GS-MFA: Genome-scale metabolic flux analysis

User Manual

Software and first draft written by Saratram Gopalakrishnan

Past contributors: Hoang Dinh, Charles Foster, John Hendry

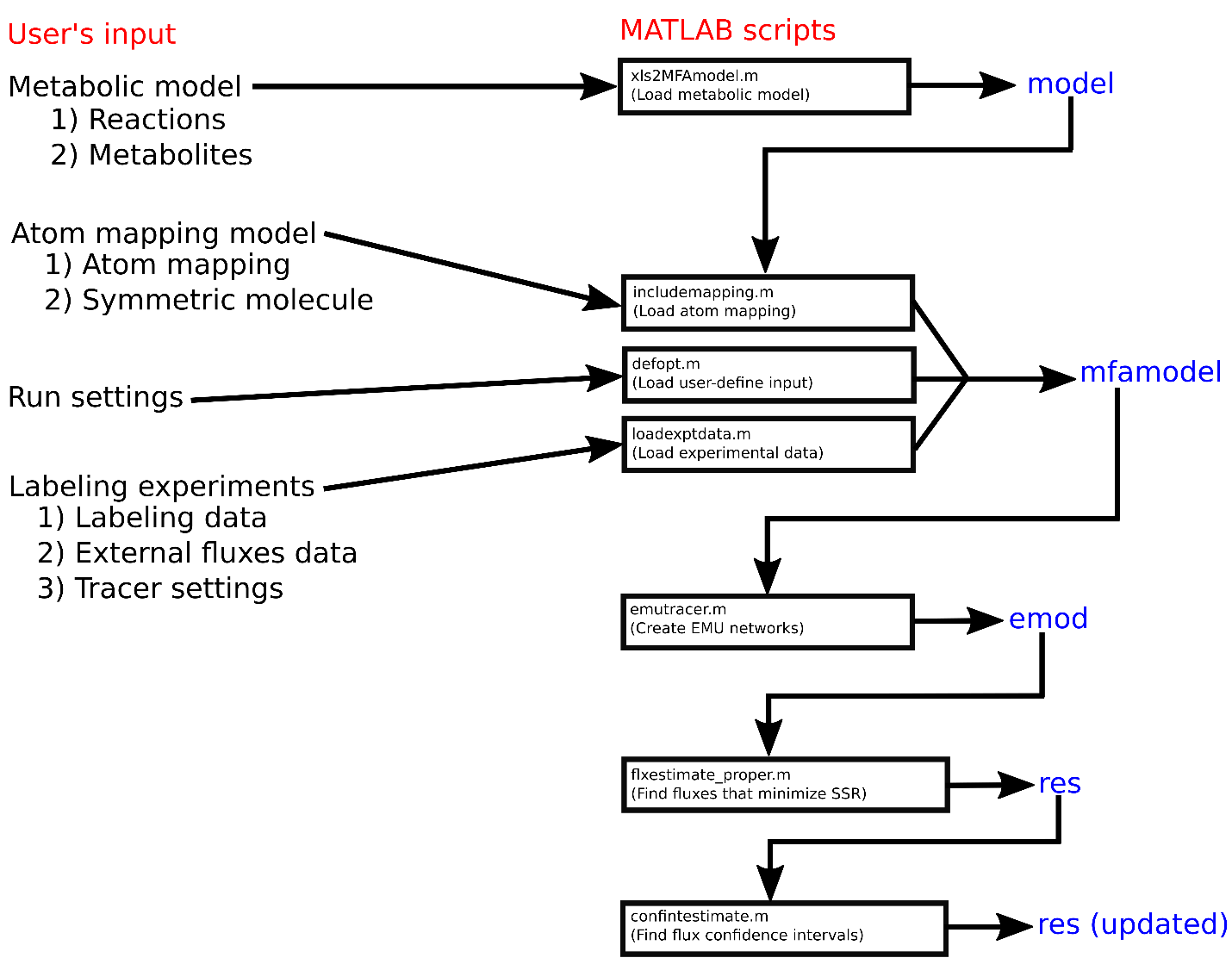
Current revision (11/30/2021) by Hoang Dinh

Link to program: <https://github.com/maranasgroup/SteadyState-MFA>

Citation: Gopalakrishnan and Maranas. 2015. Met. Eng. 32:12-22.

# 1. Introduction

GS-MFA is a software program that is built to perform 13C-metabolic flux analysis for large to genome-scale metabolic network. The scripts used the established methods described in MFA literature [1–4], summarized in Gopalakrishnan and Maranas, 2015 [5].



# 2. Program structure

The program uses MATLAB as the platform and requires the following user-defined inputs: (i) metabolic model, (ii) atom mapping model, and (iii) labeling experiments. The program identifies the flux distribution that best recapitulates experimental labeling and extracellular flux input data and estimates flux 95% confidence interval.

# 3. User manual

## 3.1. xls2MFAmodel.m

*model = xls2MFAmodel(“your\_model.xlsx”)*

Description: load metabolic model in excel format to MATLAB-specific format

Output: model[1, 2]

Excel file specifications

**“Reactions” tab**

|  |  |
| --- | --- |
| RxnID | Abbreviated name of the reaction. E.g. PGI  Use the following prefix for reactions:   * “EX\_”: exchange reactions * “DIL\_”: pool dilution reactions |
| Rxn name | Full name of the reaction. E.g. Phosphoglucose isomerase for PGI |
| Rxn Formula | Contains the stoichiometric reaction formula for the overall reaction expressed using abbreviated metabolite names. Stoichiometric coefficient is specified in parenthesis before the metabolite. Reversible reactions are indicated using a double-headed arrow (<=>) and irreversible reactions are indicated using a single headed arrow (-->).  For irreversible reactions, only in forward direction. Format all irreversible reactions into forward arrow (“-->”) and bounds = (0,1000). For example, irreversible reaction “A <-- B” should be formatted into “B -->A”.  For exchange reactions:  Unlike COBRA models, a distinguishing feature of MFA models is that reactions are not allowed to have an empty reactants or products side.   * Uptake: format reaction to “Substrate\_out --> Substrate\_e”   E.g. EX\_glc\_\_D\_e: “glc\_\_D\_out --> glc\_\_D\_e”   * Secretion: format reaction to “Byproduct\_e --> Byproduct\_out”   E.g. EX\_etoh\_e: “etoh\_e --> etoh\_out”   * Reversible: no restriction   E.g. EX\_co2\_e: “co2\_e <=> co2\_out” |
| Lower bound / Upper bound | A default lower bound of -10-8 is set for reversible reactions. Irreversible reactions have a lower bound of 0. The default upper bound for all reactions is set to 108. Alternatively, if the direction of flux through a reversible reaction is known to be in the reverse direction only, the upper bound for such reactions can be set to 0. |
| Reversibility | Indicate reaction reversibility (0 = irreversible, 1 = reversible) |
| macro/bm | Macro/bm is not used for FBA. It is used to exclude reactions with non-integer stoichiometric coefficients (such as biomass reaction, protein synthesis) when checking for correctness of the atom mapping model. (0 = metabolic reaction, 1 = macromolecular reaction) |
| Subsystem | (optional) Subsystem describes the pathway(s) to which the reaction belongs. It is included for convenience of data analysis once fluxes have been elucidated. |

**“Metabolites” tab**

|  |  |
| --- | --- |
| metid | Abbreviated name of the metabolite  Use the following suffix for metabolites:   * “\_i”: metabolically inactive pool * “\_d”: (“diluted”) metabolite at which the scripts extract model predicted MDV |
| metname | (optional) Full name of the metabolite |
| metformula | (optional) Formula of the metabolite |

## 3.2. includemapping.m

*[model, mfamodel] = includemapping(model, ‘your\_atom\_mapping.xlsx’)*

Description: Load atom mapping matrix into MATLAB model

Output: *mfamodel*

Excel file specifications

**“Atom Transition” sheet**

|  |  |
| --- | --- |
| rxnAbbreviation | Mapped to RxnID from the “Reactions” tab in the metabolic model input file |
| metabAbbreviation | Indicate metabolite selected within the reactions |
| ReactantProductFlag | Indicate whether the selected metabolite is “reactant” or “product” |
| startNodeSymbol | Represents the symbol of the element being mapped. E.g. “C” for carbon, “H” for hydrogen, “O” for oxygen, and N for nitrogen. (Currently: Only support “C”) |
| maps | Every entry in “maps” contains three pieces of information:   1. Number of carbons are contained within the metabolite being mapped. 2. The identifier of the carbons. 3. The position of the identifier.   For example, g6p is represented in a reaction:  “1,2,3,4,5,6”   1. A slot, separated by comma, is in the order of carbon. This order is consistent throughout the mapping model. For example, first slot of g6p is always that same first carbon of g6p, regardless of which reaction is being examined. 2. Number: reaction-local identity of carbon. This is local to a (specific) reaction mapping. For example, in a (specific) reaction mapping, “1” will map a carbon from another metabolite to “slot 1” of g6p. If g6p is “6,7,8,9,10,11”, this means “slot 1” of g6p is transition from a carbon from another metabolite, mapped by “6”. |

**“Symmetry” sheet**

The “Symmetry” tab contains information on the existence of a plane or point of symmetry within metabolites. This is an alternate permutation of atoms. For example, the symmetry of succinate is captured by the fact that the atom configurations C1-C2-C3-C4 and C4-C3-C2-C1 are equivalent.

|  |  |
| --- | --- |
| Metab\_name | ID of metabolite with symmetry configuration. Notes: make sure metabolites in all compartment and inactive-pool metabolite (i.e., suffix “\_i”) are added to the list. |
| default mapping | Default mapping (e.g., 1, 2, …, n where n is the number of carbons in metabolite) |
| symmetry mapping | Mapping of carbon from symmetric configuration to numbers in “default mapping” |

## 3.3. defopt.m

*[mfamodel] = defopt(mfamodel)*

Description: Load user defined run settings for subsequent program execution. Please copy the file “defopt\_source.m” to the folder of your run instance (or technically, wherever recorded in the MATLAB path), modify it to your settings, and rename the file to “defopt.m”. This is done so that the provided settings in “defopt\_source.m” does not overwrite settings in your customized “defopt.m”

Output: Settings loaded to *mfamodel*

Quick settings:

* Raw uncorrected labeling data with tracer purity (<100%) input
  + *sim\_na = true, fcor = true*
* Fully corrected labeling data
  + *sim\_na = false, fcor = false*
* See section 3.4 for experiment data file configurations[3, 4]

List of settings

|  |  |
| --- | --- |
| ss | Steady-state. Value: *true*, *false*  *true*:run stationary  *false*: instationary (time-course) MFA (please refer to another user manual for INST-MFA) |
| sim\_na | Simulate natural isotope. Value: *true*, *false*  *true*: simulate naturally occurred isotope, use for uncorrected data input  *false*: do not simulate naturally occurred labeled isotope, use |
| fcor | Fragment correction. Value: *true*, *false*  *true*: add fragment correction to predicted MDV  *false*: add fragment correction to experimental MDV (i.e., raw input data) |
| default\_sd | Default standard deviation. Value: *true*, *false*  *true*: use default of 0.05  *false*: use user input standard deviation in experiment data file (see section 3.4) |
| output\_display | Output display. Value: *true*, *false*  *true*: output details of optimization objective and variable (in MATLAB command window)  *false*: no output details |
| dfbase | Damping factor for step-length for non-linear optimization. Value: float. Default: 1e-6 |
| conf\_lvl | Confidence level for goodness-of-fit test and flux confidence interval estimation. Value: float. Default: 0.95.  Do not affect outcome of *flxestimate\_proper.m* or *flxestimate\_fast.m* step but will affect outcome of *confintestimate.m*. |
| multistart | <OBSOLETE>. Number of re-run starting at randomized starting points. Default: 10.  Do not affect outcome of *flxestimate\_proper.m*. Affect outcome of *flxestimate\_fast.m*. |
| conf\_set | Set of fluxes to estimate the confidence interval. Value: string (see below for the options). Default: ‘*minset\_main*’. For ‘*minset\_main*’ and *‘minset\_all’* options, flux coupling analysis (Burgard et al. 2004) [6] is performed to find list of fluxes coupled to external flux measurements.  List of options:   * ‘*all’*: All fluxes * ‘*main’*: All fluxes except dilution fluxes * ‘*dilution’*: All dilution fluxes (start with ‘DIL\_’) * ‘*all\_net*’: All reactions, without exchange flux estimation. For reversible reaction, only estimate interval for net flux. * ‘*all\_exch*’: All reactions, without net flux estimation. For reversible reaction, only estimate interval for exchange flux. * ‘*minset\_main*’: All fluxes in minimal set which exclude fluxes coupling to external flux measurements and dilution fluxes (see section 3.4) * ‘*minset\_all*’: All fluxes in minimal set which exclude fluxes coupling to external flux measurements (see section 3.4). This option will include non-coupled dilution fluxes. * ‘*custom*’: User defined list of fluxes |
| conf\_step | Number of steps when performing flux confidence interval estimation. Value: int. Default = 10. |

## 3.4. loadexptdata.m

*[mfamodel] = defopt(mfamodel, ‘your\_experiment.xlsx’, ‘experiment\_id’, append\_flag)*

The third and final input file contains the experimental data to be used for flux elucidation. Three pieces of information must be provided here: (a) list of MS measurements, (b) list of extracellular flux measurements, and (c) the tracer scheme used in the isotope labeling experiment.

Load multiple data examples:

*[mfamodel] = defopt(mfamodel, ‘expmt1.xlsx’, ‘expmt1\_id’, true)*

*[mfamodel] = defopt(mfamodel, ‘expmt2.xlsx’, ‘expmt2’, true)*

Functions input

|  |  |
| --- | --- |
| mfamodel | Output from previous execution |
| ‘your\_experiment.xlsx’ | Directory to excel file containing experiment data |
| ‘experiment\_id’ | Identifier of loaded experiment data. Importance for labeling data fitting from multiple experiments (e.g., using different tracers or different measuring instruments) |
| append\_flag | Append. Value: *true*, *false*  *true*:append experiment data to existed data in *mfamodel*  *false*: overwrite data in *mfamodel* |

**“MSData” sheet**

Instationary MFA requires time-course labeling data and therefore contains an additional column in the “MSData” sheet called “time” and is specified in seconds (please refer to another user manual for INST-MFA).

|  |  |
| --- | --- |
| fragment name | Name of your labeling data entry |
| EMU | Corresponding EMU from metabolic network embedded in the labeling data provided |
| fragment formula | Formula of fragment. Format: lower case for elements. (e.g., c6h12o6)  IMPORTANCE:   * For uncorrected data run, set the formula to fragment formula (e.g., GC-MS derivatized formula) * For fully corrected data run, set the formula to “c*x*” where *x* is the number of carbons in the EMU considered   E.g. fragment g6p\_d[LC-MS], EMU g6p\_d-1,2,3,4,5,6, fragment formula: c6 |
| data | Mass isotopomer distribution. From left to right: m+0, m+1, … For data list that contains m+ fraction higher than number of carbon in EMU, requires the run settings to be *sim\_na* = *true*, *fcor* = *true*. |
| error | User defined error. Number of values provided equal to number of carbons plus one (for m+0) in the corresponding EMU. This number do not equal to number of values in data column. |

**“FluxData” sheet**

|  |  |
| --- | --- |
| flux | Flux ID from *mfamodel*. It is reaction ID + ‘.f’. |
| value | Flux value  For external flux, set it to your experimentally measured value  For dilution flux, set it to 100 (corresponding to 100% of metabolite pool) |
| error | Error of flux value  For external flux, user defined (e.g., standard deviation of replicates of external flux data)  For dilution flux, set it to 0.0001 |

**“TracerData” sheet**

If the model contains other carbon sources that are not listed in this tab, they are assumed to be labeled based on natural abundance of 13C (i.e., 1.12%)

|  |  |
| --- | --- |
| Metabolite | Corresponding metabolite in the network matches with the tracer used.  IMPORTANCE: Use the metabolite corresponding to sink metabolite, which is at the outer-most edge of the network. E.g. “\_out” suffix |
| Tracer Name | User defined name of the tracer. E.g., 2-13C-Glucose |
| fraction | Molar fraction of labeled tracer provided (in total amount of substrate). |
| nCarbons | Number of carbons in the tracer |
| position | Position of labeled carbon(s) in the tracer |
| purity | Purity of tracer (e.g., 99.5%, specified by the supplier) |

## 3.5. emutracer.m

*[emod, emus] = emutracer(mfamodel)*

Trace EMUs and provide EMU networks for simulation.

## 3.6. fluxestimate\_fast.m and flxestimate\_proper.m

Input

|  |  |
| --- | --- |
| emod | Model loaded with information: Stoichiometry, EMU networks, and derivatives. |
| repeat | Number of repeats for minimization of SSR. Value: int. Default: 100. |
| randseed | <IN DEVELOPMENT> Seed for random function, allow reproducibility. |

Output

|  |  |
| --- | --- |
| res | Result of best solution corresponding to minimum SSR |
| foptCell | List of SSR values from multiple repeats (i.e., numbers of elements equal *repeat* input) |
| residualCell | List of results (identical in format to *res*) from multiple repeat (i.e., numbers of elements equal *repeat* input) |

*[res, foptCell] = flxestimate\_fast(emod)*

Suitable for a fast MFA run for model and input data debugging purpose. First, perform minimization at step size of 1e-6, number of temporary solutions equal to user‑defined *multistart* in *defopt.m*. Then, select the best solution among the list. At the best solution, continue to search for more optimal solution but with the smallest step size of MATLAB’s limit (*eps =* 2.2e-16).

*[res, foptCell, residualCell] = flxestimate\_proper(emod, repeat, randseed)*

Proper multistart run. Same principle with *flxestimate\_fast.m*. *multistart = 1* (true multistart is now called *repeat*).

## 3.7. confintestimate.m

*emod.minset = minconfset(emod)*

*[res, impres] = confintestimate(res, emod)*

First, define minimal set to *emod* by running *minconfset.m*. Then, perform flux confidence interval estimation using *confintestimate.m*.

# 4. Known errors and diagnostic suggestions

## 4.1. Errors at xls2MFAmodel.m

**Program is stuck at this step.**

Typically, this step should take at most 1 minute. This is most likely due to the empty rows (after the last row of your spreadsheet) is being read. To resolve this, go to your spreadsheet and select all the empty rows (e.g., from the empty row right after the row with data to the last empty row accessible by Ctrl + DownArrow, or you can put your cursor at the first empty row then press Ctrl + Shift + DownArrow), and delete the whole rows.

**Other suggestions**

Make sure your metabolic model is consistent throughout. Consistent IDs usage and typo-free.

## 4.2. Errors at includemapping.m

**Program is stuck at this step.**

Typically, this step should take at most 10 minutes. This is most likely due to the empty rows (after the last row of your spreadsheet) is being read. To resolve, see section 4.1 error of the same title.

**MATLAB function would raise a warning if it detected inconsistency in mappings.**

Common errors:

* Space character in “maps” column of input AMM
* Missing mapping within a reactions

Suggestions:

* Copy functions out to an empty “.m” file and run. The problematic reaction mappings will show when the errors stop the scripts.

## 4.3. Errors at defopt.m

None recorded. Make sure to select the appropriate settings depending on your input data.

## 4.4. Errors at loadexptdata.m

**Dimension mismatch**

There is likely a typo somewhere or you forgot to add a metabolite(s).

**Program is stuck at this step.**

Typically, this step should take at most 30 minutes to an hour. This is most likely due to the empty rows (after the last row of your spreadsheet) is being read. To resolve, see section 4.1 error of the same title.

## 4.5. Errors at emutracer.m

Take abnormal long time (e.g., > 30mins) or the returning results show an abnormally large number of EMUs (e.g., magnitude > 1e3):

* Cause 1: Assembly of monomer to polymer of large EMU size
  + Suggested fix: Remove polymer assembly, use monomer in metabolism instead
* Cause 2: Cycles involving large molecules
  + Example: Fatty acid biosynthesis and beta-Oxidation involving C16 to higher EMU sizes
  + Suggested fix: validate and remove pathway based on experimental evidence or replace pathways with lump reactions that do not involve EMUs (e.g., the result of fatty acid biosynthesis-degradation cycle is transhydrogenation from NADPH to NADH)
* Cause 3: Input labeling data is a cofactor or related to a cofactor
  + Example: AMP in labeling data, can be phosphorylated to ATP and involve with multiple reactions in the model
  + Suggested fix: Remove the labeling data, or dilute precursors, then synthesized diluted metabolite from diluted precursor. Remove any connections between EMU network and fitted metabolite.

## 4.6. Errors at flxestimate\_fast.m or flxestimate\_proper.m

**NaN values reported**

Something wrong with the starting point and solution search path. Need to re-initiate the whole run from the beginning (even data reloading)

## 4.7. Errors at confintestimate.m

**Take an abnormally long time**

This is problem dependent, typically happens for a large metabolic network. This happens possibly due to an unresolved flux confidence interval. Maximum wait time of 3 hours is recommended before pre-maturely terminating a confidence interval search for a particular reaction. To terminate, pause the program, execute MATLAB scripts line-by-line until you reach the line right before the termination criteria of “f > fmax”, then change your f to some value bigger than fmax to manually terminate.

# 5. References

1. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab Eng. Academic Press; 2006;8: 324–337. doi:10.1016/J.YMBEN.2006.01.004

2. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. Metab Eng. Academic Press; 2007;9: 68–86. doi:10.1016/J.YMBEN.2006.09.001

3. Crown SB, Long CP, Antoniewicz MR. Optimal tracers for parallel labeling experiments and 13C metabolic flux analysis: A new precision and synergy scoring system. Metab Eng. Academic Press; 2016;38: 10–18. doi:10.1016/J.YMBEN.2016.06.001

4. Antoniewicz MR. A guide to 13C metabolic flux analysis for the cancer biologist. Exp Mol Med. Nature Publishing Group; 2018;50: 19. doi:10.1038/s12276-018-0060-y

5. Gopalakrishnan S, Maranas CD. 13C metabolic flux analysis at a genome-scale. Metab Eng. Academic Press; 2015;32: 12–22. doi:10.1016/J.YMBEN.2015.08.006

6. Burgard AP, Nikolaev E V., Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 2004;14: 301–312. doi:10.1101/gr.1926504