**Readme**

This file presents brief descriptions of the scripts in the different folders of the model, so the user can set up its different functionalities quickly. The latest version of the model will be deposited in our github repository (<https://github.com/fjsaitua/RY-dFBA/tree/master/main%20P_pastoris%20dFBA>). Any questions of how to implement the model can be addressed to [fjsaitua@uc.cl](mailto:fjsaitua@uc.cl).

Programs that need to be installed to run the model:

1. Matlab (<http://www.mathworks.com/>)
2. Gurobi (<http://www.gurobi.com/>)
3. Cobra Toolbox for Matlab (<https://opencobra.github.io/>)
4. LibSBML (<http://sbml.org/Software/libSBML>)
5. SBLM Toolbox (<http://sbml.org/Software/SBMLToolbox>)
6. Enhanced scatter search algorithm (<http://pubs.acs.org/doi/abs/10.1021/ie801717t>)

Once these programs are correctly installed, the model may be used.

**List of scripts in alphabetical order**

***Batch Model***

* add\_it.m
  + Part of the automatic reparametrization routine HIPPO.
  + Adds a new iteration to the pending iteration group, only if it is not present already in the past group or the pending group. Returns the modified pending iterations group, and the modified tree and codes.
  + Try not to modify it
  + Last update: Benjamín Sanchez 2014 – 11 – 24.
* CrossCalibration.m
  + Recalibrates the available data with one of the candidate robust modeling structures.
  + Last Update: Francisco Saitua 2016 – 12 - 22
* decision.m
  + Part of the automatic reparametrization routine HIPPO
  + Constructs ktofix; a vector indicating which parameters should be fixed according to the sensitivity and identifiability analysis. Returns a vector with the parameters to be fixed, since they either presented a correlation bigger than 0.95 with another parameter of the model or the average sensitivity associated to those parameters was less than 0.01.
  + Try not to modify it
  + Last update: Benjamín Sánchez 2014 – 11 – 24
* find\_best.m
  + Part of the automatic reparametrization procedure HIPPO
  + Find all the model structures with no identification/sensitivity problems.
  + Try not to modify it
  + Last update: Benjamín Sánchez 2014 – 11 – 19
* fixedModifications
  + Part of the dynamic genome-scale model of *P. pastoris*.
  + Modifies the model according to parameters values. These modifications will be fixed during the whole integration. For example, here we set the maintenance ATP flux. Previous versions of the model used this script to include the stoichiometric coefficients of the biomass formation reaction as model parameters (See <https://www.ncbi.nlm.nih.gov/pubmed/25046158>)
  + Modifiable
  + Last update: Francisco Saitua 2016 – 10 – 27
* fluxMetCentral.m
  + Function for the analysis of flux distributions
  + Function that returns the central carbon metabolism fluxes of a flux vector (result of a cobra optimization) and the reactions associated to cytosolic and mitochondrial oxygen and cytosolic CO2.
  + Modifiable
  + Last update: Francisco Saitua 2016 – 12 – 19
* identificaBSB.p
  + Private script
  + Runs the identifiability analysis
  + Unmodifiable
  + See identificaHelp.m for help.
  + Claudio Gelmi PhD © All rights reserved
* intConfianza.p
  + Private script
  + Runs the determination of confidence intervals
  + It is a replace for the NLPARCI function of matlab, which is included in the Statistics toolbox.
  + Unmodifiable
  + See intconfianzaHelp for Help
  + Claudio Gelmi PhD © All rights reserved
* iteration\_complete.m
  + Key script, part of the dynamic genome-scale model of *P. pastoris*
  + Performs a complete iteration of the procedure, including parameter estimation and all the pre/post-regression analysis metrics.
  + In this script the model:
    - Initializes the Cobra Toolbox
    - Selects the solver for LP and QP problems
    - Loads the metabolic model
    - Defines the species involved in the analysis
    - Defines the lower and upper bounds of the parameters
    - Calls the function that defines the parameter estimation problem (minSquares.m)
    - Specifies the scatter search parameters (see the Egea et al. reference from 2009 in the manuscript)
    - Loads the experimental data
    - Performs the parameter estimation
    - Runs the script that performs the sensitivity, identifiability and significance tests.
  + Modifiable
  + Last update Francisco Saitua 2016 – 12 – 22
* iteration\_complete\_valid.m
  + It performs the same routines as iteration\_complete.m, but it also receives an input of the number of the model structure being tested in the cross – calibration stage.
  + Modifiable
  + Last update Francisco Saitua 2016 – 10 – 27.
* iteration\_noSS\_noCC.m
  + Part of the Reparametrization routine HIPPO
  + Does an iteration of the procedure, but skips the parameter estimation and pre/post regression diagnostics, only deciding which parameters to fix according to sensitivity and identifiability.
  + Last Update: Benjamín Sánchez 2014 – 11 – 24
* kineticConstraints.m
  + Important Script
  + Contains the kinetic expressions associated to the species included in the model and sets the lower and upper bounds of the corresponding exchange reactions.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 19
* ksensibilidadBSB.m
  + Private script
  + Runs the parameter sensitivity analysis
  + Unmodifiable
  + See ksensibilidadHelp.m for help.
  + Claudio Gelmi PhD © All rights reserved
* minSquares.m:
  + Important script, used by the scatter search algorithm to evaluate the objective function used in parameter estimation (minimization of normalized squared residuals).
  + It also keeps track of the iterations performed by the scatter search algorithm during parameter estimation and save the partial results in the matrix checkpoint.mat
  + Modifiable (specially if you want to try a different objective function for model calibration)
  + Last Update: Francisco Saitua 2016 – 12 – 22
* minSquares2.m:
  + Integrates the dynamic model and returns the simulation output, weighted appropriately. To be used in the lsqcurvefit function to determine model residuals, which are used in the determination of confidence intervals in the scripts reg\_analysis\_complete.m and reg\_analysis\_onlyCC.m.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* permited\_it.m
  + Returns a boolean if the proposed iteration is allowed or not. Customizable for each dynamic model.
  + It could be used to reduce the number of modeling structures explored by HIPPO. For instance, one can set that the algorithm should not explore structures that fix all the parameters associated to one of the state variables.
  + Last Update: Francisco Saitua 2016 – 12 – 22.
* printResults.m
  + Displays the model simulation & experimental results in a 2x3 graph.
  + Last Update: Francisco Saitua 2016 – 12 – 22.
* pseudoSteadyState.m
  + Important script, corresponding to the dynamic block of the model.
  + Used in the parameter estimation and reparametrization stages.
  + Contains the system of ordinary differential equations that represent the mass balances of the species considered in the model.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 22.
* pseudoSteadyState\_simulationMOMA.m
  + Similar script to pseudoSteadyStatem, but this version was designed to run dynamic single gene deletions employing the minimization of metabolic adjustment as objective function.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 22
* recalculate\_CC.m
  + Part of the reparametrization algorithm HIPPO
  + Calculates for all solutions of the reparametrization the CCs, and discards the ones that have any CC > 2.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 22
* reg\_analysis\_complete.m
  + Performs al the regression analysis for a parameter set: significance, sensitivity and identifiability analysis, and AICc calculation. Returns all the different regression analysis outputs.
  + Modifiable, the first entry of the identificaBSB and ksensibilidadBSB functions correspond to the state variables to be considered in the analysis.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* reg\_analysis\_noSS\_noCC.m
  + Part of the reparametrization algorithm HIPPO
  + Updates the sensitivity (Ms) and Correlation (Mc) matrices, removing columns and rows of the original one that are fixed in this iteration.
  + Try not to modify it.
  + Last Update: Benjamín Sánchez 2014 – 11 – 24
* reg\_analysis\_onlyCC.m
  + Part of the reparametrization algorithm HIPPO
  + Calculates the Coefficients of Confidence (CC) values for a given estimation.
  + Last Update: Benjamín Sánchez 2014 – 11 – 29
* reparam\_dFBA.m
  + Main function for reparametrization of the dFBA model with the HIPPO algorithm
  + Try not to modify it.
  + Last Update: Benjamín Sánchez 2014 – 11 – 29
* RPP\_dFBA.m
  + Main function to use, to understand the logic of model calibration and reparametrization, please read this script first.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 29
* run\_PPdFBA.m
  + Main script to run simulations with the batch model.
  + It can be set up to run normal simulations or dynamic single gene deletions.
  + To be used with the main script RunMOMABatch.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 29
* Run\_RPP\_dFBA.m
  + Main script for model calibration and reparametrization. This function calls the RPP\_dFBA function
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 29
* solveFBA.m
  + Important Script
  + Solves a QP problem to find metabolic fluxes via FBA. It considers a trade-off between two metabolic objectives: Maximize biomass formation and minimize the absolute flux sum.
  + Uses the solveCobraQP command to specify the optimization problem.
  + If desired, the last section the script could become available to save the flux distribution throughout a cultivation
  + Last Update: Francisco Saitua 2016 – 12 – 22
* solveFBA\_simulationMOMA.m
  + Important Script
  + Solves two sequential QP problems to find the fluxes of a single gene deletion mutant using the minimization of metabolic adjustment as objective function.
  + If desired, the last section the script could become available to save the flux distribution throughout a cultivation
  + Last Update: Francisco Saitua 2016 – 12 – 22

***Fed-batch model***

* add\_it.m: Same function as in the batch model
* CrossCalibration: Same function as in the batch model
* decision.m: Same function as in the batch model
* feed\_function.m
  + Part of the dynamic genome-scale model of *P. pastoris*.
  + Returns the flux in [L/h] associated to the feed function of a fed-batch culture, which is specified in this script.
  + Different feeding policies could be eincluded
  + Modifiable
  + Last update: Francisco Saitua 2016 – 10 – 27
* find\_best.m: Same function as in the batch model
* fixedModifications.m: Same function as in the batch model
* identificaBSB.p: Same function as in the batch model
* intconfianza.m: Same function as in the batch model
* iteration\_complete.m
  + Same function as in the batch model.
  + More information is included to the model:
    - Feed concentration
    - Feed start time
    - Sampling Rate
    - Parameters associated to the fedbatch version of the model
  + Last Update: Francisco Saitua 2016 – 12 – 22
* iteration\_complete\_valid: Same function as in the batch model
* iteration\_noSS\_noCC: Same function as in the batch model
* kineticConstranits: Same function as in the batch model
* ksensibilidadBSB.p: Same function as in the batch model
* minSquares.m: Same function as in the batch model
* minSquares2.m: Same function as in the batch model
* permited\_it.m Same function as in the batch model
* printResults.m: Same function as in the batch model
* pseudoSteadyState.m
  + Same function as in the batch model but now this includes a variable volume and the sampling rate of the experiment in the mass balances.
  + Modifiable
  + Last Update: Francisco Saitua 2016 – 12 – 22
* recalculate\_CC: Same function as in the batch model
* reg\_analysis\_complete: Same function as in the batch model
* reg\_analysis\_noSS\_noCC.m: Same function as in the batch model
* reg\_analysis\_only\_CC.m: Same function as in the batch model
* reparam\_dFBA: Same function as in the batch model
* RPP\_dFBA.m: Same function as in the batch model
* run\_PPdFBA: Same function as in the batch model
* Run\_RPP\_dFBA: Same function as in the fed-batch model
* solveFBA: Same function as in the fed-batch model
* Try\_Validation: Same function as in the fed-batch model

***Simulations***

Dynamic Flux Balance Analysis

* NodeAnalysis:
  + Main function to analyze flux distributions
  + Determines and presents as graph bars the fluxes producing and consuming a certain metabolite, in order to build flux distributions
  + It uses a “Metabolic movie” derived from the simulation of a dynamic culture (by activating the lines of code from the solveFBA scripts).
  + Last Update: Francisco Saitua 2016 – 12 – 22
* printMajorFluxes\_Consumption.m
  + Finds and prints the fluxes that consume the metabolite (node) being analyzed.
  + If the model is assessed using the biomass maximization as objective function (LP problem), extremely high fluxes may appear (because the participate in type III pathways). This is avoided by determining flux distributions considering the minimization of total fluxes.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* printMajorFluxes\_Production.m
  + Finds and prints the fluxes that produce the metabolite (node) being analyzed.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* StackedBarPlot.m
  + Constructs the graphs that organize the flux information in different stages of the cultivation.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* writePieChartData.m
  + Writes the information of the Node Analysis to an excel spreadsheet.
  + Last Update: Francisco Saitua 2016 – 12 – 22

Dynamic MOMA

* Analyze\_dMOMA\_output.m
  + Ranks and plots the performance of single gene deletion mutants according to how much they increase the final concentration of the protein of interest.
  + It uses as input matrices containing the value of the state variables of the model at the end of a batch cultivation.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* NodeAnalysis\_MOMA.m
  + Analog to NodeAnalysis.m of the dynamic flux balance analysis section, but instead of comparing different fermentation times, it analyzes the flux producing or consuming metabolites of interest among single gene deletion mutants.
  + It is the main script for the analysis of the metabolic impact of single gene deletions.
  + Last Update: Francisco Saitua 2016 – 12 – 22
* printMajorFluxes\_Consumption:
  + Analog to the function of the Dynamic Flux Balance Analysis
* printMajorFluxes\_Production
  + Analog to the function of the Dynamic Flux Balance Analysis
* RunMOMABatch.m
  + Runs the dynamic Single Gene Deletions using the Minimization of the Metabolic Adjustment as objective function to simulate the mutant’s metabolic behavior.
  + It uses scripts located in the folder containing the batch model, but you can work from the Dynamic MOMA directory.
  + Important Script, it saves the profiles of the mutants along with the parental strain.
  + Last Update: Francisco Saitua 2016 – 12 – 2013
* StackedBarPlot:
  + Analog to the function of the Dynamic Flux Balance Analysis
* writePieChartData
  + Analog to the function of the Dynamic Flux Balance Analysis
* Xticklabel\_rotate.m
  + Function to rotate the tick on the X axis in plots

Evaluation of Feeding Policies

* feedFunction.m
  + Analog to the function of the fed-batch model, with the exception that it considers the μSET profile being simulated.
  + Simulates the addition of feeding solution to the culture (in L/h).
  + Last Update: Francisco Saitua 23 – 12 – 23
* fixedModifications:
  + Script already described
* get\_O2\_for\_input\_times:
  + Using the flux distribution during the culture and the sampling time, this script determines the instantaneous oxygen uptake rate for the different experimental times
  + Last Updated: Francisco Saitua 2016 – 12 – 23
* kineticConstraints:
  + Script already described, but sets the specific production of the recombinant protein being analyzed according to a function that depends on the growth rate.
  + Last Updated: Francisco Saitua 2016 – 12 – 22
* pseudoSteadyState:
  + Script already described.
* qp\_seurm
  + Function that determines the specific volumetric productivity of HAS depending on the growth rate of the culture.
  + Last Updated: Francisco Saitua 16 – 12 – 23
* run\_PPdFBA
  + Script already explained
* solveFBA
  + Script already described
* stopEvents
  + Important script
  + Sets the stop conditions of the simulations that evaluate different feeding policies
  + In our case, the stop conditions were 2:
    - The increase of the culture volume beyond 1L (maximum capacity of our setup)
    - The oxygen demand from P. pastoris was higher than what our setup could offer
  + This script was included in the “odeset” command of the ode15s integrator
  + Last update: Francisco Saitua 23 – 12 – 22
* Test\_muset\_fedbatch
  + Key script of the simulations
  + Evaluates different feeding strategies and calculates the volumetric productivity and the final concentration of a product of interest given a feeding profile.
  + The evaluation of several feeding strategies helps to determine which one would maximize the production of a compound of interest.
  + Last Updated: Francisco Saitua 2016 – 12 – 22