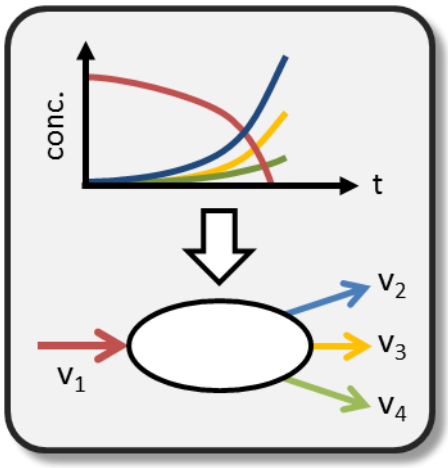

PhysioFit

Software Version 1.0

http://github.com/MetaSys-LISBP/PhysioFit/

User Manual

Version 1.0

UMR5504, UMR792 Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés Group MetaSys: Metabolic Systems
Author: Pierre Millard <pierre.millard@insa-toulouse.fr>

LISBP, INSA
CNRS, INRA, INSA - 135 Avenue de Rangueil - 31077 TOULOUSE Cedex 4 - FRANCE
Tel/Fax : +33 (0) 561 559 689 - http://metasys.insa-toulouse.fr

CONTENTS

INTRODUCTION

PhysioFit is a scientific tool designed to *i*) quantify exchange (production and consumption) fluxes and *ii*) cell growth rate during (batch) cultivations of microorganisms. Fluxes are estimated from time-course measurements of extracellular metabolites and biomass concentrations. An important assumption is that cells are in metabolic (pseudo) steady-state.

PhysioFit includes the following features:

- Calculation of growth rate and extracellular (uptake and production) fluxes.
- Lag before growth (e.g. due to adaptation to a novel environment) can be taken into account and estimated.
- Non-enzymatic degradation of some carbon sources (e.g. DHA or glutamine) can be estimated and taken into account when calculating exchange fluxes.
- Sensitivity analyses are performed to estimate the precision of the calculated fluxes.
- Evaluation of the goodness of fit and visual inspection of the fitted curves.

LICENSING

The original version of PhysioFit software was developed in the MetaSys team (http://metasys.insa-toulouse.fr) in the LISBP, Toulouse, FRANCE.

If you publish results obtained with PhysioFit, please provide the link to the PhysioFit repository.

Software and documentation author: Pierre Millard, INRA, France <pierre.millard@insa-toulouse.fr>

METHODS

Initial concentrations of species (i.e. concentrations of biomass - X - and metabolites - M_i - at $t = 0$) and fluxes (exchange fluxes - q_{M_i} - and growth rate - μ -) are estimated by fitting to time-course measurements of experimental data. Simulations, i.e. calculation of $X(t)$ and $M_i(t)$, are performed using analytical functions, as detailed below.

The general model, taking into accounts both degradation and lag phase, can be described using the following system of ordinary differential equations:

$$\frac{dX}{dt} = \begin{cases} 0 \; if \; t < t_{lag} \\ \mu \cdot X \; else \end{cases} \qquad (eq. 1)$$

$$\frac{dM_i}{dt} = \begin{cases} -k \cdot M_i \; if \; t < t_{lag} \\ -k \cdot M_i + X \cdot q_{M_i} \; else \end{cases} \qquad (eq. 2)$$

with q_{M_i} being positive (negative) when M_i is produced (consumed). The sign can thus be used to automatically identify products and substrates in high throughput workflows for automated functional analysis of metabolic systems.

Integrating equations 1-2 provides the following analytical functions:

$$X(t) = \begin{cases} X_0 \; if \; t < t_{lag} \\ X_0 \cdot e^{\mu \cdot (t - t_{lag})} \; else \end{cases} \qquad (eq. 3)$$

$$M_i(t) = \begin{cases} M_i^0 \cdot e^{-k \cdot t} \; if \; t < t_{lag} \\ q_{M_i} \cdot \frac{X_0}{\mu + k} \cdot \left(e^{\mu \cdot (t - t_{lag})} - e^{-k \cdot (t - t_{lag})} \right) + M_i^0 \cdot e^{-k \cdot t} \; else \end{cases} \qquad (eq. 4)$$

In the absence of a lag phase (i.e. $t_{lag} = 0$), equations 3-4 simplifies to:

$$X(t) = X_0 \cdot e^{\mu \cdot t} \qquad (eq. 5)$$

$$M_i(t) = q_{M_i} \cdot \frac{X_0}{\mu + k} \cdot (e^{\mu \cdot t} - e^{-k \cdot t}) + M_i^0 \cdot e^{-k \cdot t} \qquad (eq. 6)$$

In the absence of degradation (i.e. $k = 0$), equation 4 simplifies to:

$$M_i(t) = \begin{cases} M_i^0 \; if \; t < t_{lag} \\ q_{M_i} \cdot \frac{X_0}{\mu} \cdot \left(e^{\mu \cdot (t - t_{lag})} - 1 \right) + M_i^0 \; else \end{cases} \qquad (eq. 7)$$

In the absence of both degradation and lag (i.e. $t_{lag} = 0$ and $k = 0$), equations 3-4 simplifies to:

$$X(t) = X_0 \cdot e^{\mu \cdot t} \qquad (eq. 8)$$

$$M_i(t) = q_{M_i} \cdot \frac{X_0}{\mu} \cdot (e^{\mu \cdot t} - 1) + M_i^0 \qquad (eq. 9)$$

Parameter estimation is performed using the *nlsic* optimization algorithm (see Sokol *et al.*, Bioinformatics, 2012 for details) by minimizing the following cost function:

$$residuum = \sum_i \left(\frac{sim_i - meas_i}{weight_i}\right)^2 \qquad (eq. 10)$$

where sim is the simulated data, $meas$ denotes measurements, and $weight$ is a weighting factor (e.g., standard deviation on measurements).

A global sensitivity analysis (Monte-Carlo approach) is implemented to evaluate the precision of the estimated parameters, plots are generated for visual inspection of the fitting quality, and a khi2 test is performed to assess the statistical goodness of fit.

Finally, PhysioFit includes routines to estimate the first-order degradation constants from time-course metabolite concentrations measured in the medium without cells. In this situation, simulations are performed using the following equation:

$$M_i(t) = M_i^0 \cdot e^{-k \cdot t} \qquad (eq. 11)$$

REQUIREMENTS AND DEPENDENCIES

PhysioFit was developed on Windows but can be used on Linux (tested on Ubuntu 12 and Mandriva 2011), Windows (tested on XP and 7) and MacOS (not tested yet) platforms.

This software requires installing:

- `R` 3.0+
- The following R packages: `nnls`, `numDeriv`, `RColorBrewer`. These packages can be installed by running the following command in an R console:

```
install.packages(c("nnls", "numDeriv", "RColorBrewer"))
```

The `nlsic.R` script is required for optimization and must be included in the `physio_fit.R` directory. PhysioFit is freely available and is distributed under open-source license at:

http://github.com/MetaSys-LISBP/

If you are not used to install system wide environments like R, ask some help from your local computer service. We don't provide support for installation.

INSTALLATION

Unpack the content of `PhysioFit-vX.Y.zip` (where X.Y is the version number) somewhere on your disk. If you want to make PhysioFit available system wide and install it in a protected directory, you need administrative privileges. Otherwise, PhysioFit will be available only in your personal session.

USER'S MANUAL

I. Required information and data files

PhysioFit requires two inputs:

- Time-course measurements of extracellular metabolites and biomass concentration.
- Precision on measurements.

Time-course measurements of extracellular metabolites and biomass concentration must be provided in a tabulated text file ('`.txt`' extension), with the following columns:

`time`	experimental times
`X`	biomass concentration
`meta_y`	metabolite y concentration (1 column per metabolite)

An example of the input file structure is shown in Figure 1.

```
time	X	Glc	Ace	Lac	Man
1.5	0.056	12.4	1	0.001	15
2	0.059	10.9	NA	0.07	14
3	0.078	9.60	NA	0.39	12
4	0.135	7.80	NA	1.75	9
5	0.233	4.10	5.8	3.23	5
6	0.451	0.10	8.3	4.92	0.1
```

Figure 1. Structure of PhysioFit input file. The first column (must be named `time`) contains sampling times, the second column (must be named `X`) contains biomass concentration, and other column contains metabolites (substrates or products).

Units of estimated parameters depend on the units of experimental data, e.g. fluxes will be given in mmol/g_{DW}/h if metabolite concentrations are in mM, times in h, and biomass concentration in g_{DW}/L.

PhysioFit can handle missing data which must be given as `NA`s in the input file.

Weighting factors (e.g. experimental standard deviation on measurements) used to calculate cost during optimization can be provided in a file with the same format (and named `xxx_sd.txt`, where `xxx` is the name of data file), except for the `time` column which must be removed. Weighting factors may also be provided as an argument (see the Usage section).

An example of input files (`example.txt` and `example_sd.txt`) ready to run is provided with the software in the `Example` folder.

II. PhysioFit usage

To run the calculation, the user must:

- Go to PhysioFit directory, and load PhysioFit:

```
setwd("D:/PhysioFit_directory")
source("physio_fit.R")
```

- Go to the working directory containing input file(s)

```
setwd("D:/Data_directory/Experiment_1")
```

- Run the calculation

```
res <- physio_fit("example")
```

The function `physio_fit(datfile, ...)` takes the following arguments:

`datfile`	name of input file (with or without `'.txt'` extension)
`mc`	boolean, Monte-Carlo sensitivity analysis is performed if True (default = True)
`it`	number of Monte-Carlo iterations (default = 50)
`pos`	negative concentrations of noisy datasets generated during Monte-Carlo iterations are set to 0 if True (default = True)
`vini`	initial value for fluxes and concentrations (default = 1)
`upcf`	upper constraints on fluxes (default = 50)
`locf`	lower constraints on fluxes (default = -50)
`upcc`	upper constraints on initial concentrations (default = 50)
`locc`	lower constraints on initial concentrations (default = 1e-6)
`weight`	weight matrix used for residuum calculation, can be: • a matrix with of same dimensions as the measurements matrix (but without the `time` column) and containing weights (e.g. SDs on measurements) • a named vector containing weights for <u>all</u> the metabolites provided in the input file • NULL (by default), in this case the matrix is automatically loaded from the file `xxx_sd.txt` - where `xxx` is the data file name - if this file exists, otherwise `weight` is constructed from `sd_X` and `sd_M` arguments
`sd_X`	standard deviation on biomass concentration (default = 0.002), used only if `weight=NULL`
`sd_M`	standard deviation on metabolite concentrations (default = 0.5), used only if `weight=NULL`
`save`	boolean, save results (in txt & pdf files) if True (default = True)
`col`	vector of colors (hexadecimal codes) for the plot (default = NULL, in this case colors are automatically generated using from the `Set1` colors of the `RColorBrewer` package)
`alpha`	alpha channel (degree of transparency) used to plot the confidence intervals on the fit, must be provided as a 2-character hexadecimal code (default = 33)
`deg`	a named vector containing first-order degradation constants for metabolites which are degraded non-enzymatically in the medium

`lag` boolean, a lag phase is assumed (and the lag time estimated) if True (default = False)
`summary` boolean, results of the khi-2 test are displayed if True (default = False)

PhysioFit returns a list with the following items (type):

`$sys` information on the system (list)
- `$to_est` names of estimated parameters (vector)
- `$params` parameters (vector)
- `$nconc` names of all concentrations (vector)
- `$nflux` names of fluxes (vector)
- `$metab` names of metabolites (vector)
- `$weight` standard deviation on measurements (matrix)
- `$te_upc` upper bounds of free parameters (vector)
- `$te_loc` lower bounds of free parameters (vector)
- `$u` linear constraint matrix, u*par>=co (matrix)
- `$co` constraints vector (vector)
- `$times` experimental times (vector)
- `$data_meas` experimental data (matrix)
- `$nb_par` number of free parameters (integer)
- `$nb_conc` number of extracellular metabolites (integer)
- `$deg` degradation constants (vector)
- `$lag` estimation of lag phase (boolean)

`$opt` optimization results returned by nlsic (list)

`$khi2test` results of the khi2 test for goodness of fit evaluation (list)
- `$'khi2 value'` names of parameters to estimate (float)
- `$'data points'` number of measurements (integer)
- `$'fitted parameters'` number of free parameters (integer)
- `$'degrees of freedom'` degrees of freedom of the system (integer)
- `$'khi2 reduced value'` khi2 reduced value, i.e. `'khi2 value'/'data points'` (float)
- `$'p-value'` p-value of the test, i.e. $P(\chi^2 \leq 0.95)$ (float)
- `$'conclusion'` conclusion of the test (text), i.e. "*At level of 95% confidence, the model [fits/does not fit] the data good enough with respect to the provided measurement SD*."

`$sens` mean, median, standard deviation and 95% confidence intervals on the free parameters, estimated using a Monte-Carlo analysis (matrix)

PhysioFit also includes routines to estimate the first-order degradation constants (*k*) of metabolites which are degraded non-enzymatically. In this case, the input file must contain time course concentrations of each metabolite measured in the medium without cells, and the *X* column must not be included. The input file structure is the same as the input file of the function `physio_fit`. To estimate *k*, run the following command:

```
res_deg <- estimate_k("example_blk")
```

This function returns a list containing the same elements as the output of `physio_fit`, with an additional (named) vector `$deg_cst`, which contains the degradation constants for each metabolite and can be passed directly to `physio_fit`:

```
res <- physio_fit(..., deg=res_deg$deg_cst)
```

III. Results files

Results are stored in the following files (where `xxx` is the name of the input file) created in the `xxx_res` folder of the working directory:

`xxx.pdf`	plot of simulated and experimental data (with the 95% confidence intervals on the fit if `mc=TRUE`)
`xxx_res.txt`	optimal solution and goodness of fit (khi2 test), results of the sensitivity analysis (mean, median, SD, 95% confidence intervals) are provided if `mc=TRUE`
`xxx_log.txt`	detailed information on the system (list of metabolites, initial parameters, constraints, etc) and nlsic results (parameters, jacobian, residuals, etc)

LICENSE FOR PHYSIOFIT SOFTWARE

Educational Community License, Version 2.0 [OSI Approved License]

Educational Community License Version 2.0, April 2007

The Educational Community License version 2.0 ("ECL") consists of the Apache 2.0 license, modified to change the scope of the patent grant in section 3 to be specific to the needs of the education communities using this license. The original Apache 2.0 license can be found at: http://www.apache.org/licenses/LICENSE-2.0

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed. Any patent license granted hereby with respect to contributions by an individual employed by an institution or organization is limited to patent claims where the individual that is the author of the Work is also the inventor of the patent claims licensed, and where the organization or institution has the right to grant such license under applicable grant and research funding agreements. No other express or implied licenses are granted.

4. Redistribution.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

You must give any other recipients of the Work or Derivative Works a copy of this License; and

You must cause any modified files to carry prominent notices stating that You changed the files; and

You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and

If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions. Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to

damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing the Work of Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

APPENDIX: How to apply the Educational Community License to your work

To apply the Educational Community License to your work, attach the following boilerplate notice, with the fields enclosed by brackets "[]" replaced with your own identifying information. (Don’t include the brackets!) The text should be enclosed in the appropriate comment syntax for the file format. We also recommend that a file or class name and description of purpose be included on the same "printed page" as the copyright notice for easier identification within third-party archives.

Copyright [yyyy] [name of copyright owner] Licensed under the Educational Community License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

http://www.osedu.org/licenses/ECL-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

PhysioFit (hereafter "the program") is a scientific software for experimental design of ^{13}C-MFA experiments.

All its parts are freely distributed according to GPL2 terms http://www.gnu.org/licenses/gpl-2.0.txt (permission to execute, to copy, to modify and to redistribute; obligation to preserve the original copyright notice and to keep your modifications OpenSource only if you redistribute them alone or as part of PhysioFit)

WARRANTY

BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

COPYRIGHT 2018, INRA