trained with one data set, and validated with an independent data set of fed-batch fermentations. The PSO-SVM was compared with the SVM algorithm. In general, the results show that the PSO-SVM is an efficient alternative for monitoring fermentations.

VI.1. Introduction

The monitoring of certain variables is one of the main issues of optimizing fermentation because the yields of the targeted products can be estimated. Usually, it is difficult to measure online substrate, biomass, and product concentrations in the bio-process. The so-called "soft-sensors" are an alternative for on-line estimation. Soft sensors are software based sophisticated monitoring systems, which can relate the infrequently measured process variables with the easily measured (Wang *et al*., 2010). In this way, these soft-sensors assist in making the real-time prediction of the unmeasured variables (Desai *et al*., 2006).

Several software sensors have been proposed such as, modeling of mechanisms (Ramkrishna, 2003), adaptive observers (Bastin and Dochain, 1990), and Support vector Machine (SVM) (Wang *et al*., 2010). The first two required a detailed model of the process. Model development is sometimes a complicated task because it is not feasible to find a trade-off between simplicity and precision. Furthermore, SVM is a simpler and powerful alternative based on statistical learning with small samples. This method shares many of its features with the artificial neural networks, but it proposes some additional characteristics (Vapnik *et al*., 1996). It has good generalization ability of the regression function, robustness of the solution, and sparseness of the regression. Moreover, SVM also provides an explicit knowledge of the data points, which are important in defining the regression function. This methodology has been scarcely used in bio-processes/technology applications (Liu *et al*., 2010; Nadadoor *et al*., 2012; Ou Yang *et al*., 2015; Wang *et al*., 2010) including fermentation of penicillin, and the estimation of biomass, among others. Nevertheless, the realization of the advantages depends on the correct choice of parameters.

Particle Swarm optimization (PSO) is a global and efficient method of parameter estimation. This is an evolutionary computation technique developed by Kennedy and Eberhart in 1995 (Eberhart and Kennedy, 1995) and motivated by the behavior of organisms such as fishing schooling and bird flock. The algorithm is a validated evolutionary computation way of searching the extremum of function, which is simple in application and quick in convergence (Eberhart and Yuhui Shi, 2001).

In this context this work presents an extension of the SVM by PSO to estimate the parameters of the SVM in order to reach faster the optimal solution. This new PSO-SVM based soft-sensor is applied to estimate lipids, biomass, and citric acid concentrations from fed-batch fermentation cultures. The methodology of SVM and PSO-SVM is presented in section VI.2. Section VI.3 holds the results of the soft-sensor with the training and validation data. Finally, section VI.4 reports the conclusions and some perspectives of this work.

VI.2. Methods

VI.2.1. Support Vector Machines

The support vector machine theory relies on the idea to trap the input data into a feature space via nonlinear mapping. SVM maps the inner product of the feature space on the original space via kernels. Based on a given set of training data where $x \in R^N$ is the vector of the model inputs, and $y \in R$ is the vector of the scalar outputs (Vapnik *et al.*, 1996). The objective of the regression analysis is to determine a function that predicts accurately the desired outputs y in the form

$$y = w^T.\phi(x) + b \tag{III.81}$$

where $\phi(x): R^N \rightarrow R^M$ is the high dimensional feature space which is non-linearly mapped from the input space x. The vector w and the coefficient b are estimated by

$$\min_{w,\,\xi,\,\xi^*} J = \frac{1}{2} w^T.w + C.\sum_{i=i}^{N}\left(\xi_\iota + \xi_i^*\right) \tag{III.82}$$

Subject to:

$$\begin{cases} y_i - w^T.\phi(x_i) - b \le \varepsilon + \xi_i \\ w^T.\phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i,\ \xi_i^* \ge 0 \end{cases} \tag{III.83}$$

In equation (III.82) the first term is the regularized term, and the second term is the empirical error (risk) measured by the insensitive ε-loss function enabling to use less data points to represent the decision function given by equation (III.81). The variables ξ_i and ξ_i^* are the slack variables that measure the deviation of the support vectors from the boundaries of the ε-zone. The constant C is the regularization constant. It determines the trade-off between the empirical risk and the regularized term. The term ε is called the tube size and it is equivalent to the

approximation accuracy placed on the training data points. Both parameters will determine the efficiency of the estimation (Jianlin *et al*., 2006).

In order to simplify the minimization problem, Lagrange multipliers are introduced as follows:

$$L = J - \sum_{i=1}^{N} \alpha_i \left\{ \xi_i + \varepsilon - y_i + w^T . \phi(x_i) + b \right\} - \sum_{i=1}^{N} \alpha_i^* \left\{ \xi_i^* + \varepsilon + y_i - w^T . \phi(x_i) - b \right\} - \sum_{i=1}^{N} \left(\eta_i . \xi_i - \eta_i^* . \xi_i^* \right) \quad \text{(III.84)}$$

where the parameters α, α^*, η, and η^* are the Lagrange multipliers. According to the Karush-Kuhn-Tucker (KKT) of quadratic programming, the dual equation that can be obtained (Liu *et al*., 2010) is:

$$\min_{\alpha, \alpha^*} W = \frac{1}{2} \sum_{i,j=i}^{N} \left(\alpha_\iota - \alpha_i^* \right) . \left(\alpha_j - \alpha_j^* \right) . K\left(x_i, x_j \right) + \varepsilon . \sum_{i=i}^{N} \left(\alpha_\iota + \alpha_i^* \right) - \sum_{i=i}^{N} \left(\alpha_\iota - \alpha_i^* \right) . y_i \quad \text{(III.85)}$$

Subject to

$$\begin{cases} \sum_{i=i}^{N} \left(\alpha_\iota - \alpha_i^* \right) = 0 \\ 0 \le \alpha_\iota, \alpha_i^* \le C; \quad i = 1,2,\ldots,N \end{cases} \quad \text{(III.86)}$$

Therefore, the final regression function given in equation (III.81)is rewritten as

$$y = \sum_{i=i}^{N} \left(\alpha_\iota - \alpha_i^* \right) . K\left(x_i, x_j \right) + b \quad \text{(III.87)}$$

where $K\left(x_i, x_j \right) = \phi(x_i) . \phi\left(x_j \right)$ is the kernel function.

VI.2.2. Kernel functions

Different kernels can be selected to construct different types of SVM. The kernel can be any symmetric function satisfying the Mercer's condition. Typical examples include (Yan *et al*., 2004):

Polynomial kernels $K\left(x_i, x_j \right) = \left[(x . x_i) + 1 \right]^d$;

Radial Basis Function (Gaussian) kernels $K\left(x_i, x_j \right) = \exp\left(- \left(\left\| x - x_i \right\|^2 / 2\sigma^2 \right) \right)$;

The most used Kernel function is the Radial Basis Function (RBF) because it can classify multi-dimensional data, and it has fewer parameters to be set in comparison with other kernels. The parameter σ represents the width of the RBF.

VI.2.3. PSO – SVM

The accuracy of the results of SVM is linked to the setting of user parameters such as *C*, *σ,* and *ε.* The parameter *C* determines the trade-off between model complexity and model degree to which deviations larger than *ε* are tolerated (Liu *et al*., 2010). Kernel parameter *σ* determines the kernel width and relates to the input range of the training data set. If *C* is too large the estimation accuracy rate is very high in the training phase, but very low in testing phase. If *C* is too small, the estimation accuracy is unsatisfied, and the model is useless. An excessively large value for parameter *σ* leads to over fitting, while a disproportionately small value to under fitting. It is thus seen that SVM parameters are a combinatorial optimization problem, in which every combination generates a new solution function for equation (III.87). This optimization problem can be solved by cross verification trial which can result on a local minimum and costs high calculation times. Another solution is the use of global algorithms, such as particle swarm optimization (PSO) (Eberhart and Kennedy, 1995).

PSO is a simple algorithm implying a small quantity of parameters to be adjusted with fast convergence. In this paper, the objective function of PSO is given by the Akaike Information Criterion (AIC) to reduce the problem to a selection of best models (Burnham and Anderson, 2004). The *AIC* formula is given by

$$\mathrm{AIC} = n\left(\ln\left(2\pi\right)+1\right)+n.\ln\left(\frac{SSE}{n}\right)+2.p \qquad \text{(III.88)}$$

where *n* is the sample size, *p* is the number of factors in the model, *SSE* is the squared error: $SSE = 1/n.\sum_{i=1}^{n}\left(y_i - y_i^{*}\right)^2$, where y_i correspond to the off-line data, and y_i^{*} are the estimates of the model.

VI.3. Results and Discussion

The method PSO-SVM was applied to construct a soft-sensor of biomass, lipids, and citric acid concentrations from the lipid fermentation of *Yarrowia lipolytica* growing on glucose. Two independent data sets performed in fed-batch cultures (Cescut, 2009) were used in this work. Three variables were available from on-line measurements (Figure III.35(1) added base to control pH, (2) volume, (3) partial oxygen pressure (pO2).

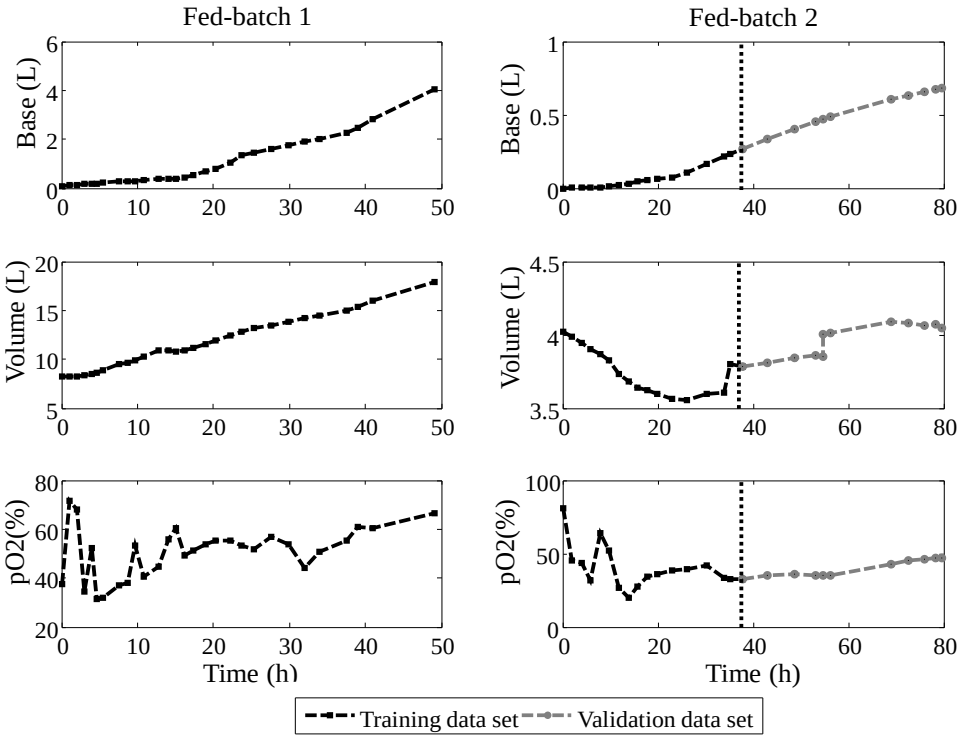


Figure III.35. Available on-line measurements for the fed-batch cultures

The combination of several variables was tested to design the soft-sensors (Table III.21). The discrimination of measurement was considered by the AIC values and the visualization of the results. Table III.21 presents the AIC values for training (AICt) and validation (AICv) for lipid concentration, where the combination of Base - pO2 reported low values for both sets. This methodology was also applied for the biomass concentration and the citric acid mass soft-sensors (Results not shown). The selected measurements for the PSO-SVM based soft-sensors are shown in Figure III.35.

Table III.21. Combinations to determine the best set for validation of lipid concentration.

Variable	AICt	AICv	Variable	AICt	AICv
Base	-673.89	67.24	pO2 – Base/Volume	-691.27	39.68
pO2	-144.35	20.14	pO2*Volume	-690.85	19.49
Volume	-690.63	2.46	pO2/Volume	-690.82	23.20
Base - pO2	-691.14	5.43	Base - pO2*Volume	-690.85	12.41
Base - Volume	-690.79	12.75	Base - pO2/Volume	-690.84	24.01
pO2 – Volume	-690.79	12.75	Base*Volume	-690.77	24.12
Base - pO2 - Volume	-690.81	15.44	pO2 – Base*Volume	-691.79	26.28
Base/Volume	38.51	103.61			

The kernel function was the RBF Gaussian function. The PSO follows Clerc's constriction factor (Eberhart and Yuhui Shi, 2001) using both learning factors (*i.e.* c_1 and c_2) at 1.5, with an initial inertia weight of 0.72. The population size was settled to 30 to consider a moderate search within the grid. The maximum number of iterations was fixed to 20 where after several runs the AIC variation was less than 0.001 in the AIC value.

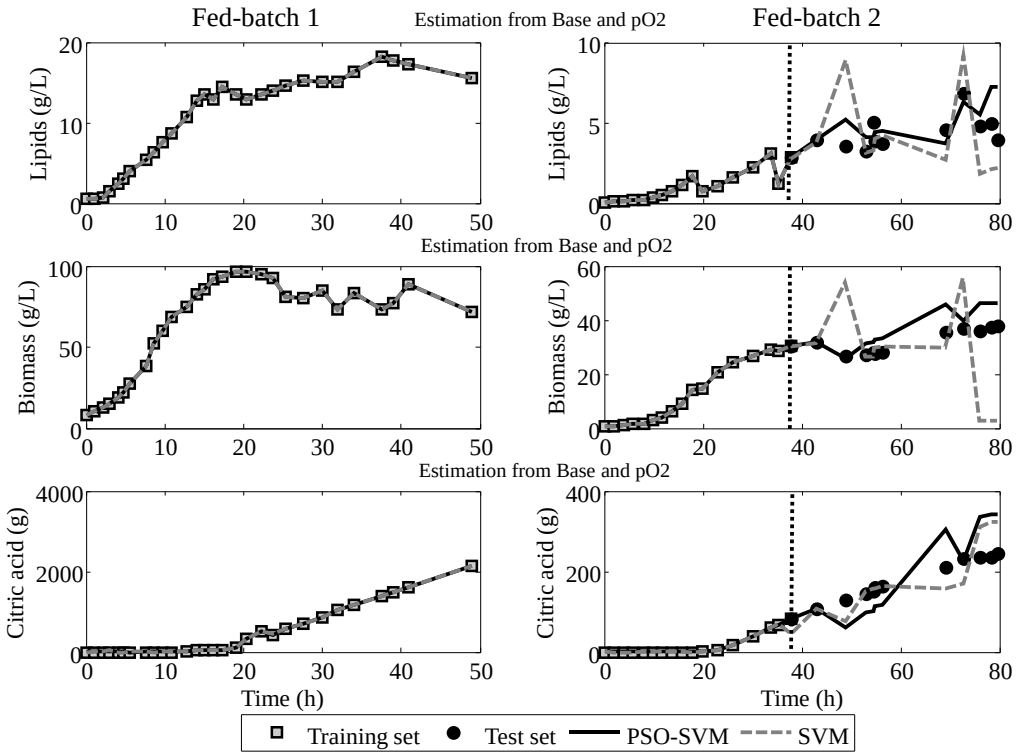


Figure III.36. Training and test results of soft-sensor based on PSO-SVM.

The two independent data sets (Cescut, 2009) were used to train and to validate the soft-sensor based on PSO-SVM. The first data set (Fed-batch 1, 29 data points), and the 60% of the other (Fed-batch 2, 16 data points between t = 0 - 38h) were used to train and to generate the support

vectors of the PSO-SVM. In all the cases of the soft-sensor for lipid and biomass concentrations, and citric acid mass the used measurements were the added base to control pH and the pO2.In order to test the ability of the PSO-SVM, the 40% left of the second data set (12 data points from t = 38 - 79h) was employed to validate the regression of the soft-sensor. In addition, standard SVM was used to compare and to illustrate the improvement of PSO-SVM. The parameters for standard SVM were taken from (Wang *et al*., 2010).

Figure III.36 displays the results of the training and test sets based on SVM and PSO-SVM. The dashed line separates the results of the training from the test sets. The values of C, σ, and ε and the numerical comparison of results are presented on Table III.22, where the *AIC* and *SSE* are used as performance indicators. The three soft-sensors reflect an acceptable fitting with the training data set with both PSO-SVM and SVM. However, *AIC* values for SVM are higher than the ones from PSO-SVM. This reflects that the selection of the parameters C, σ, and ε by PSO improves the fitting. The values of ε were small reflecting that the width of the tube surrounding the data is also small. This value allows having good results in the training sets, but it increases the risk of poor generalization performance of the estimator.

Table III.22. Comparison of performance evaluation of SVM and PSO-SVM.

Variable		Parameters		AICt		AICv		SSEv	
		SVM	PSO-SVM	SVM	PSO-SVM	SVM	PSO-SVM	SVM	PSO-SVM
Lipids	C	15000	3220	-484.8	-691.14	17.57	5.43	2.48	1.61
	ε	0.001	0.001						
	σ	1.5	0.724						
Biomass	C	15000	4260	-450.6	-690.61	140.6	79.30	200.5	22.47
	ε	0.001	0.001						
	σ	1.5	0.38						
Citric acid	C	15000	9160	107.62	-687.4	188.11	200.45	1095.1	1701.4
	ε	0.001	0.001						
	σ	1.5	0.58						

Regarding the results on the test data, lipid and biomass concentrations reflect acceptable results for PSO-SVM, where the estimation follows the behavior of the state variables. Otherwise, SVM shows a highly disturbed behavior on those concentrations. This is depicted by the higher AICv and SSE for SVM. The soft-sensors for citric acid mass presented a similar regression varying around the values all over the test data. Nevertheless, the AICv and SSE for PSO-SVM were higher than SVM exhibiting a better performance of SVM. This is due to the parameters

selection by PSO which was defined to find the best training set. The CPU times to do the training (t) with PSO-SVM were 317.7, 458.6, and 586.04 seconds for lipids, biomass, and citric acid respectively. The time to calculate the state variables (v) was 0.064s for the three of them, which is an encouraging value to couple the soft-sensor on-line.

VI.4. Conclusions

This work introduced a soft-sensor based on PSO-SVM for monitoring fermentation processes. The addition of PSO to SVM reduced the uncertainty generated by manually choosing the SVM parameters or the cross fitting computation time. The PSO-SVM method was successfully implemented to estimate lipids and biomass concentrations in fed-batch cultures. The results proved that PSO-SVM is an attractive and simple alternative for monitoring. Moreover, this method implemented a well trade-off between the quality of the approximation of the given data and the complexity of the approximating function. Future work will focus on increasing the fitness quality and extending the usefulness of the soft-sensor in other experimental working conditions.

VI.5. References

Bastin, G., Dochain, D., 1990. On-line Estimation and Adaptive Control of Bioreactors. Elsevier.

Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol. Methods Res. 33, 261–304. doi:10.1177/0049124104268644

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Desai, K., Badhe, Y., Tambe, S.S., Kulkarni, B.D., 2006. Soft-sensor development for fed-batch bioreactors using support vector regression. Biochem. Eng. J. 27, 225–239. doi:10.1016/j.bej.2005.08.002

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. MHS'95. Proc. Sixth Int. Symp. Micro Mach. Hum. Sci. 39–43. doi:10.1109/MHS.1995.494215

Eberhart, R.C., Yuhui Shi, 2001. Particle swarm optimization: developments, applications and resources, in: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat.

No.01TH8546). IEEE, pp. 81–86. doi:10.1109/CEC.2001.934374

Jianlin, W., Tao, Y.U., Cuiyun, J.I.N., 2006. On-line estimation of biomass in fermentation process using support vector machine. Chinese J. Chem. Eng. 14, 383–388.

Liu, G., Zhou, D., Xu, H., Mei, C., 2010. Model optimization of SVM for a fermentation soft sensor. Expert Syst. Appl. 37, 2708–2713. doi:10.1016/j.eswa.2009.08.008

Nadadoor, V.R., De la Hoz Siegler, H., Shah, S.L., McCaffrey, W.C., Ben-Zvi, A., 2012. Online sensor for monitoring a microalgal bioreactor system using support vector regression. Chemom. Intell. Lab. Syst. 110, 38–48. doi:10.1016/j.chemolab.2011.09.007

Ou Yang, H.-B., Li, S., Zhang, P., Kong, X., 2015. Model penicillin fermentation by least squares support vector machine with tuning based on amended harmony search. Int. J. Biomath. 8, 1550037. doi:10.1142/S1793524515500370

Ramkrishna, D., 2003. On modeling of bioreactors for control. J. Process Control 13, 581–589.

Vapnik, V., Golowich, S.E., Smola, A., 1996. Support vector method for function approximation, regression estimation, and signal processing. Annu. Conf. Neural Inf. Process. Syst. 281–287. doi:10.1007/978-3-642-33311-8_5

Wang, X., Chen, J., Liu, C., Pan, F., 2010. Hybrid modeling of penicillin fermentation process based on least square support vector machine. Chem. Eng. Res. Des. 88, 415–420. doi:10.1016/j.cherd.2009.08.010

Yan, W., Shao, H., Wang, X., 2004. Soft sensing modeling based on support vector machine and Bayesian model selection. Comput. Chem. Eng. 28, 1489–1498. doi:10.1016/j.compchemeng.2003.11.004

VII. DISCUSSION AND CONCLUSIONS

To assess the optimization of lipids, the first interest was modeling lipid accumulation in two cases, where the yeast could react differently (*i.e.* nitrogen limitation, and nitrogen deficiency cultures). Five models were developed with this aim. These models could be classified into two groups: macroscopic models and Dynamic Metabolic Models (DMM).

Regarding the macroscopic models, two models were presented: an unstructured model based on empirical kinetic equations and a quota model built on the Droop quota model (Droop, 1968). It is worth noting that the Droop model had been successfully used for modeling microalgae in mixotrophic, autotrophic, and heterotrophic conditions; however, in this work it was extended for yeast. Both models were efficiently calibrated and validated with two independent data sets taking into account the two cases above mentioned (*i.e.* nitrogen limitation, and nitrogen deficiency cultures). Moreover, overflow metabolism was included to represent the shift of this yeast towards the production of citric acid. This was modeled through Boolean indicators to represent the activation/deactivation of the overflow. Comparison results demonstrated that both models had equivalent fitting and prediction capabilities.

However, the quota based model presented a better description of extracellular nitrogen dynamics, which leads to a better prediction of overflow metabolism making it an interesting and better option for control purposes. This could be explained by the addition of the structured variable to account for the nitrogen quota, which allows identifying and confirming that at certain N/C ratio (0.77 molN.Cmol^{-1}) *Yarrowia lipolytica* shifts to citric acid production (Robles-Rodriguez et al., 2017). These two models could be straightforward implemented, but the generalization of their results could be limited when the environmental conditions of their design change, and/or when other metabolites would be studied, which imply the addition of extra equations with their respective kinetic expressions.

On the other hand, the development of DMM required a more detailed construction basis as explained in CHAPTER II. The use of large metabolic networks could cause computational burden and a great number of parameters to be identified. For this, a reduced metabolic network was proposed comprising 59 reactions and 51 metabolites, where special attention should be payed to the catalytic biomass (biomass without accumulated lipids) equation. The metabolic network was decomposed into Elementary Modes (EMs) based on the quasi-steady state

assumption (QSSA). As the number of EMs is large, the developed reduction method in this thesis allows a simple selection of EMs through polygons maximizing the convex hull. The advantages of this method are its applicability to cases where experimental data is not available and its simple application. Moreover, it was proven that the smallest polygon (triangle) could properly represent the whole metabolic network. Nevertheless, in some cases the selection of a small vertex polygon could not guarantee the maximization of convex hull, and a trade-off between generalization of the solution and number of EMs must be considered.

The three dynamic metabolic models proposed in this manuscript were: Hybrid Cybernetic Model (HCM), Macroscopic Bioreaction Model (MBM), and Fuzzy MBM. They were based on QSSA and they shared a common structure regarding the cultures in three phases: growth phase, lipid accumulation phase, and concomitant production of citric acid and lipids. HCM was the only one accounting with cellular regulation by cybernetic "enzymes", which allowed knowing the individual dynamics of the selected EMs. Nevertheless, these "enzymes" are lack of biological sense and add to the model an important number of parameters to estimate. MBM demonstrated to be a straightforward approach to reproduce experimental data and to identify kinetics, but it presented bad generalization for control purposes due to its time dependency for the identification of metabolic shifts. Fuzzy MBM was a new proposition to maintain the highlighting features of MBM and to overcome the dynamic generalization of metabolic shifts. This was performed by considering the dynamics of the ratio of the uptake rates of nitrogen and glucose (*rN/rC*), which was introduced into membership functions describing each phase.

The performance of these models was acceptable for the three different cultures of *Yarrowia lipolytica* from glucose: one under nitrogen limitation, and two under nitrogen deficiency conditions. The statistical comparison of our results showed that Fuzzy MBM and HCM had similar prediction abilities (comparable *AIC*). The drawbacks of these models are the generalization to mutant strains, which could imply the deletion of some reactions considered into our metabolic network.

From a general point of view, the five models could be compared regarding the building goals, provided insights, and prediction capabilities. The models were developed to be applied for control purposes. Even though, the microorganisms usually face disturbances, which cause metabolic shifts. These shifts could be better explained by the dynamic metabolic models, which present more robustness and reliability in comparison to others (Young *et al.*, 2008).

Moreover, these dynamic metabolic models provided more insights about the dynamic behavior of metabolism. They are able to reproduce metabolic fluxes, which could be useful to identify the active pathways in order to select targets for metabolic engineering. In addition, dynamic metabolic models have proven to be great tools for enhancing productivity (Song *et al.*, 2014). However, if the models are compared in the sense of *AIC*, the Quota Model will be preferable because this model can reproduce the data with a lower number of parameters. Nevertheless, the selection of the proper model should be carefully performed depending on the modeling goals to find an adequate trade-off between knowledge and complexity. Based on these thoughts, it was proposed to use two models for the control strategies: the Quota model, and the Hybrid Cybernetic Model.

In order to continue towards optimization and aiming at testing control strategies experimentally, state estimators were proposed. Observers were first to be considered, but after the observability tests (results not shown) the dynamic metabolic models resulted to be non-observable due to the lack of on-line state variables. Consequently, our proposition was based on data-based software sensors, such as Support Vector Machine (SVM). This algorithm is well known for having explicit knowledge of the regression function and a good generalization of the solution (Hsu *et al.*, 2008; Vapnik *et al.*, 1996). Furthermore, data-based software sensors could help to easily extend the range of application of these software sensors. One of those extensions being the use of mutant strains of *Yarrowia lipolytica*, where even the proposed dynamic metabolic models could fail. To our knowledge, only one work (Nadadoor *et al.*, 2012) have tried to estimate oil content in microalgae by Raman spectrophotometry, but not from simple measurements. Nevertheless, the novelty of our proposition is not only the application, but the addition of a global optimization algorithm Particle Swarm Optimization (*i.e.* PSO) to select the characteristic parameters of SVM based on the calculation of the *AIC*.

The PSO-SVM method was successfully implemented to estimate the on-line prediction of the concentrations of lipids, biomass, and citric acid in fed-batch cultures. The most important advantage of this method is the use of several input data that can be link to describe the behavior of the aimed state variable. This method links the input variables with the estimated variables, but the detail of those interactions and the relationships with other variables cannot be predicted at the level that models do. Obtaining accurate results highly depend on the training data set, which must contain a wide range of operating conditions to provide a better generalization of

the solution. This last could be seen as a disadvantage that is avoided when a great number of data is available. Therefore, the data sets and its quality could hamper the use of this method.

Our results in the concentrations of lipids, biomass and citric acid proved PSO-SVM as an effective and promising estimation method. Moreover, this algorithm happened to be fast on implementation, where an average of 450 seconds was needed for the training, whereas the validation was performed in 0.048s. This is a great feature for the real time application where corrections of the estimations could be carried out in situ. The on-line estimation of these state variables could then be used for the monitoring of the fermentation, and also to determine the initial conditions of the models that will be used for the optimal control at each time step.

The principal goal of this work concerned the optimization of the bioprocess. This often includes the use of two important criteria: yield and productivity. Yield is the ratio of the formed amount of a target product to the consumed amount of a substrate, and thus dimensionless. On the other hand, productivity is the substrate consumption rate multiplied by yield, and thus has the unit of rate. These two quantities are not necessarily proportional, and often show conflicts. Nevertheless, in the case of lipid accumulation, it is more useful to refer to lipid content fraction instead of yield relating the concentrations of the produced lipids with the total biomass. These two criteria were selected to modulate the *N/C* ratio that highly defines the metabolic shifts of lipid accumulation in *Yarrowia lipolytica*. The multi-objective optimization was solved in different manners by using first an objective function with constant weights, and after by evaluating the impact of those weights.

Firstly, the Quota model was used to optimize lipid fraction and productivity, where two objectives were weighted based on an optimal constant flow rate of both input variables. The optimization was solved by the pattern search algorithm through piecewise linear functions (PWL). Three cases with different numbers of PWL were tested to solve the two-folded optimization problem: (i) a constant flow rates, (ii) 2 PWL, and (iii) 5 PWL. The PWL functions were defined with open time intervals to optimize the time length of each linear function. The parameters to compute these PWL functions were bounded between zero and the optimized final time. Moreover, a sum constraint was added to keep the number of PWL functions (*i.e.* two and five) as defined. Therefore, the optimization problem in this case was extended to minimize the final time and the slopes and time intervals of the n-PWL.

The use of a quota model provided a volumetric productivity of 0.95 $g.(L\text{-}h)^{-1}$ with a lipid fraction of 0.23 $g.g^{-1}$ with a negligible concentration of citric acid after 30.5 h. Results of the optimization highlighted the need of regulation and the proper managing of the N/C ratio that can lead to the production of citric acid. This was shown by the fact that the use of piecewise linear functions increased 1.1 times productivity and 1.04 times lipid content fraction without citric acid production. In general, the optimization with the PWL functions suggested increasing linear functions for almost the whole time of the culture, with a strong decrease at the end. However, the changes of slope within each case were different. The larger number (5 PWL) provided better results with a difference of 5% of improvement in both objectives. This points out that the use of a larger number than 5 PWL functions could not be worth trying. Even if good and promising results were obtained, changes in the control strategy could be envisaged by having other manipulated variables, or even performing robustness tests by adding disturbances.

For the second case of study, the Hybrid Cybernetic Model was used as it was the first to be identified. In addition, the comparison of the results of the DMMs in the sense of AIC showed similar results for HCM and Fuzzy MBM. The optimization was split into two regarding the phases on which the dynamic metabolic models were constructed. In a first phase, growth was optimized to ensure having a reasonable quantity of catalytic biomass (lipid-free biomass). In a second phase, lipid fraction and productivity were optimized by using a continuously weighted objective function. The weights were obtained from the optimal dynamic trajectories of Pareto fronts. The methodology employed for the optimization based on Pareto front is a new proposition, which performs a dynamic search of the Pareto optimal set (*i.e.* calculation of the Pareto front every time step). This was accomplished by coupling the static unconstrained Multi-Objective Particle Swarm Optimization (MOPSO) to generate the ensemble of Pareto optimal points with a dynamic constraint optimization with the Pattern Search algorithm. The use of this methodology allowed the maximization of both criteria simultaneously, in other words, treating both with the same degree of importance. In order to highlight the salient features of this method, it was compared against constant flow rates and against an optimization without a Pareto weighted objective function solved by Pattern Search.

Results proved that the Pareto free optimization favored lipid fraction, whereas the MOPSO-Pattern method traded-off well both objectives. An increase of almost 40% was achieved for productivity, whilst only a 4% was improved on lipid fraction. In addition, the robustness of

this methodology was tested by reducing the glucose feeding concentration by half, where the results showed and adequate response to the disturbance. However, other disturbances could be added such as nitrogen feeding concentrations, or this could be also considered as other manipulated variables.

Although the MOPSO-Pattern methodology was effective, it presented two disadvantages to be assessed. The first was the large calculation time of the Pareto front. This issue was solved by the use of a reduced model, which was able to shorten the computation time to 80% of the actual time. The second corresponded to the selection of an optimal trajectory based on a unique point from the Pareto front. In this case, it was decided to use the maximum distance from the origin as the manner to find this optimal point. However, this distance depends on the geometry of the Pareto front, which could be affected by the presence of convex regions. For the results, volumetric productivity of 0.778 $g.(L\text{-}h)^{-1}$ and lipid fraction of 0.196 $g.g^{-1}$ were achieved in 21.5 h with 5 $g.L^{-1}$ of citric acid. It is noteworthy that an extra disturbance was considered reducing the feeding glucose concentration by half for 8 hours.

Although our results for both optimizations are merely theoretical, they are quite promising considering that literature results obtained with the same strain and the same initial conditions seized 0.25 $g.g^{-1}$ on lipid fraction and 0.23 $g.(L\text{-}h)^{-1}$ of volumetric productivity with 120 $g.L^{-1}$ of citric acid after 60 h of fermentation (Cescut, 2009). An experimental validation of the proposed control strategies should be performed to confirm our results. Nevertheless, the two control strategies presented are a step forward towards the optimization of lipid accumulation.

VIII. REFERENCES

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Droop, M.R., 1968. Vitamin B12 and marine ecology. The kinetics of uptake, growth and inhibition in monochrysis Lutheri. J. Mar. Assoc. 689–733.

Geng, J., Song, H.-S., Yuan, J., Ramkrishna, D., 2012. On enhancing productivity of bioethanol with multiple species. Biotechnol. Bioeng. 109, 1508–17. doi:10.1002/bit.24419

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2008. A Practical Guide to Support Vector Classification.

BJU Int. 101, 1396–400. doi:10.1177/02632760022050997

Nadadoor, V.R., De la Hoz Siegler, H., Shah, S.L., McCaffrey, W.C., Ben-Zvi, A., 2012. Online sensor for monitoring a microalgal bioreactor system using support vector regression. Chemom. Intell. Lab. Syst. 110, 38–48. doi:10.1016/j.chemolab.2011.09.007

Ramkrishna, D., Song, H.-S.S., 2012. Dynamic Models of Metabolism : Review of the Cybernetic Approach. AIChE J. 58, 986–997. doi:10.1002/aic

Robles-Rodriguez, C.E., Bideaux, C., Guillouet, S.E., Gorret, N., Cescut, J., Uribelarrea, J.-. L., Roux, G., Molina-Jouve, C., Aceves-Lara, C.., 2017. Dynamic metabolic modeling of lipid accumulation and citric acid prodiction by Yarrowia lipolytica. Comput. Chem. Eng. 100, 139–152. doi:10.1016/j.compchemeng.2017.02.013

Song, H.-S., Cannon, W., Beliaev, A., Konopka, A., 2014. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. Processes 2, 711–752. doi:10.3390/pr2040711

Vapnik, V., Golowich, S.E., Smola, A., 1996. Support vector method for function approximation, regression estimation, and signal processing. Annu. Conf. Neural Inf. Process. Syst. 281–287. doi:10.1007/978-3-642-33311-8_5

Young, J.D., Henne, K.L., Morgan, J.A., Konopka, A.E., Ramkrishna, D., 2008. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. Biotechnol. Bioeng. 100, 542–559. doi:10.1002/bit.21780

CHAPTER IV.

CONCLUSIONS AND PERSPECTIVES

CHAPTER IV. CONCLUSIONS AND PERSPECTIVES

The main objective of this work was to optimize the production of microbial lipids, also called Single Cell Oils (SCO), which could be considered as precursors towards biofuels production. The approach of the optimization was regarded from the point of view of automatic control, which included the development of models and software sensors for the implementation of control strategies.

In a first step, a literature review on lipid accumulation (CHAPTER I) was presented to compare the performances of several oleaginous microorganisms (accumulating more than 20% w/w), where yeasts are among the microorganisms with the highest conversion yields, and the highest lipid contents (Ratledge, 2004). The oleaginous yeast *Yarrowia lipolytica* was chosen based on its 36% w/w lipid content, and its fully sequenced genome (Barth and Gaillardin, 1997). This yeast accumulates lipids under nitrogen exhaustion and carbon excess conditions. However, when growing from glucose under the same conditions, it can also produce citric acid.

Previous optimization studies were focused on the construction of mutant strains able to accumulate higher quantities of lipids and the changes on experimental conditions (*e.g.* type of limitation, cultures modes). Nevertheless, other alternatives for optimization, such as the application of automatic control have been scarcely applied to lipid accumulation by oleaginous yeast (Muñoz-Tamayo *et al*., 2014). Therefore, we proposed the use of optimal control strategies to contribute at solving this problem. In this context, this study addressed three main axes: modeling, state estimation, and control. These axes were examined in the bibliographic review of CHAPTER II with an especial interest on dynamic modeling of metabolism, which was the core part of this PhD thesis. To assess the optimization of lipids, the first interest was modeling lipid accumulation cultures of *Yarrowia lipolytica* from glucose under nitrogen limitation, and nitrogen deficiency conditions. Different model frameworks could be used to capture the salient features of the system depending on the attained goals (Gernaey *et al.*, 2010). Those objectives could include understanding of system characteristics, prediction of system behavior in new conditions, and discovery of new strategies for process improvement. The basic phenomenological input to such models was the conversion rates (*i.e.* kinetics of various biochemical reactions) of various process components displayed in a dynamic formulation based on differential equations.

In this work, five models were developed: two macroscopic models and three Dynamic Metabolic Models (DMM). The main assumption of the models in our study was based on the fact that lipid accumulation could be seen as the set of three events: growth, lipid accumulation, and citric acid production. In the DMM, the first assumption was to consider each "event" as different metabolic phases, whereas in the kinetic models it was taken into account by overflow metabolism. Therefore, the main idea was focused on identifying the shifts enhancing the produced metabolites. However, when several models are employed, a discrimination strategy is needed to assess the relative effectiveness. In case that models show similar performance in fulfilling a specific goal, the model with the best parsimony would be preferred. The goodness of fit and parsimony of the model can be measured using theoretical information and statistical tools (*i.e. AIC*) (Burnham and Anderson, 2004).

The *AIC* values showed similar values between the models, making difficult the choice of the best model. From the DMM, Fuzzy MBM and HCM were similar in the sense of AIC, whilst from the macroscopic models the quota model was preferred for its better managing of nitrogen source. Regarding model features, the DMM provide more insights about the dynamic behavior of metabolism (Song et al., 2014a). In this context, HCM and the Quota model were used on control strategies.

In order to implement a control strategy, monitoring of the process variables is needed. For this reason, three soft sensors were built on Support Vector Machines (SVM), whose parameters were optimized by PSO. The coupling of these two allowed an effective and promising estimation. Our results in the concentrations of lipids, biomass and citric acid with PSO-SVM were compared with a non-tuned SVM where the method proved to be better. The principal advantage of this method was its simplicity and fastness on application, which are important features on a real time scheme.

The developed models and soft sensors facilitate thus fulfilling the main goal of this work consisting on the optimization of lipid accumulation. This was addressed by the maximization of two criteria: lipid productivity and lipid content fraction. Two different strategies were used to solve the optimization. The first used a constant weighed objective function solved by piecewise linear functions by integrating the quota model. The second used a dynamically weighted objective function, where a Pareto front was calculated every time step with a reduced

HCM. Our results are theoretical, but promising for a real application due to the attained values of 0.778 – 1.01 $g.(L\text{-}h)^{-1}$ and 0.196 – 0.22 $g.g^{-1}$ for lipid productivity and lipid content fraction, whereas the literature values were reported from 0.08 to 0.27 $g.(L\text{-}h)^{-1}$, and from 0.17 to 0.25 $g.g^{-1}$, respectively (Aggelis and Komaitis, 1999; Beopoulos et al., 2009a; Cescut, 2009; Ochoa-Estopier, 2012).

It is noteworthy that this study is the first to propose models for describing lipid accumulation and citric acid production based on metabolism. Moreover, this work provides the first insights towards the optimization of lipid accumulation by automatic control approach.

From our point of view, this work leads to several perspectives in order to improve and complement the three axes of this work: modeling, state estimation, and control. The natural next step towards the optimization is the experimental validation of the control strategy. However, this involves several tasks to be previously completed.

In a first time, supplementary software sensors must be developed to estimate online other state variables (*i.e.* feeding concentrations, glucose and nitrogen titers) that are also required to be incorporated into the model for the optimization control. This could be achieved by the addition of other on-line measurements such as the *RQ*, gas composition, and semi-online measurements like optical density (*OD*). Moreover, the use of a larger quantity of measurements could promote a better estimation of our method and will enhance having a better prediction of the state variables.

Then, the coupling of the control strategy to the soft sensors should be validated in simulations to test the accurate link between the soft sensors and the control law. Finally, the experimental set up of our control strategy, coupled to soft sensors, will allow the valorization of our propositions for the enhancement of the optimal operation of this process.

Alternative improvements and extensions could be performed to the control strategy, such as the inclusion of a classification method to select the optimal point in the Pareto front. This will support the selection of the optimal trajectory. This could be performed by the addition of decision making aid approaches such as the multi-attribute utility theory. However, a computationally fast method should be chosen to guarantee the proper application of the proposed control strategy.

A fusion of sensors could be envisaged by the evaluation of the different combinations of sensors performed by SVM in order to have more flexibility for monitoring. Furthermore, this could be coupled to fault detection to guarantee the availability of reconstructed state along fermentation time. Moreover, the inclusion of fault detection will grant the exposure of failures or miscalculation by the sensors.

Concerning modeling, the dynamic metabolic models are able to predict intracellular fluxes, which are highly relevant to systems biology and could be used for applications focusing on monitoring metabolic behavior for metabolic engineering (Sauer, 2006). However, it would be compelling to validate flux distribution by methods using 13C labeling (omics). This will allow corroborating our metabolic network to highlight the importance of the models and could be of great aid for metabolic engineering *Yarrowia lipolytica*.

On the other hand, it would be interesting to compare the model simulation results with a more complete stoichiometric matrix, such as the detailed metabolic model of Pan and Hua, (2012). Moreover, this could also help verifying if flux distribution is affected by the reduction. Finally, the dynamic metabolic models (HCM and Fuzzy MBM) could be easly extended for mutant strains of *Yarrowia lipolytica* and the use of other substrates (*e.g.* glycerol, xylose). This will allow testing the robustness of the model providing a generalized version with better prediction capabilities.

The results presented in this manuscript provide some advances on the understanding of the complex behavior of *Yarrowia lipolytica* in lipid fermentations. However, this work could also motivate undertaking further experiments where the transient dynamics of internal metabolites could be followed by advanced analytical tools. The increasing knowledge and the strong trend of omics data (*e.g.* fluxomics) could be a conceivable option to complete the provided information and to improve the dynamic models. In this way, it is believed that the insights provided by this study could be completed to better understanding this interesting and promising oleaginous microorganism.

This manuscript contributed to the study of optimization through process control. However, it is believed that the synergy of the proposed control strategies and the soft sensors coupled to

metabolic engineered strains could furnish the better alternative towards the optimal production of lipids as precursors for biofuels.

Finally, I would like to close by citing two quotes: the first from the statistician George Edward Pelham Box and the second by the father of informatics Alan Turing.

"All models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be born in mind "

George Edward Pelham Box

"Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two facilities, which we may call intuition and ingenuity".

Alan Turing

REFERENCES

Aggelis, G., Komaitis, M., 1999. Enhancement of single cell oil production by Yarrowia lipolytica growing in the presence of Teucrium polium L. aqueous extract. Biotechnol. Lett. 21, 747–749. doi:10.1023/A:1005591127592

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol. Rev. 19, 219–37.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol. Methods Res. 33, 261–304. doi:10.1177/0049124104268644

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences

Appliquées de Toulouse.

Gernaey, K. V., Lantz, A.E., Tufvesson, P., Woodley, J.M., Sin, G., 2010. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. Trends Biotechnol. 28, 346–54. doi:10.1016/j.tibtech.2010.03.006

Muñoz-Tamayo, R., Aceves-Lara, C.-A., Bideaux, C., 2014. Optimization of lipid production by oleaginous yeast in continuous culture, in: IFAC Proceedings Volumes. Cape Town, South Africa, pp. 6210–6215. doi:10.3182/20140824-6-ZA-1003.02125

Ochoa-Estopier, A., 2012. Analyses systématique des bascules métaboliques chez les levures d'intérêt industriel: application aux bascules du métabolisme lipidique chez Yarrowia lipolytica. Institut National des Sciences Appliquées de Toulouse.

Pan, P., Hua, Q., 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS One 7, e51535. doi:10.1371/journal.pone.0051535

Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. Biochimie 86, 807–15. doi:10.1016/j.biochi.2004.09.017

Sauer, U., 2006. Metabolic networks in motion: 13C-based flux analysis. Mol. Syst. Biol. 2, 62. doi:10.1038/msb4100109

Song, H.-S., Cannon, W., Beliaev, A., Konopka, A., 2014. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. Processes 2, 711–752. doi:10.3390/pr2040711

APPENDIX .

MATERIALS AND METHODS

APPENDIX. MATERIALS AND METHODS

SUMMARY OF THE CHAPTER

In this chapter the description of the experimental data sets used along this work is addressed. These data sets were used to propose some assumptions of the different models, to validate the models and to train the soft-sensors developed.

The experimental data of three cultures are presented in this appendix. The first was reported by Cescut, (2009). The other two were performed by other members of the team FAME (Fermentation Advances and Microbial Engineering, Team EAD8) at LISBP.

The strain and the culture medium are first presented. The description of the analytical methods to determine the major metabolites of these cultures is presented. Finally, the three experimental cultures consisting on one fed-batch and two sequential batch modes are analyzed with the sake of modeling.

I. STRAIN AND CULTURE MEDIUM

The microorganism used was *Yarrowia lipolytica* wild type W29. This strain was obtained from the Microbiology and Molecular Genetics Laboratory (Paris-Grignon, France). The optimal growth conditions of this yeast are: 28 °C and pH of 5.6 (Faure, 2002). The formula of the biomass in exponential growth is $CH_{1.744}O_{0.451}N_{0.132}$ corresponding to a molecular weight of 24.67 $g.Cmol^{-1}$ (Cescut, 2009). This formula represents 92% of the total biomass weight where 8% are ashes.

The strain was maintained at 4 °C on plates of YPD/glucose agar medium with 10 $g.L^{-1}$ of yeast extract, 20 $g.L^{-1}$ bactopeptone, 20 $g.L^{-1}$ glucose and 20 $g.^{L-1}$ agar. Pure sterile glycerol was added to the culture broth to reach a concentration of 30 % (v/v) for cryopreservation of the cells. After sterile glycerol addition, the glycerol stocks were stored in sterile vials at 80 °C. These frozen stock cultures were used to inoculate pre-cultures for fed-batch fermentation (Ochoa-Estopier and Guillouet, 2014).

Precultures were carried out on a 5-mL tube of YPD medium at 28 °C for 16 h at 100 rpm. The culture in growth phase was transferred into a 1 L Erlenmeyer flask containing 170 mL of mineral medium at pH 5.6.

The composition of mineral medium was: K_2HPO_4: 3 $g.L^{-1}$; $(NH_4)_2SO_4$: 3 $g.L^{-1}$; NaH_2PO_4,H_2O: 3 $g.L^{-1}$ $MgSO_4,7H_2O$: 1 $g.L^{-1}$; $ZnSO_4,7H_2O$: 0.04 $g.L^{-1}$; $FeSO_4,7H_2O$: 0.0163 $g.L^{-1}$; $MnSO_4,H_2O$: 0.0038 $g.L^{-1}$ $CoCl_2,6H_2O$: 0.0005 $g.L^{-1}$; $CuSO_4,5H_2O$: 0.0009$g.L^{-1}$; $Na_2MoSO_4,2H_2O$: 0.00006 $g.L^{-1}$; $CaCl_2,2H_2O$: 0.023 $g.L^{-1}$; H_3BO_3: 0.003 $g.L^{-1}$; and 10 mL of vitamins solution (Cescut, 2009).

Vitamins solution contained: d-biotin: 0.05 $g.L^{-1}$, hydrochloride thiamine: 1 $g.L^{-1}$, panthotenic acid: 1 $g.L^{-1}$, hydrochloride pyridoxol: 1 $g.L^{-1}$, nicotinic acid: 1 $g.L^{-1}$, p-aminobenzoic acid: 0.2 g. L^{-1}, myo-inositol: 25 $g.L^{-1}$. Before sterilization, pH of this medium was adjusted to 4.5 with H_3PO_4 solution and at the working pH (5.6) with ammoniac solution.

II. ANALYTICAL METHODS

II.1. Biomass concentration

Biomass concentration during yeast growth was evaluated by spectrophotometry at 600 nm in a spectrophotometer HITACHI U-1100. Cell dry weight determination, culture samples (5–10 ml) were harvested by filtration on a 0.45 µm membrane (Sartorius) and dried at 200 mm Hg and 60 °C during 48 h until a constant weight was achieved.

II.2. Organic acids and Sugar

Culture supernatant was obtained by centrifuging (Min-iSpin Eppendorf, USA) the fermentation culture in Eppendorf tubes at 3200 g for 3 min. The supernatant was then filtrated by a polyamide membrane with porosity of 0.45 µm (SARTORIUS®, Germany). The determination of the alcohols, organic acids and sugars concentrations from supernatants was performed by HPLC (Cescut, 2009; Ochoa-Estopier, 2012; Ochoa-Estopier and Guillouet, 2014). HPLC analyses were carried out using a column Aminex HPX-87H (300 mm *7.8 mm) with the following conditions: temperature of 50 °C, with 5 mM H_2SO_4 as eluent (flow rate of 0.5 ml.min^{-1}) and dual detection (refractometer and UV at 210 nm). Compounds were identified and quantified with respect to standards.

II.3. Ammonium

An ammonium ion electrode (PH/ISE meter model 710A + Ammonia Gas Sensing Electrode Model 95-12, Orion Research Inc. Boston USA) was used to quantify the residual ammonia concentration in the culture medium. 5 mL of culture broth were filtered on a polyamide membrane with pore-size diameter of 0.45 µm (Sartorius AG, Germany) and mixed with 200 µl of ISA buffer solution (Ionic Strength Adjustor: NaOH 5 mol.L^{-1}; disodium EDTA, 0.05 mol.L^{-1} and methanol 1 g.L^{-1}). The concentration of NH_4 was determined by correlating a calibration curve on mV with the NH4 concentration in a range of 10^{-1}–10^{-4} mol.L^{-1} NH_4Cl.

II.4. Lipid Extraction

Two analytical methods were used to determine the content of total lipids among the three cultures presented before. In the followings paragraphs, these methods are briefly described.

II.4.1. Culture under nitrogen limitation

The procedure used for total lipids extraction in the nitrogen limitation culture (Cescut, 2009) consisted in three consecutive extractions in methanol/chloroform at different concentrations (2:1, 1:1, and 1:2 v/v), where 500 mg of dry mass cells in 15 mL represented one volume. Each extraction step consisted of incubation for 24 h at room temperature on roller mixer. As the extraction could "extract" non-lipid contaminants (sugars, amino acids, proteins and salts), the organic phase was rocking against a KCl solution (0.08 $g.L^{-1}$) for 15 min and then a liquid/liquid separation technique was used after centrifugation (5000 g, 10 min).

Finally, lipids were recovered as dry material after solvent evaporation in a Rota vapor (35 °C). Total lipids content was quantified by gravimetric method. Lipid extracts were kept in chloroform at –20 °C. Fatty acids were methylated into fatty acid methyl ester (FAME) using trimethyl sulfonium hydroxide (TMSH, 0.2 M in methanol, Macherey-Nagel, Germany). Gas Chromatography analysis were carried out on a Hewlett-Packard 5890 gas chromatograph equipped with a 50 m*250 µm*25 µm WCOT Fused Silica column (VARIAN, USA) and a FID detector. The carrier gas was nitrogen with a 50 $mL.min^{-1}$ flow. Oven temperature was varied in different gradients: first 9 $°C.min^{-1}$ for a range between 50-75 °C, after 13 $°C.min^{-1}$ for 75-140 °C and 140-180 °C, then 1.5 $°C.min^{-1}$ for 180-240 °C, and finally 4.5 $°C.min^{-1}$ for an injection temperature of 140 °C, detector temperature 250 °C. Identification of methyl esters was based on the comparison of retention times with known standards.

II.4.2. Cultures under nitrogen deficiency

Total lipids extraction in the nitrogen deficiency cultures was based on the digestion-methylation method (Browse *et al.*, 1986). The procedure consisted on three general steps: sample preparation, digestion/methylation, and purification. Sample preparations were centrifuged at 3200 g for 5 min at 4 °C (Centrifuge 5810R, Rotor F-34-6-38, Adapters

Reference 5804 775.007 and 5804 776.003, Eppendorf®, US). Culture supernatants were separated from the cell pellet and were kept at -20 °C to quantify extracellular metabolites. The cell pellets were washed with a 0.9 % NaCl saline solution at 4 °C and centrifuged at 790 g for 10 min at 4 °C (Centrifuge 5810R, Rotor A-4-62, Adapters Reference 5810 763.002 and 5810 755.000 Eppendorf®, US). Washed cells were kept at -81 °C for at least 8 h before lyophilisation and freeze-dried for 120 h.

20 mg of freeze-dried cells were suspended for the digestion methylation with 2 mL of 2.5 % Methanol/H_2SO_4 solution (v/v) containing 1 $mg.mL^{-1}$ antioxidant (2.6-di-tert-buthyl-4-methylphenol, (BHT)). The tubes were incubated in an agitated water-bath at 85 °C for 120 min. After that time, sealed Pyrex tubes were kept at room temperature.

For the purification step, 1.5 mL of 5 % NaCl (v/v) solution with 1 $mg.mL^{-1}$ BHT, and 2 mL of n-Hexane containing 1 $mg.mL^{-1}$ BHT were added to the tubes to be mixed twice for one minute by a vortex mixer (Multi Reax shaker, Heidolph, Germany). The obtained samples were centrifuged at 790 g for 10 min at 4 °C. The collected organic phase was further analyzed by gas chromatography.

Gas Chromatography analysis were carried out on an Agilent Hewlett-Packard 6890 Series gas chromatograph equipped with a column Varian CP7419 L*ID*OD*Film Thickness 50 m*0.25 mm*0.36 mm*0.25 µm WCOT Fused Silica column (VARIAN, USA) and a FID detector. The carrier gas was nitrogen with a flowrate of 20 $mL.min^{-1}$ Oven temperature was varied in different gradients: first 10 °$C.min^{-1}$ for a range between 50-140 °C, after 20°$C.min^{-1}$ for 140-190 °C, and finally 25 °$C.min^{-1}$ for 190-275 °C, for an injection temperature of 140 °C, detector temperature 250 °C. The identification and calculation of fatty acid methyl esters were based on the comparison of retention times. All chemical compounds and solvents used in lipid analysis had the highest analytical grade (Sigma-Aldrich).

III. EXPERIMENTAL DATA

The results from three cultures of *Yarrowia lipolytica* on glucose under nitrogen deficiency or limitation conditions were used. The experimental set-up of the process for the cultures can be generalized as presented in Figure A0.1. Glucose, salts, and ammonium solutions, were the

main feeding streams of the bioreactor. Other additions were antifoam and base and acid solutions to control pH. As displayed in Figure A0.1, the measurements of pH, pO_2, temperature, pressure, air composition, and the several additions were monitored by an acquisition system. The variant details within experiments will be further addressed in the description of each culture.

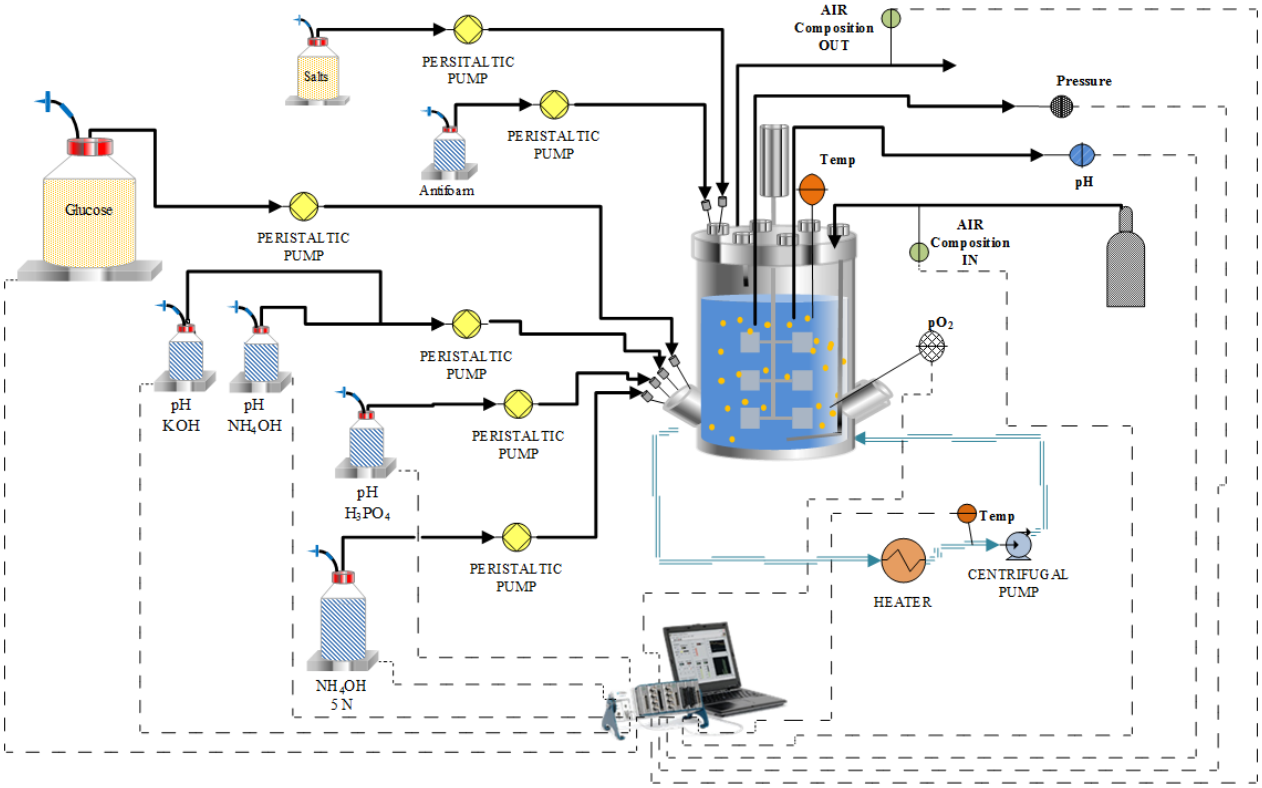


Figure A0.1. Experimental set-up for the cultures of *Yarrowia lipolytica* on glucose: Solid lines represent the flow of the inlets and the outlet; dashed lines are the on-line measures stored in the acquisition system.

The results of the cultures are expressed as concentrations ($g.L^{-1}$) of the different metabolites, where some definitions are needed.

- Total biomass: The measured biomass concentration including lipids.
- Catalytic biomass: Total biomass without lipids.
- Lipid fraction: It relates the concentration of total lipids over the concentration of total biomass ($g.g^{-1}$).

The first experiment is a fed-batch culture where glucose and nitrogen were continuously fed into the bioreactor. In this culture, lipid accumulation was performed under nitrogen limitation.

The other two experiments are seen as sequential batch cultures with glucose pulses and no additions of the nitrogen source. These two are then regarded as cultures where lipid accumulation was obtained under nitrogen deficiency. A detailed explanation of the obtained results is reported within this section.

III.1. Culture under nitrogen limitation

This data set correspond to a Fed-batch culture reported by Cescut (Cescut, 2009). This experiment was conducted in a 20 *L* bioreactor Figure A0.2. Start-up medium had the same composition as the inoculation medium without $(NH_4)_2SO_4$. Operating conditions were controlled to a temperature of 28 °C, and a pH value of 5.6 regulated by a 5 *M* KOH solution. Monitoring and control was performed by a home-made software developed in our laboratory by P. Perlot (Béghin Say) and J.-L. Uribelarrea (INSA) (see Cescut, 2009). This software allowed the on-line acquisition of stirring speed, pH, temperature, percentage of dissolved oxygen, base and antifoam additions.

The inlet and outlet gas compositions in carbon dioxide (CO_2) and oxygen (O_2) were measured by a fermentation gas monitor system (LumaSense technologies Europe). The system combines a multipoint sampler 1309 with a gas analyzer (INNOVA 1313).

Figure A0.2. Bioreactor of 20*L* used for the nitrogen limitation culture.

The input and output oxygen flow rates were measured online, as well as the agitation. For the Fed-batch culture, two main streams were used: glucose (730 $g.L^{-1}$) and ammonia (5 *M*), which were fed continuously as displayed in Figure A 0.3. The glucose and nitrogen feeding rates, as well as the pH control additions were measured by weighing balances as in Figure A0.1. Other measured variables were the inlet and outlet gas composition, gas flow rate, broth temperature, pressure, and aeration flow.

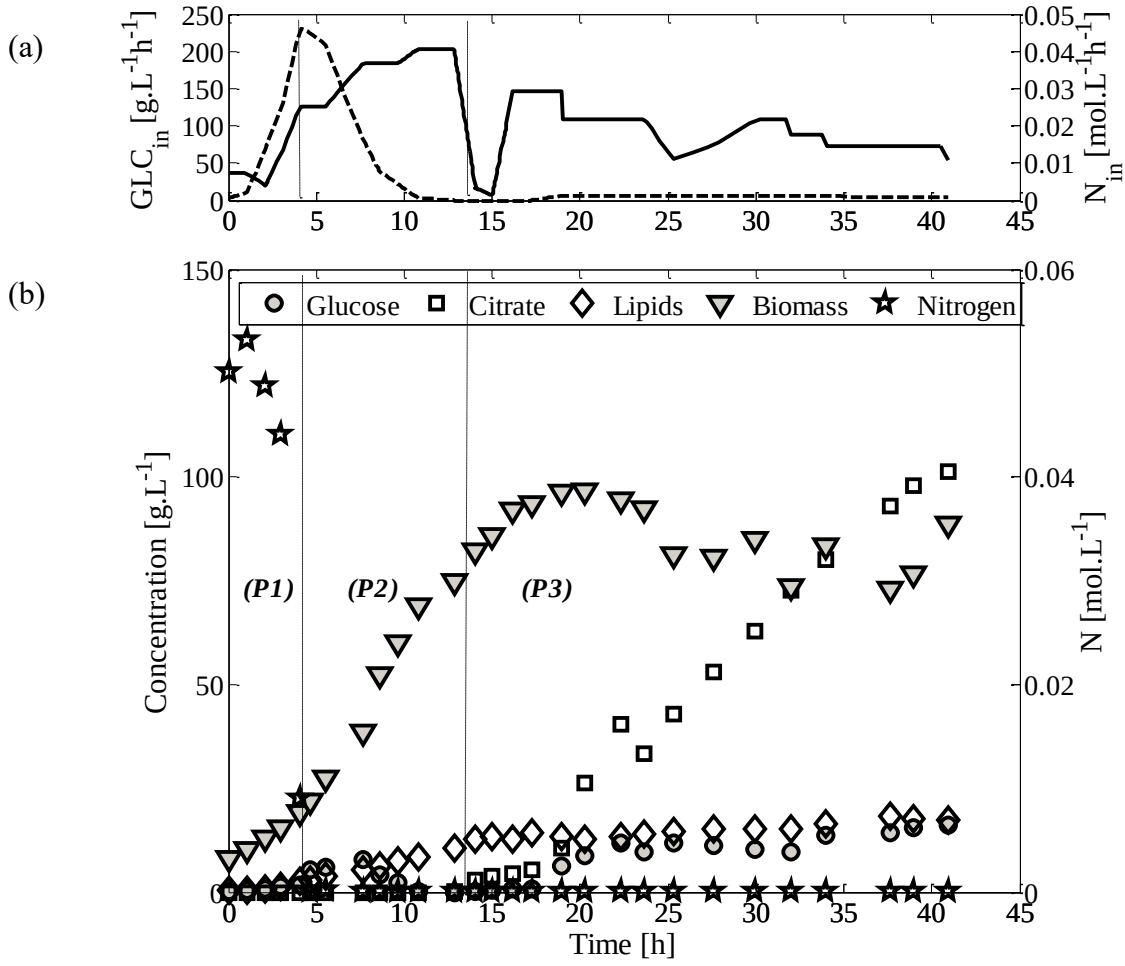


Figure A 0.3. Data of Fed-Batch culture of *Yarrowia lipolytica* on glucose under nitrogen limitation (JC) (Cescut, 2009)**. (a) Input flow rates of (-) Glucose and (--) Ammonia. (b) Profile of the major culture metabolites.**

This culture can be divided into three phases, settled as purposes of modeling, by assuming that the concentrations of intracellular metabolites were at quasi-steady state (Figure A 0.3) (see CHAPTER III, section III).

P1. Growth phase: Glucose was exponentially fed to obtain a constant specific growth rate. In this phase, all nutritional requirements of the yeast were included in the initial medium allowing growth without any limitation. Nitrogen limitation was set up at the end of this phase by shifting from the ammonia solution to a KOH solution for pH regulation.
P2. Lipid accumulation: Carbon feed rate was increased by a pump control algorithm defined in different step changes, whereas nitrogen feed rate presented a detriment in order to modulate the N/C ratio and to provide the needed carbon for residual growth (see Figure A 0.3). In this phase, residual growth occurred simultaneously with lipid accumulation. This was achieved
P3. Citric acid production: Nitrogen feed was at minimum with a controlled carbon feed rate. This caused a linear reduction of the N/C ratio leading to concomitant excretion of citric acid with lipid accumulation.

In Figure A 0.3 (b), nitrogen limitation was fixed at 4 h when the biomass concentration was about 1 $Cmol.L^{-1}$ (23 $g.L^{-1}$). The maximum total lipid concentration (18.2 $g.L^{-1}$) is achieved in phase *P3*. The maximal concentration of citric acid was 120 $g.L^{-1}$ at the end of the culture.

III.2. Cultures under nitrogen deficiency

Two cultures of *Yarrowia lipolytica* with a lipid accumulation phase under nitrogen deficiency were employed in this work. The experiments were performed out by other members of the team FAME at LISBP. The cultures were carried out as sequential batches with glucose pulses without ammonia feeding. Both experiments were performed in a 5 *L* bioreactor (Figure A0.4). In order to clearly identify the two cultures, here both are distinguished as experiment 1 and 2.

With regards on comparing the different experiments, the phases defined in section III.1 were also assumed. However, only two of them were supposed: Phase *P1* or Growth phase where all the nutrients are available in the medium and Phase *P3* where citric acid is produced concomitantly with lipid accumulation.

Figure A0.4. Bioreactor of 5 *L* used for cultures under nitrogen deficiency.

III.2.1. Sequential-Batch Experiment 1

For the first experiment, the bioreactor contained the medium described in section I with a different initial concentration of 0.038 mol.L^{-1} $(NH_4)_2SO_4$. Operating conditions were controlled to a temperature of 28 °C and a pH value of 5.6 regulated by a 5 *M* KOH solution. The bioreactor was monitored with the acquisition software MFCS/win 2.1®. Glucose was fed by pulses at 35 h and 55 h as depicted in Figure A0.5 (a).

In this culture, residual ammonium concentration was not measured. Figure A0.5 (b) shows that lipid accumulation began after 20 h, when citric acid was also produced. The maximum citric acid concentration was attained at the end of the culture (60 g.L^{-1}), whereas total lipid concentration achieved a maximum value near to 5 g.L^{-1} corresponding to a lipid content of 0.133 g.g^{-1}.

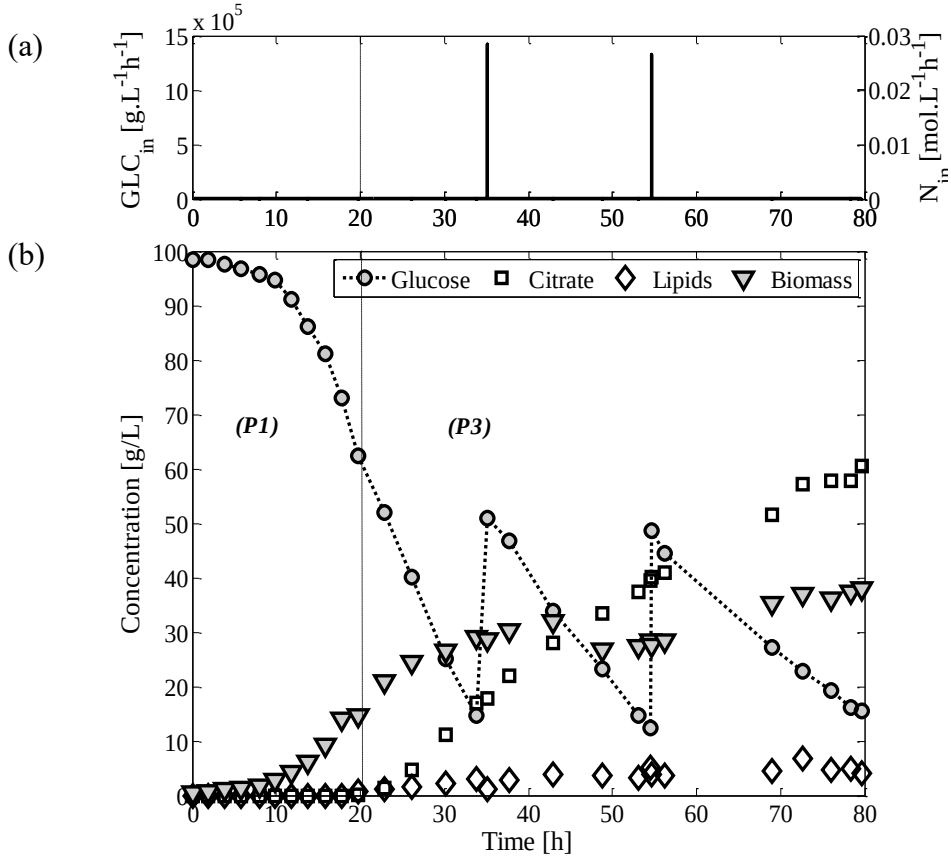


Figure A0.5. Data of Sequential-Batch culture (Exp.1) of *Yarrowia lipolytica* on glucose under nitrogen deficiency (Cochot, 2014). (a) Input flow rates of (-) Glucose pulses and (--) Ammonia. (b) Profile of the major metabolites of the culture.

III.2.2. Sequential-Batch Experiment 2

For the second experiment, an initial concentration of 0.068 mol.L^{-1} $(NH_4)_2SO_4$ was used, which differs from the medium presented in section I. As well as in the experiment presented in section III.2.1, the bioreactor was monitored with the acquisition software MFCS/win 2.1®. Operating conditions were controlled to a temperature of 28 °C and a pH value of 5.6 regulated by a 5 *M* KOH solution.

Only one pulse of glucose solution was used in this experiment at 19.3 h to increase glucose concentration up to 100 g.L^{-1}, as illustrated in Figure A0.6 (a). The maximum lipid content was found at 28 h corresponding to 0.102 g.g^{-1}. The maximum citric acid concentration was around 18 g.L^{-1} at the end of the culture.

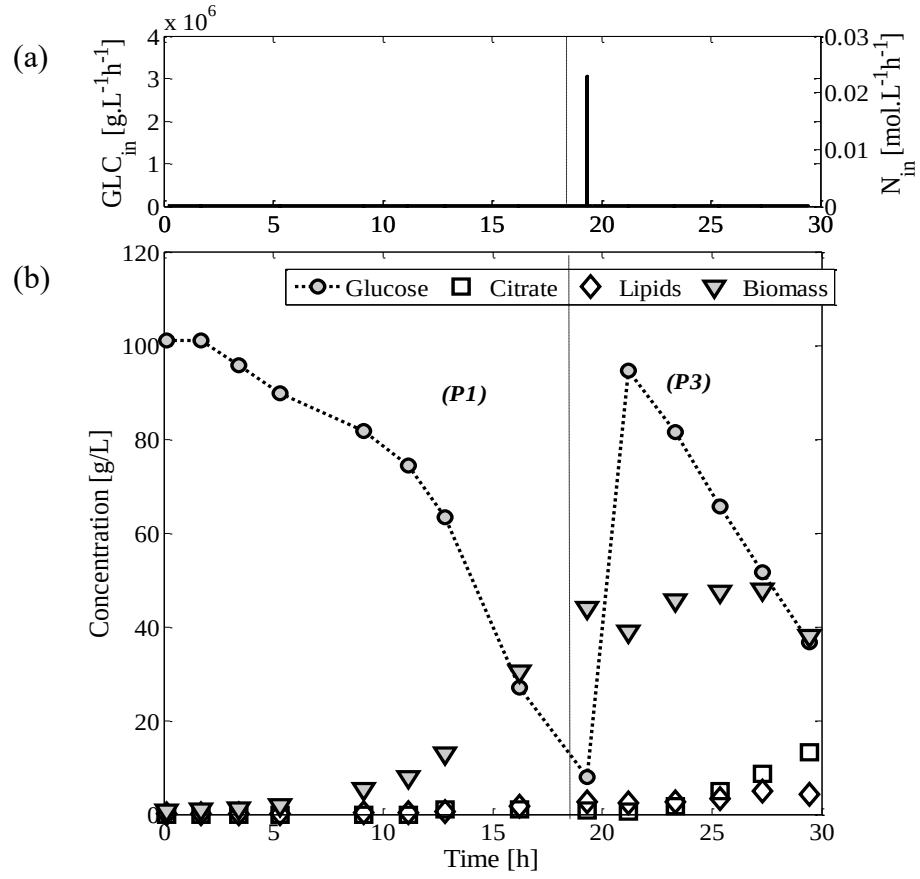


Figure A0.6. Data of Sequential-Batch culture (Exp.2) of *Yarrowia lipolytica* on glucose under nitrogen deficiency (Manson, 2015). (a) Input flow rates of (-) Glucose pulses and (--) Ammonia. (b) Profile of the major metabolites of the culture.

IV. CONCLUSIONS

The proper analysis of these experimental data is the basis of the hypothesis of the works developed within this thesis. The differences among the cultures represent a challenge for modeling lipid accumulation. However, the diversity of the data could extent the application of the models with the aim of including the variation between lipid accumulation under nitrogen deficiency and limitation.

V. REFERENCES

Browse, J., McCourt, P.J., Somerville, C.R., 1986. Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal. Biochem. 152, 141–145. doi:10.1016/0003-2697(86)90132-6

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Ochoa-Estopier, A., 2012. Analyses systématique des bascules métaboliques chez les levures d'intérêt industriel: application aux bascules du métabolisme lipidique chez Yarrowia lipolytica. Institut National des Sciences Appliquées de Toulouse.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

RESUME

RESUME

Dans les dernières années, l'intérêt sur les impacts environnementales due aux émissions de CO2 par l'utilisation des ressources fossiles a augmenté fortement (Fernando *et al*., 2006; Naik *et al*., 2010).. En 2010, 43 % des émissions en CO2 ont été produites à partir de la combustion de combustibles et 36 % à partir de la consommation de charbon (CO_2 Emissions, 2012). Par conséquent, afin de réduire les émissions de gaz à effet de serre, de nombreux pays ont adopté des mesures gouvernementales visant à augmenter la production et l'utilisation des biocarburants (*e.g.* Energy Independence and Security Act of 2007) (Vicente et al., 2009). Ce fait a encouragé la recherche envers le développement de nouvelles alternatives pour la production de biocarburants.

L'industrie de l'aviation a récemment fait enquête sur de nouveaux itinéraires pour produire des carburants renouvelables. Par exemple, l'industrie de l'aviation européenne a encouragé la recherche sur la production de bio-kérosène afin d'atteindre une cible de 2 millions de tonnes de biocarburants produits en Europe d'ici 2020 (European Advanced Biofuels Flight Path Initiative). Cette industrie s'est engagée pour accomplir plusieurs objectifs ambitieux afin d'améliorer sa performance environnementale, y compris une réduction de 50 % en émissions nettes de carbone dans le domaine de l'aviation d'ici 2050 (Babau *et al*., 2013). Le bio-kérosène produit devra être mélangé aux carburants fossiles afin d'abaisser les changements dans les moteurs d'aéronefs (25 ans de durée de vie moyenne) et d'atteindre une performance identique à celle obtenue avec les Jet A1. Parmi les différentes solutions émergentes pour le remplacement des biocarburants certifiés par l'ASTM, les acides gras et les esters hydro transformés (HEFA en anglais) sont de major intérêt (Standard ASTM D1655). Les HEFA sont produites à partir d'huiles qui sont convertis en kérosène paraffinées biosynthétique par hydrotraitement. Ces réactions physico-chimiques sont bien connus et, par conséquent, le défi réside dans la production d'huiles et la rechercher des matières premières (Llamas *et al*., 2012).

Une alternative prometteuse pour la production bio-kérosène est l'utilisation de microorganismes oléagineux. La production des lipides microbiennes comme précurseur des biocarburants (carburants de deuxième génération) présentent de nombreux avantages sur les huiles végétales car ils ne nécessitent pas de terres arables, et cette production n'est pas en compétition avec des aliments puisque la biomasse peut être utilisé comme source de carbone (Vicente *et al*., 2009). Les microorganismes dites oléagineux peuvent accumuler une haute

teneur en lipides (plus de 20% w/w) (Ratledge, 2004) parmi lesquels, les champignons et les levures sont les meilleurs candidats. Ces deux ont des rendements de conversion élevés et une forte teneur en lipides ayant profils semblables aux celles de biocarburants commerciales (Meng et al., 2009).

Dans ce contexte, ce travail de thèse a été réalisé dans le cadre du projet Probio3 qui s'intéresse à la production biocatalytique de bio-lipides à partir des ressources renouvelables. Le projet Probio3 envisage de créer une filière qui a pour ambition le développement intensive de la production de lipides pour qu'il soit robuste et de performances reproductibles afin de réduire les impacts environnementales causés par les émissions de CO_2. Ces lipides produits par des microorganismes oléagineux peuvent être des précurseurs pour la production de biocarburants et produits de la chimie verte.

Yarrowia lipolytica est une des levures oléagineuses étudies dans ce projet car elle est capable d'accumuler jusqu'à 36% de sa propre masse en lipides sous la forme de triglycérides avec une composition prometteuse pour la production de biocarburants (Ratledge and Wynn, 2002). *Yarrowia lipolytica* a été étudiée dans les procédés de fermentation sous diverses conditions d'exploitation en essayant d'atteindre des hautes quantités de lipides (Beopoulos et al., 2009). Cette levure peut être utilisée pour la production de surfactants chimiques, acides organiques et des lipides à partir de différentes substrats comme le glucose, le glycérol, les acides gras, entre autres. Vu que son génome est entièrement séquencé (Barth and Gaillardin, 1997), l'ingénierie métabolique chez cette levure peut être obtenu facilement pour optimiser la production de lipides. L'accumulation des lipides chez *Yarrowia lipolytica* est déclenchée par un épuisement de l'azote et/ou un excès de la source carbonée. Toutefois, le rendement de conversion des lipides est entravé par la production de l'acide citrique.

Dans le cadre du projet, l'un des objectifs principaux est de diriger les flux du carbone vers les lipides pour atteindre une maximisation qui englobe trois critères : la productivité, la teneur en lipides et le rendement lipide/substrat. Ces critères peuvent être aboutis en utilisant plusieurs alternatives comme:

- le génie génétique pour le développement des souches modifiées dont divers travaux ont été faits (Friedlander et al., 2016; Ledesma-Amaro and Nicaud, 2016; Nicaud, 2012; Rakicka et al., 2015).,

- le génie de bioprocédés dont différentes modes de cultures ont été testés (Ochoa-Estopier and Guillouet, 2014; Papanikolaou et al., 2006; Papanikolaou and Aggelis, 2011), et
- la commande de bioprocédés où peu de travaux ont été reportés sur l'accumulation des lipides par les levures.

Cette thèse essaye de compléter la dernière. Pourtant, ce travail se focalise sur la modélisation, le suivi des variables d'état (capteurs logiciels) et la commande. Dans ce travail, nous nous intéressons d'abord à l'élaboration d'un modèle capable de décrire les basculements métaboliques de *Yarrowia lipolytica* ayant un bon compromis entre le degré de description du métabolisme et l'applicabilité du modèle. Ensuite, nous nous consacrons à la reconstruction des variables d'état par des capteurs logiciels (software sensors), pour finalement proposer et évaluer des stratégies de commande. La Figure R0.1 décrit la méthodologie suivi dans ce travail de thèse.

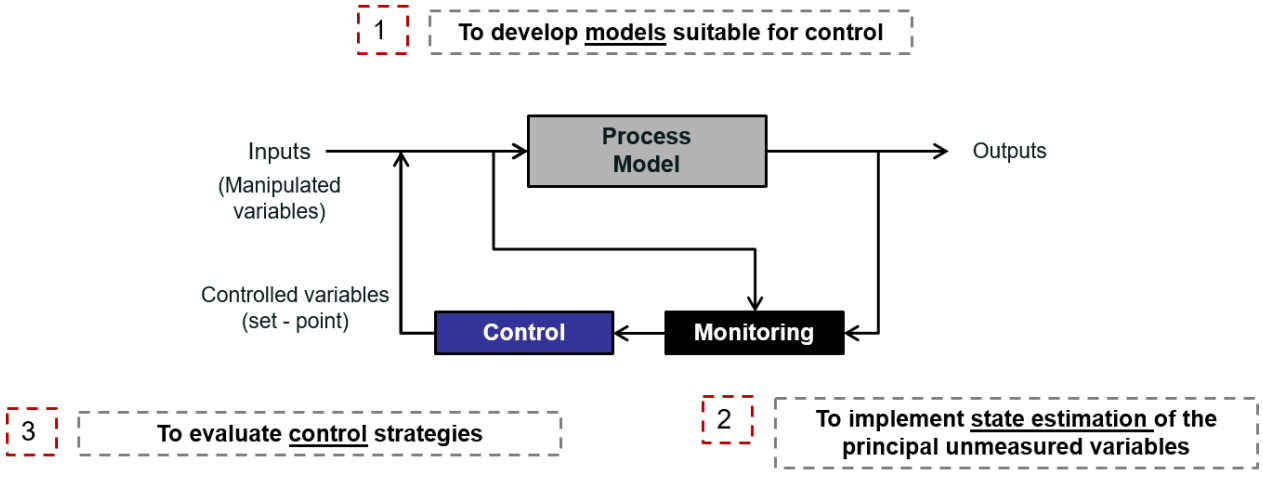


Figure R0.1. Stratégie de cette thèse.

Ce travail de thèse est divisé en quatre chapitres. Le premier chapitre est une analyse bibliographique de la production de biocarburants qui peuvent être produite par des micro-organismes oléagineux comme les levures. Un aperçu général du métabolisme avec un intérêt particulier sur l'accumulation des lipides est présenté dans une deuxième section de ce chapitre.

Le deuxième chapitre examine les modèles existants qui décrivent l'accumulation de lipides et introduit l'approche de modélisation métabolique dynamique. La modélisation métabolique est

divisée en deux méthodologies: l'analyse métabolique statique et l'adaptation dynamique du métabolisme. Différentes approches sont présentées avec un intérêt spécial sur la manière de gérer les basculements métaboliques. Dans ce chapitre, nous comparons plusieurs stratégies de contrôle commande qui sont proposé pour optimiser les bioprocédés. Enfin, plusieurs capteurs logiciels, utilisés pour la surveillance en ligne, sont décrites pour connaitre ses avantages et inconvénients.

Le troisième chapitre contient les principaux résultats de ce travail de thèse. Il est formé par cinq sections présentées comme des publications. La première section porte sur une nouvelle méthodologie pour la réduction des réseaux métaboliques afin d'utiliser les avantages d'un modèle dynamique. La deuxième section développe trois modèles dynamiques basées sur un réseau métabolique réduit. La méthodologie utilisée pour deux de ces modèles a été déjà rapporté dans la littérature, tandis que le troisième est une nouvelle proposition. La troisième section présente le développement de deux modèles macroscopiques, l'un d'eux est appliqué dans une stratégie de commande pour optimiser la productivité des lipides et le contenu lipidique. La quatrième section expose une stratégie de commande basée sur le modèle prédictive multi-objectif (Pareto) qui est appliquée à l'accumulation de lipides. La loi de commande utilise un modèle métabolique dynamique pour prédire les objectifs contradictoires. La cinquième section montre un capteur logiciel développé pour prédire la dynamique en ligne des trois variables d'état.

Le dernier chapitre englobe la discussion des résultats obtenus et il propose quelques perspectives sur la poursuite des travaux à faire. Une annexe décrit le matériel et les méthodes pour l'obtention de données expérimentales utilisées tout au long de ce travail.

I. ACCUMULATION DES LIPIDES

Ce chapitre est une revue bibliographique consacré à l'accumulation lipidique avec plusieurs sections. La première section montre une analyse de la production de biocarburants produits par les micro-organismes oléagineux, comme la levure. Dans une deuxième section de ce chapitre, un aperçu général du métabolisme est présenté avec un intérêt particulier sur l'accumulation des lipides.

La Figure R0.2 montre une voie métabolique réduite où les substrats sont transformés dans le précurseur Acétyle-CoA à partir duquel, sous une limitation nutritionnelle et un excès du carbone, différent produits peuvent être produits comme l'acide citrique et les lipides. Les lipides produits peuvent être classés comme structuraux et accumulés. Cette différence est visualisée dans la droite de la Figure R0.2 dont la biomasse est divisé en deux parties: une appelée biomasse catalytique qui contient les lipides structuraux, et une autre avec les lipides accumulés qui sont majoritairement composés par des triacyleglycérols (TAGs). Alors, la biomasse totale (cellule) est définie comme la somme de la biomasse catalytique et les lipides accumulés.

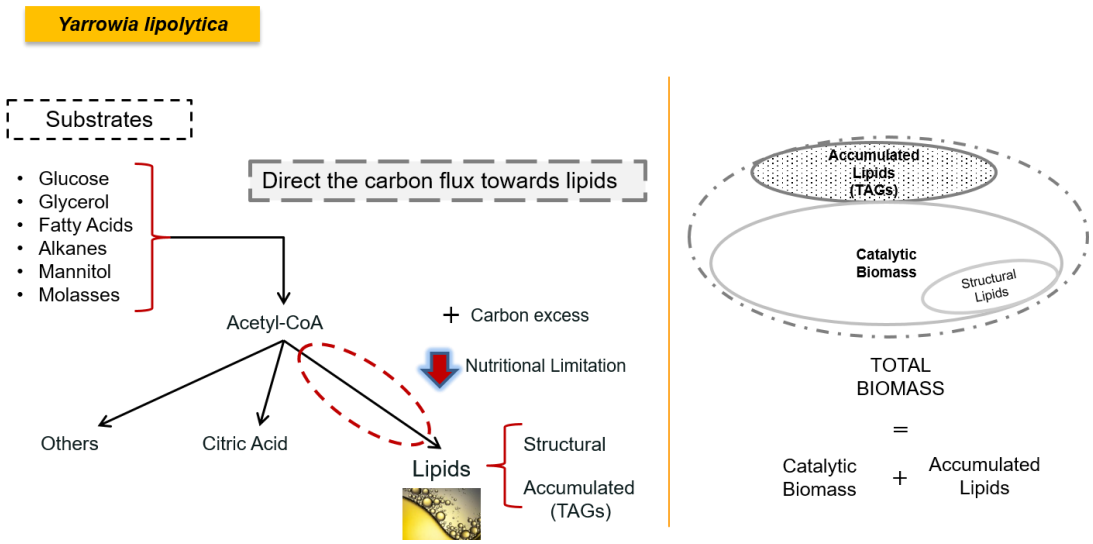


Figure R0.2. Résumé des voies métaboliques chez *Yarrowia lipolytica*.

Dans ce premier chapitre, un réseau métabolique réduit est aussi proposé prenant en compte 2 compartiments: le cytosol et la mitochondrie. Ce réseau est utilisé dans les chapitre III dans des approches de modélisation.

Une étude sur les fermentations avec *Yarrowia lipolytica* a rapporté que l'accumulation de lipides est déclenchée par une limitation nutritionnelle. Cela pourrait conduire à la production d'autres sous-produits tels que l'acide citrique. La production de l'acide citrique a été augmentée lorsque la source carbonée a été le glucose. L'utilisation d'un co-substrat (par exemple glycérol, stéarine) était efficace pour réorienter le métabolisme vers l'accumulation de lipides.

Plusieurs facteurs influençant l'accumulation de lipides ont été présentés. Parmi ces facteurs, la source de carbone (*i.e.* le glucose), le pH, et la température ont été fixés dans ce travail de thèse, et donc l'aspect principaux à maitriser était le rapport N / C. En outre, les cultures continues étaient les plus efficaces en termes de teneur en lipides (gLIP.gX-1) parmi les différentes conditions de culture ont été examinées. Les cultures pré-alimentées avec du glucose, par exemple, étaient meilleures pour fournir à la fois une bonne productivité et un teneur lipidique, ce qui sont deux des critères importants à prendre en compte dans l'optimisation de la production de lipides. La culture en fed-batch a été choisi parce qu'elle permet au même temps de récupérer les cellules accumulant des lipides avec un bon contrôle de la contamination du milieu et la maîtrise du rapport N / C, signalé comme le facteur le plus influant sur l'accumulation de lipides.

II. LA MODELISATION ET L'OPTIMISATION DES BIOPROCEDES

Dans le chapitre II, les modèles existants pour l'accumulation de lipides sont décrits. Puis, les outils d'optimisation où différentes stratégies de contrôle sont comparées. Enfin, plusieurs capteurs logiciels, utilisés pour la surveillance en ligne, sont décrites afin d'analyser leurs avantages et inconvénients.

Il est à noter que dans la littérature un seul modèle a été utilisé pour décrire les lipides et l'acide citrique. Dans ce modèle, la production d'acide citrique a été supposée être la conséquence d'un changement de substrat en glucose, qui a causé un débordement de carbone métabolique (excès de carbone) (Papanikolaou et al., 2006). Même si la construction du modèle est intéressante pour l'application, les résultats n'ont montré qu'une performance adéquate pour les lipides structuraux, car les expériences rapportées n'ont pas signalé de lipides accumulés lorsque le glucose était utilisé comme substrat. Ceci motive le développement des modèles plus performantes permettant de décrire la production couplée des lipides et de l'acide citrique.

Une grande partie du chapitre II examine l'approche de modélisation métabolique dynamique, qui est considéré en deux étapes : l'analyse métabolique statique, et l'adaptation dynamique du métabolisme. Différentes approches sont présentées avec un intérêt spécial sur leurs manières de gérer les changements métaboliques, ces approches s'appuient sur l'hypothèse de régime quasi-permanent. Cependant, certains métabolites pourraient être considérés comme accumulés

en modifiant certaines équations (voir la section II.3.3). Ce fait est très intéressant pour la modélisation des lipides. Par conséquent, cette approche de modélisation pourrait être utilisée pour coupler le métabolisme au comportement dynamique des métabolites primaires vers l'optimisation de la production de lipides.

En ce qui concerne l'optimisation des procédés, plusieurs types de commande ont été analysés, remarquant que les fermentations nécessitent de l'optimisation de plusieurs objectifs (par exemple productivité, rendement, temps). Par conséquent, l'optimisation multi-objective est utilisée dans ce travail (voir CHAPITRE III.V). Cependant, ce type d'optimisation trouve quelques défis, spécialement lors de la pondération des objectifs. Les algorithmes basés sur le front de Pareto ont été présentés comme de bonnes alternatives, dont le MOPSO a été mis en évidence grâce à sa rapide convergence et à sa grande distribution de l'essaim. Cet état de l'art encourage la proposition d'une commande multi-objective pour optimiser la productivité des lipides et la fraction de contenu lipidique.

Enfin, nous montrons l'importance de surveiller en ligne les états pour optimiser les cultures. Les capteurs logiciels basés sur des modèles, également connus sous le nom d'observateurs, ont été décrits avec leurs nombreuses applications aux bioprocédés. Bien que leur utilisation, ils dépendent d'un modèle bien identifié et observable, et leur vitesse de convergence est définie sur les conditions de fonctionnement. D'autre part, les capteurs logiciels basés sur des données ont gagné en popularité sur des applications industrielles où SVM a été souligné comme un algorithme intéressant pour son simplicité. Il est donc possible de croire que cette méthode pourrait être directement applicable à l'estimation en ligne des variables les plus importantes impliquées dans les fermentations lipidiques.

III. RÉSULTATS

Ce chapitre est présenté sous forme de publications dont la première section résume les cinq publications et finalement.

La première publication montre une analyse d'un réseau métabolique à l'aide des modes élémentaires. Dans cette section, nous avons proposé une méthodologie visant à réduire le nombre de modes élémentaires, avec un compromis entre la quantité et la qualité de la

réduction. Cette réduction est réalisée grâce à l'analyse convexe dans le domaine de rendement où les points dites extrêmes peuvent représenter tous les points qui sont à l'intérieur de ce cône convexe. Les modes élémentaires sont normalement représentés par des enveloppes convexes dont les modes qui sont autour cette enveloppe sont appelés de modes extrêmes. Donc, nous avons proposé l'utilisation de polygones avec trois, quatre et cinq modes élémentaires extrêmes. Différents cas ont été considérés lorsqu'on disposait de données expérimentales ou non. Les différents polygones ont été comparés avec des données de la littérature pour valider la méthode. Nos résultats ont démontré qu'aucune différence significative n'est affichée parmi les polygones proposés. Par conséquent, le plus petit polygone, qui dans ce cas était le triangle, était préférable pour cette méthode.

L'objectif de cette thèse était l'optimisation de l'accumulation des lipides par *Yarrowia lipolytica*, pour cela trois modèles dynamiques ont été proposés dans la deuxième publication. Dans ces modèles, trois phases ont été définies dans le but de modéliser les phénomènes qui produisent lors de l'accumulation des lipides, III.2.4 du chapitre I. Les trois modèles partagent une structure commune décrite par les bilans de matières et les cinétiques, mais elles divergent dans la manière de réguler le métabolisme, cette-à-dire le basculement entre chaque phase métabolique. Les modèles utilisés ont été le Modèle Hybride Cybernétique (HCM) qui utilise des enzymes cybernétiques pour réguler chaque phase; le Modèle de Bioréactions Macroscopiques (MBM) qui considère que une phase est liée à une autre par un activation-désactivation fixée; et une nouvelle approche appelée Fuzzy MBM dans lesquels les basculements sont décrits par des règles de la logique floue qui prennent en compte le ratio d'absorption de N/C.

Les trois modèles ont été calibrés, validées et comparées par des critères statistiques, comme l'AIC, prouvant que le modèle Fuzzy MBM et HCM ont des résultats comparables. Ces deux modèles peuvent être utilisés pour prédire d'autres informations liées à une fermentation, comme le quotient respiratoire, le ratio de N/C, et la dynamique des flux métaboliques d'un réseau métabolique.

Deux modèles macroscopiques ont été proposés en parallèle aux modèles métaboliques de la dernière dynamique section. Le premier était un modèle basé sur des cinétiques de type Monod, appelé "non structurée " et le deuxième était basé sur le modèle de Droop (modèle quota). Ce dernier contient une variable structurée (le quota) pour tenir compte des rendements dépendant

du temps. En lieu d'utiliser des phases, ces deux modèles considèrent le concept de 'overflow' pour tenir compte de la production concomitante d'acide citrique et des lipides. Ce phénomène a été identifié par un indicateur booléen qui a été calculé sur des cinétiques. Dans ce travail, un problème d'optimisation est aussi traité visant à maximiser deux critères: la teneur en lipides et la productivité. Les apportes d'azote et du carbone ont été les variables sur lesquelles la commande a réagis pour gérer le rapport N/C en approximant la solution par de fonctions linéaires par morceaux. Trois cas d'étude sont été testés : un avec un débit constant et deux avec 2 et 5 fonctions linéaires par morceaux. Les résultats de la simulation sont montrés dans le Table R0.1. Il est apprécié que l'optimisation permettre de obtenir des productivités jusqu'à 0.95 $g.L^{-1}.h^{-1}$ et des teneurs de 0.23 $g.g^{-1}$. Ces résultats sont prometteurs par rapport aux valeurs de la littérature.

Table R0.1. Performance of the control strategy

	t_f **(h)**	$V(t_f)$ **(L)**	$CIT(t_f)$ **($g.L^{-1}$)**	$X_L(t_f)$ **($g.L^{-1}$)**	$X_f(t_f)$ **($g.L^{-1}$)**	$P_L(t_f)$ **($g.L^{-1}.h^{-1}$)**	$w_L(t_f)$ **($g.g^{-1}$)**	$w_P.P_L + P_W$
Constant flow rate	20.38	11.39	3.26	16.87	74.85	0.827	0.192	0.783
2 PWL functions	27.04	14.02	0.86	26.93	105.88	0.996	0.212	0.888
5 PWL functions	30.25	14.99	0.39	30.36	113.56	1.004	0.221	0.914

En ce qui concerne l'optimisation de l'accumulation des lipides, les deux critères à maximiser sont souvent contradictoires et donc il est nécessaire d'attribuer une pondération adéquate à chacun des objectifs. La quatrième publication porte sur une stratégie pour traiter ce problème en utilisant les solutions des fronts de Pareto non dominées calculés depuis l'algorithme d'essaim particulier (PSO). Néanmoins, le calcul des fronts de Pareto est statique, et d'ailleurs PSO ne tient pas compte l'utilisation de contraintes qui sont nécessaires pour une mise en œuvre réaliste. Par conséquent, le présent article fait un couplage de Multi-Objective PSO avec un algorithme de recherche pattern pour rendre l'optimisation dynamique. Une modèle réduite depuis le HCM est aussi proposé pour diminuer le temps de calcul des fronts de Pareto.

Dans cette publication, également que l'ultérieur, l'optimisation a visé à trouver les meilleures trajectoires pour les apportes des sources de carbone et d'azote afin de maximiser la teneur

lipidique et sa productivité. Cette stratégie de commande a été effectuée en deux étapes: d'abord en favorisant la croissance de la levure où la maximisation de la biomasse est atteinte. Dans un deuxième temps, la levure accumule des lipides. Trois cas d'étude ont été analysés, le premier dans lequel la fermentation continue avec le même débit que celui qui était utilisé pendant la croissance, et les autres deux propose une commande modèle prédictive avec deux algorithmes: pattern search et notre proposition MOPSO pattern search. Les résultats de l'optimisation sont montrés dans le Table R0.2 où on constate que le cas des débits constants n'est pas optimisé. L'utilisation d'une pondération variable par MOPSO-Pattern a permis d'obtenir une amélioration de 36% à niveau de la productivité, mais juste un 4% pour la teneur lipidiques qui est reflétés dans un incrément de 39% dans la production de lipides, ce qui démontre l'effectivité et l'importance de la pondération variable.

Table R0.2. Comparison of Objectives in the control strategy

Variable	P_L (g/h)	w_L (g/g)	V (L)	LIP (g)	$t_{f,growth}$ (h)	$t_{f,prod}$ (h)
Growth	1.1	0.05	10.55	23.4	4.5	21.5
Pattern Search	7.68	0.188	12.14	160.6	4.5	21
MOPSO Pattern	10.43	0.196	13.40	223.3	4.5	21.5

La dernier publication se consacre au développement des capteurs logiciels pour le suivie des variables d'état. Dans cet article, trois capteurs logiciels indépendants du modèle ont été proposés afin d'éviter l'ajout des incertitudes du modèle à la boucle de commande. Les capteurs logiciels ont été basés sur la méthodologie de Machines à Vecteurs de Supporte (SVM) qui ont été optimisés par un algorithme de PSO. SVM est une méthodologie de l'intelligence artificielle fondée sur l'apprentissage des machines qui reposent sur la qualité et la disponibilité des données. Afin de prouver leur efficacité, nous avons utilisés deux jeux de données indépendantes: un qui était utilisé pour l'apprentissage, tandis que l'autre était utilisé pour la validation. Les résultats ont démontré un bon rapport entre simplicité et la qualité de l'ajustement. Même si les expériences sont nécessaires pour valider cette méthode, leurs résultats ont prouvé l'efficacité pour assurer le suivi des lipides, la biomasse et l'acide citrique.

IV. CONCLUSIONS ET PERSPECTIVES

L'objectif principal de cette thèse était d'optimiser la production de lipides microbiens qui sont considérés comme précurseurs pour la production de biocarburants d'un point de vue de la commande automatique. Dans ce contexte, cette thèse a porté sur trois axes principaux : la modélisation, le suivi des variables d'état et la commande. Ces axes ont été examinés dans l'analyse bibliographique avec un intérêt particulier sur la modélisation dynamique du métabolisme, qui était un élément essentiel de cette thèse.

Cinq modèles ont été proposés, dont deux modèles macroscopiques et trois modèles métaboliques dynamiques (DMM). L'hypothèse principale des modèles dans notre étude était fondée sur le fait que pendant une fermentation avec des levures oléagineuses trois phénomènes peuvent apparaitre: la croissance, l'accumulation des lipides et la production d'acide citrique. Pour cela, nous avons proposé un réseau métabolique réduit comprenant 59 Réactions et 51 métabolites, où une attention particulière devrait être portée à l'équation de la biomasse catalytique (Biomasse sans lipides accumulés). Le réseau métabolique a été décomposé en modes élémentaires (EMs) basés sur l'hypothèse de régime quasi permanent (QSSA). Le nombre d'EMs, étant grande (1944 EMs) pour être utilisé, a été réduit grâce à une méthode de réduction qui maximise un polygone autour de l'enveloppe convexe représentant la solution. En outre, il a été prouvé que le plus petit polygone (triangle) pourrait représenter correctement l'ensemble réseau métabolique.

La différence la plus grande entre les trois modèles métaboliques dynamiques proposées dans cette thèse a été la manière dans laquelle ils effectuent les basculements métaboliques. Il est à noter que HCM était le seul modèles qui tient en compte la régulation cellulaire, ce qui est fait par les "enzymes cybernétiques ", qui a permis à connaître dynamique de chaque EMs sélectionné. Néanmoins, ces "enzymes" ont l'absence de sens biologique et elles ajoutent au modèle un important nombre de paramètres à estimer. MBM, en quant à lui, a démontré d'être une approche simple pour reproduire les données expérimentales et pour identifier les cinétiques, mais il a présenté mauvaise généralisation à des fins de contrôle en raison de sa dépendance temporelle pour l'identification de changements métaboliques. Finalement, Fuzzy MBM, étant une nouvelle proposition de cette thèse, a maintenu les bonne caractéristiques qui MBM disposait tout en mettant en place des changements dynamiques entre les phases à l'aide

de la logique floue et du rapport de l'absorption d'azote et de glucose des taux (rN/rC). La performance de ces modèles était acceptable pour les trois cultures différentes de *Yarrowia lipolytica* à partir de glucose. Les avantages de ces modèles sont la reconstruction d'autres métabolites ou des autres indicateurs (*i.e.* le quotient respiratoire RQ) sans avoir des équations supplémentaires, ils sont aussi capables de reproduire les flux métaboliques ce qui pourrait être utile pour identifier les voies actives afin de sélectionner des cibles en ingénierie métabolique. En outre, les modèles métaboliques dynamiques ont monté des excellents performances outils pour l'amélioration de la productivité (Song et al., 2014). Par contre, ses inconvénients sont la généralisation vers des souches mutantes, ce qui pourrait impliquer la suppression de certaines réactions considérées dans notre réseau métabolique.

En ce qui concerne les modèles macroscopiques, nous avons présenté un modèle non structuré basé sur des équations cinétiques de Monod et un modèle construit sur le modèle de quota de Droop (Droop, 1968). Il est important à noter que le modèle de Droop avait été utilisé que pour les microalgues. Toutefois, dans le cadre de ce travail, il a été prorogé pour la levure ce qui justifie l'originalité du travail. Les deux modèles ont été calibrés et validé avec trois jeux des données indépendantes. Puis, ce modèles ont été capables de représenter le basculement métaboliques vers l'overflow qui produit l'acide citrique. Cet effet a été modelé par l'ajout des indicateurs Booléens. Même si les deux modèles ont présenté des résultats comparables, le modèle quota a mieux décrit la dynamique de l'azote et du glucose ce qui l'a fait un meilleur choix pour des objectives de contrôle commande. En plus, nous avons démontré les bons habilités de prédiction du modèle quota pour reproduire des expériences de la littérature et sa robustesse pour s'adapter aux souches mutantes.

En ce qui concerne la surveillance des variables d'états, notre proposition a été basée sur les capteurs logiciels obtenus par de Machines à Vecteurs de Support (SVM) (Hsu et al., 2008; Vapnik et al., 1996). La nouveauté de notre proposition est non seulement l'implémentation sur les fermentations lipidiques, mais aussi l'ajout d'un algorithme d'optimisation globale comme l'essaim de particules (PSO) pour sélectionner les paramètres caractéristiques de SVM basée sur le calcul de l'AIC. L'obtention de résultats exacts dépend fortement sur le jeu de données de formation, qui doit contenir un large éventail de conditions d'exploitation pour fournir une meilleure généralisation de la solution. Cette dernière pourrait être considérée comme un inconvénient qui est évité lorsqu'un grand nombre de données est disponible. Par conséquent, la quantité de données et sa qualité pourraient entraver l'utilisation de cette méthode. Nos

résultats pour estimer les concentrations des lipides et la biomasse ont prouvé que la méthodologie que nous proposons de combiner PSO-SVM est une méthode d'estimation efficace et prometteuse. Cet algorithme est rapide sur la mise en œuvre, où 7 minutes ont été nécessaires pour l'apprentissage, alors que la validation a été effectuée dans 0.05 secondes. Cette méthodologie peut être utile pour l'application en temps réel où des corrections des estimations pourraient être réalisées in situ. Un dernier avantage de cette méthode est sa capacité d'extrapolation pour d'autres microorganismes ou des souches modifiées vue qu'elle ne dépende pas des modèles.

L'optimisation de l'accumulation lipidique qui entraine la maximisation de trois critères: la teneur lipidique, la productivité et le rendement. Dans le cadre de cette thèse les deux premiers ont été utilisés. Ces deux critères ne sont pas nécessairement proportionnels et sont souvent contradictoires. Des études théoriques ont été élaborés pour optimiser les critères avant mentionnées à travers de la définition de trajectoires pour les débits d'alimentation optimale de glucose et d'azote. Ces deux apports ont été sélectionnés pour moduler le ratio N/C qui définit les basculements métaboliques de l'accumulation de lipides en *Yarrowia lipolytica*. L'optimisation multi-objective a été traitée dans plusieurs manières avec deux modèles.

Tout d'abord, nous avons utilisé le modèle de quota pour optimiser la productivité de la fraction lipidique, et par l'intermédiaire de fonctions linéaire par morceaux. Dans ce cas, les deux objectifs ont été pondérés en fonction d'un débit constant de optimale les deux variables d'entrée. Pour le deuxième cas d'étude, nous avons utilisé le modèle cybernétique hybride (HCM). L'optimisation a été divisée en deux phases basées sur la construction des modèles dynamiques métaboliques. Dans une première phase, la croissance a été optimisée pour assurer une concentration de biomasse catalytique (biomasse libre de lipides). Dans une deuxième phase, la teneur lipidique et la productivité ont été optimisées en utilisant une fonction objective pondérée en permanence. Les poids ont été obtenus depuis le calcule dynamique des fronts de Pareto. La méthodologie utilisée pour l'optimisation basée sur Pareto est une nouvelle proposition de cette thèse, qui effectue une recherche dynamique de l'optimum de l'ensemble du Pareto (c'est-à-dire le calcul du front de Pareto à chaque étape). Cela a été accompli par l'accouplement d'un algorithme d'essaim de particules Multi-Objective sans contraintes (MOPSO) pour générer l'ensemble de points optimales (Pareto front) avec un algorithme de recherche de Pattern (PS) qui es un optimiseur dynamique qui tient en compte les contraintes du système. L'utilisation de cette méthode a permis la maximisation des deux critères

simultanément et les traiter avec le même degré d'importance. La robustesse de notre méthode a été testée en simulation en réduisant le glucose concentration d'alimentation de moitié, où les résultats ont montré une réponse adéquate à la perturbation. Toutefois, d'autres perturbations pourraient être ajoutées comme les changements des concentrations d'apport en azote ou les concentrations pourraient être également considérés comme des autres variables manipulées. L'un des inconvénients de ce méthodes était le temps de calcul ce qui a été résolu avec une réduction de modèle qui a permis de raccourcir le temps de calcul à 80% du temps réel.

En ce qui concerne les résultats des stratégies de contrôle, l'utilisation des différents modèles a donné des résultats similaires. Dans le premier cas qui a considéré le modèle quota, nous avons obtenu une productivité volumétrique de 0,95 g. $(L\text{-}h)^{-1}$ avec une fraction lipidique de 0,23 $g.g^{-1}$ et une concentration négligeable de l'acide citrique après 30,5 h. Pour les deuxième cas avec le modèle HCM, la productivité volumétrique a été de 0,778 g. $(L\text{-}h)^{-1}$ et la teneur lipidique de 0,196 $g.g^{-1}$ obtenus en 21,5 h avec 5 $g.L^{-1}$ d'acide citrique. Il convient de noter que lorsque HCM a été utilisée, une perturbation a été considérée: la réduction de l'alimentation de la concentration de glucose par la moitié. Cette perturbation a été maintenue pour 8 heures, plus d'un tiers de la durée de fonctionnement, ce qui pourrait avoir une incidence sur les résultats. Même si nos résultats sont théoriques, ils sont assez prometteuses compte tenu que les résultats obtenus dans la littérature avec la même souche et les même conditions initiales ont saisi 0,25 $g.g^{-1}$ en teneur lipidique et 0.23 $g.(L\text{-}h)^{-1}$ en productivité volumétrique avec 120 $g.L^{-1}$ d'acide citrique après 60 h de fermentation (Cescut, 2009).

Enfin, il important de noter que cette thèse est la première à proposer des modèles pour décrire l'accumulation des lipides et la production d'acide citrique basé sur le métabolisme. En outre, cette thèse fournit les premières indications vers l'optimisation de l'accumulation de lipides par la commande automatique.

De notre point de vue, ce travail conduit à plusieurs perspectives afin d'améliorer et de compléter les trois axes de cette thèse de doctorat: la modélisation, la surveillance d'états et la contrôle. La prochaine étape naturelle vers l'optimisation est la validation expérimentale de la stratégie de commande. Cependant, cela comporte plusieurs tâches à être accomplies antérieurement.

Concernant la modélisation, les modèles dynamiques métaboliques sont capables de prédire les flux intracellulaires, ce qui pourraient être utilisé pour des applications de l'ingénierie métabolique (Sauer, 2006). Toutefois, il serait important de comparer la distribution des flux obtenus avec des modèles dynamiques avec des méthodes de marquage au 13C (omics). Cela permettra à corroborer notre réseau métabolique pour souligner l'importance de ces modèles.

D'autre part, il serait intéressant de comparer les résultats du modèle avec une matrice stœchiométrique plus complète, comme celle utilisé par Pan and Hua, (2012). En outre, cela pourrait également contribuer à vérifier si la distribution de flux est affectée par la réduction. Enfin, les modèles dynamiques métaboliques (HCM et Fuzzy MBM) pourraient être étendu aux souches mutantes de *Yarrowia lipolytica* et l'utilisation d'autres substrats (*i.e.* le glycérol, la xylose). Ceci permettra de tester la robustesse du modèle fournissant une version généralisée avec de meilleures capacités d'estimation.

Des capteurs logiciels supplémentaires doivent être élaborés pour estimer d'autres variables d'état (*i.e.* concentrations d'alimentation, le glucose et les titres d'azote) qui sont également nécessaires pour être incorporés dans le modèle avant l'optimisation. Cet objectif pourrait être atteint par l'ajout d'autres mesures en ligne tels que le RQ, la composition du gaz, et des mesures demie en ligne comme la densité optique (DO). En outre, l'utilisation d'une plus grande quantité de mesures pourrait promouvoir une meilleure estimation des états. Puis le couplage de la commande et les capteurs logiciels doivent être validés avec des simulations pour trouver le délai entre le temps de décision et d'implémentation par des actionneurs. Enfin, la mise en œuvre expérimentale de notre stratégie de commande couplée aux capteurs permettra de valider nos propositions.

Autres améliorations pourraient être effectuées pour la stratégie de commande, comme l'inclusion d'une méthode de classification pour sélectionner le point optimal dans le front de Pareto. Cela va prendre en charge la sélection de la trajectoire optimale. Ceci pourrait être effectué par l'ajout d'un algorithme de prise de décision comme la théorie des multi-attribue.

En ce qui concerne la surveillance, une fusion des capteurs pourrait être envisagée pour évaluer les différentes combinaisons de variables effectués par SVM. Cette méthodologie pourrait être couplé à la détection de défauts afin d'assurer estimation des états tout au long d'une fermentation.

Les résultats présentés dans cette thèse fournissent certaines idées pour la compréhension des fermentations lipidiques avec *Yarrowia lipolytica*. Ces travaux pourraient aussi motiver la réalisation d'autres expériences où la dynamique transitoire de métabolites internes pourrait être suivie par d'outils d'analyse omics (*e.g.* fluxomiques).

Cette thèse a également contribué à l'étude d'optimisation par la commande des procédés. On croit que la synergie des stratégies de commande proposées et les capteurs logiciels associés aux souches d'ingénierie métabolique pourraient fournir la meilleure alternative vers l'optimisation de la production de lipides comme précurseurs des biocarburants.

V. RÉFÉRENCES

Babau, M., Cescut, J., Allouche, Y., Lombaert-Valot, I., Fillaudeau, L., Uribelarrea, J.-L., Molina-Jouve, C., 2013. Towards a Microbial Production of Fatty Acids as Precursors of Biokerosene from Glucose and Xylose. Oil Gas Sci. Technol. – Rev. d'IFP Energies Nouv. 68, 899–911. doi:10.2516/ogst/2013148

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol. Rev. 19, 219–37.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2009. Yarrowia lipolytica as a model for bio-oil production. Prog. Lipid Res. 48, 375–387. doi:10.1016/j.plipres.2009.08.005

Cescut, J., 2009. Accumulation d'acylglycérols par des espèces levuriennes à usage carburant aéronautique: physiologie et performances de procédés. Institut National des Sciences Appliquées de Toulouse.

Droop, M.R., 1968. Vitamin B12 and marine ecology. The kinetics of uptake, growth and inhibition in monochrysis Lutheri. J. Mar. Assoc. 689–733.

Fernando, S., Adhikari, S., Chandrapal, C., Murali, N., 2006. Biorefineries: Current Status, Challenges, and Future Direction. Energy & Fuels 20, 1727–1737. doi:10.1021/ef060097w

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E.H., Consiglio, A.L., MacEwen, K., Crabtree, D. V., Afshar, J., Nugent, R.L., Hamilton, M.A., Joe Shaw, A., South, C.R., Stephanopoulos, G., Brevnova, E.E., 2016. Engineering of a high lipid producing

Yarrowia lipolytica strain. Biotechnol. Biofuels 9, 77. doi:10.1186/s13068-016-0492-3

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2008. A Practical Guide to Support Vector Classification. BJU Int. 101, 1396–400. doi:10.1177/02632760022050997

Ledesma-Amaro, R., Nicaud, J.-M., 2016. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. Prog. Lipid Res. 61, 40–50. doi:10.1016/j.plipres.2015.12.001

Llamas, A., García-Martínez, M., Al-Lal, A.-M., Canoira, L., Lapuerta, M., 2012. Biokerosene from coconut and palm kernel oils: Production and properties of their blends with fossil kerosene. Fuel 102, 483–490. doi:10.1016/j.fuel.2012.06.108

Meng, X., Xian, M., Xu, X., Zhang, L., Nie, Q., Yang, J., Xu, X., Zhang, L., Nie, Q., Xian, M., 2009. Biodiesel production from oleaginous microorganisms. Renew. Energy 34, 1–5. doi:10.1016/j.renene.2008.04.014

Naik, S.N., Goud, V. V., Rout, P.K., Dalai, A.K., 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. 14, 578–597. doi:10.1016/j.rser.2009.10.003

Nicaud, J.M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.

Ochoa-Estopier, A., Guillouet, S.E., 2014. D-stat culture for studying the metabolic shifts from oxidative metabolism to lipid accumulation and citric acid production in Yarrowia lipolytica. J. Biotechnol. 170, 35–41. doi:10.1016/j.jbiotec.2013.11.008

Pan, P., Hua, Q., 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS One 7, e51535. doi:10.1371/journal.pone.0051535

Papanikolaou, S., Aggelis, G., 2011. Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. Eur. J. Lipid Sci. Technol. 113, 1031–1051. doi:10.1002/ejlt.201100014

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142. doi:10.1007/s00284-005-0223-7

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., Nicaud, J.M., 2015. Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. Biotechnol. Biofuels 8, 104. doi:10.1186/s13068-015-0286-z

Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. Biochimie 86, 807–15. doi:10.1016/j.biochi.2004.09.017

Ratledge, C., Wynn, J.P., 2002. The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. Adv. Appl. Microbiol. 51, 1–51.

Sauer, U., 2006. Metabolic networks in motion: 13C-based flux analysis. Mol. Syst. Biol. 2, 62. doi:10.1038/msb4100109

Song, H.-S., Reifman, J., Wallqvist, A., 2014. Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. PLoS One 9, e112524. doi:10.1371/journal.pone.0112524

Vapnik, V., Golowich, S.E., Smola, A., 1996. Support vector method for function approximation, regression estimation, and signal processing. Annu. Conf. Neural Inf. Process. Syst. 281–287. doi:10.1007/978-3-642-33311-8_5

Vicente, G., Bautista, L.F., Rodríguez, R., Gutiérrez, F.J., Sádaba, I., Ruiz-Vázquez, R.M., Torres-Martínez, S., Garre, V., 2009. Biodiesel production from biomass of an oleaginous fungus. Biochem. Eng. J. 48, 22–27. doi:10.1016/j.bej.2009.07.014