

DISCOPOLIS : an algorithm for uniform sampling of metabolic flux distributions via iterative sequences of linear programs

Philippe Bogaerts*, Marianne Rooman**

*3BIO-BioControl, **3BIO-BioInfo, Université Libre de Bruxelles
Av. F.-D. Roosevelt 50 C.P. 165/61, B-1050 Brussels, Belgium
(e-mail: philippe.bogaerts@ulb.ac.be ; mrooman@ulb.ac.be)

Abstract: Mathematical models of metabolic networks are often underdetermined systems with more unknown fluxes than available equality constraints describing mass balances and external flux measurements. After reduction of the flux space based on the available equality constraints, the admissible reduced fluxes belong to a convex polytope defined by the intersection of half-planes representing the inequality constraints (e.g., upper and lower bounds of the fluxes). Random uniform sampling of this polytope allows building marginal distributions for each flux and computing the mean solution representative of the mean metabolism exhibited by the studied organism. This contribution proposes a new algorithm based on DIScrete Sampling of CONVex POLYtopes via Linear program Iterative Sequences (DISCOPOLIS), in which the linear programs are iteratively used to constrain the solutions inside the polytope, taking into account all the previously estimated fluxes. The solutions are weighted to ensure sampling uniformity.

© 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Metabolic network, Metabolic Flux Analysis, Flux Balance Analysis, Flux Variability Analysis, Most Accurate Fluxes, uniform sampling, hit-and-run methods, underdetermined systems, constrained-based modeling

1. INTRODUCTION

The analysis of metabolic networks can be tackled with several methodologies. The most common one is Metabolic Flux Analysis (MFA) (Stephanopoulos et al., 1998) which aims at determining the values of the metabolic fluxes based on algebraic linear equations representing the mass balances of the intracellular metabolites and inequality constraints corresponding to lower and upper bounds for the fluxes. The quasi steady-state approximation, based on the assumption that intracellular metabolites do not accumulate, allows reducing the system of mass balances to a simple system of algebraic equations. The latter is classically augmented with the measurements that are available for some fluxes, usually describing exchanges between the inside and the outside of the cells. In most cases met in this constrained-based modeling context, the number of equations describing mass balances and measurements is less than the number of unknown fluxes, hence leading to an underdetermined system. Minimum and maximum values of each flux can be determined with linear programs. This methodology corresponds to Flux Variability Analysis (FVA) (Mahadevan and Schilling, 2003) or Flux Spectrum Approach (FSA) (Llaneras and Picó, 2007). The minimum and maximum flux values can also be obtained through complex analysis leading to the decomposition of all the admissible flux distributions as combinations of Elementary Flux Modes or Extreme Pathways (Klamt and Stelling, 2003). Another approach for dealing with system underdeterminacy is Flux Balance Analysis (FBA) (Orth et al., 2010a), which assumes an

optimal metabolic behavior of the cells, described with an objective cost function consisting of a linear combination of some fluxes. The choice of the objective cost function to be minimized or maximized is here a key issue. One of the most commonly used choices corresponds to biomass growth maximization. It is worth noting that the optimal value of the cost function can often be reached with different flux distributions, hence keeping an underdetermined system. Recently, another kind of optimal solution, based on the concept of Most Accurate Fluxes (MAF) (Mhallelm Gziri and Bogaerts, 2019) has been proposed with the advantages, on the one hand, to require no assumptions regarding an optimal biological behavior and, on the other hand, to guarantee a unique solution with a very low computational load.

Besides these approaches, randomly sampling the set of existing solutions to the underdetermined system allows building marginal distributions for each flux and, especially, determining the mean value which should be representative of the mean metabolic behavior of the cell under specific conditions. After reduction of the flux space based on the available equality constraints, the admissible reduced fluxes belong to a convex polytope defined by the intersection of half-planes representing the inequality constraints. There are several methods to sample uniformly such a polytope. The most simple and intuitive one is the rejection technique (Rubinstein, 1982) which consists in, firstly, determining a regular shape, e.g. a rectangular parallelepiped corresponding to the minimum and maximum values of the fluxes, that encloses the polytope of solutions and, secondly, uniformly

sampling the enclosing shape and keeping only the samples which satisfy the constraints defined by the half-planes. This method is not well suited to complex networks as the fraction of samples to be rejected increases dramatically with the dimension of the solution space. Hit-and-run methods (Smith, 1984) follow random walks which determine samples included in the solution space. These algorithms converge to sets of uniformly distributed samples. However, a well-known drawback of hit-and-run methods is that, if the polytope of solutions has an irregular shape with some highly elongated directions, the samples get often stuck in some parts of the polytope, which prevents the algorithm to converge to a global uniform sampling of the whole set. To tackle this problem, other methods have been proposed, e.g. the artificial centering hit-and-run method (ACHR) (Kaufman and Smith, 1998), which has been implemented in the COBRA toolbox (Schellenberger et al., 2011). However, this method has not been proven to converge to uniformly distributed samples. Recently, the coordinated hit-and-run with rounding (CHRR) method (Haraldsdóttir et al., 2017) has been added to the COBRA toolbox. Its key feature is to preprocess the solution polytope by, firstly, determining the ellipsoid with largest volume which can be inscribed in the polytope and, secondly, transforming it to a unit ball. After this rounding step, a classical coordinated hit-and-run is applied in the unit ball and the obtained samples are projected in the original space through the inverse of the transform that was applied to the ellipsoid. The hit-and-run becomes then very efficient given that all the points within the ball are solutions to the problem and the shape is perfectly regular. However, as will be shown below on a toy example, the points that are lost by the ellipsoid can bias the mean of the samples, which necessarily corresponds to the center of the ellipsoid.

In this contribution, a new method is presented, which consists in determining each sample of the distribution by choosing first randomly the order of the fluxes to be determined, and then fixing iteratively their values using uniform random sampling on intervals defined by modified linear programs. The latter take into account all the inequality constraints defined, on the one hand, by the half-planes forming the original polytope and, on the other hand, by the equality constraints corresponding to the flux values fixed in the previous iterations. Although this algorithm does not - strictly speaking - lead to uniform samples, we recover uniformity by attaching weights to the samples, which correct for the reduction of the sampling space upon fixing some flux values. This weighting procedure yields a rigorous determination of the mean value of the flux distribution.

The paper is organized as follows. Section 2 recalls the definition of the solution polytope for a metabolic network. Section 3 presents the DISCOPOLIS algorithm. Section 4 illustrates its use, first, on a toy example which shows the bias obtained with the rounded polytope of the CHRR method and the unbiased result obtained with DISCOPOLIS and, secondly, on the core metabolic network of *Escherichia coli* (Orth et al., 2010b). Conclusions and perspectives are proposed in Section 5.

2. THE POLYTOPE OF SOLUTIONS FOR THE FLUX DISTRIBUTION OF A METABOLIC NETWORK

2.1 Metabolic Flux Analysis (MFA) and Flux Variability Analysis (FVA)

We consider the general case of a metabolic network linking m intracellular metabolites through n metabolic fluxes.

Assuming that intracellular metabolites do not accumulate inside the cell (and neglecting the dilution effect of intracellular concentrations due to cell growth), the system of mass balances of the intracellular metabolites reduces to a simple set of m algebraic homogenous equations with n unknowns

$$Nv = 0 \quad (1)$$

where $N \in \mathfrak{R}^{m \times n}$ is the stoichiometric matrix and $v \in \mathfrak{R}^n$ the metabolic fluxes (in $\text{mol} \cdot \text{cell}^{-1} \cdot \text{h}^{-1}$). Other linear equality constraints can usually be added to system (1), typically resulting from measurements of some exchange fluxes. All the equality constraints which link the metabolic fluxes (mass balances (1), measurements of exchange fluxes, etc.) will be concatenated in

$$A_e v = b_e \quad (2)$$

with $A_e \in \mathfrak{R}^{n_e \times n}$, $b_e \in \mathfrak{R}^{n_e}$.

The metabolic fluxes are also subject to inequality constraints which can be concatenated in

$$Av \leq b \quad (3)$$

with $A \in \mathfrak{R}^{n_i \times n}$ and $b \in \mathfrak{R}^{n_i}$. These constraints mainly include lower and upper bounds for the fluxes but may also contain additional inequalities based on some biological assumptions, e.g. regarding overflow metabolism (Richelle et al., 2016; Bogaerts et al., 2017).

MFA aims at solving system $\{(2),(3)\}$ which is, in most cases, underdetermined with the number of equations (n_e) lower than the number of unknown fluxes (n). Minimum and maximum values can be determined for each flux $v(i)$, $i \in [1, n]$, by solving $2n$ linear programs (LPs)

$$v_i^{MIN,MAX} = \underset{v}{\text{Min, Max}} v_i \quad \forall i \in [1, n] \quad (4)$$

subject to $\{(2),(3)\}$.

2.2 Elimination of the Equality Constraints and Definition of the Polytope of Solutions

Let us define $A_0 \in \mathfrak{R}^{n \times (n-n_e)}$ as the matrix whose columns represent the orthonormal basis for the null space of A_e , i.e. the set of all $v \in \mathfrak{R}^n$ such that $A_e v = 0$. Using the equality $A_e A_0 = 0$, any solution v of (2) can be decomposed into

$$v = v_0 + A_0 q \quad (5)$$

where v_0 is a particular solution of (2), i.e. $A_e v_0 = b_e$, and $q \in \mathfrak{R}^{n_q}$.

In this subspace of dimension n_q , the inequality constraints (3) become

$$A' q \leq b' \quad (6)$$

with

$$A' = A A_0 \quad (7)$$

$$b' = b - A v_0 \quad (8)$$

A particular solution v_0 is the parsimonious solution obtained with the quadratic program (QP)

$$v_0 = \underset{v}{\text{Min}} v^T v \quad (9)$$

subject to $\{(2),(3)\}$.

In this new subspace of dimension n_q , the original problem is now fully described by the set of inequalities (6). The intersection of the corresponding half-planes defines the convex polytope of solutions for the reduced flux distribution q . Note that, regarding the notations used in the following sections, we will systematically define the solution polytope with the set of inequalities (2) instead of (6), assuming that the appropriate reduction procedure $\{(5),(6),(7),(8)\}$ has been applied.

3. THE DISCOPOLIS ALGORITHM

3.1 The Rejection Algorithm

We first recall in Fig. 1 the rejection algorithm which will help introducing the DISCOPOLIS algorithm in the next section.

Input : solution polytope defined by A and b ; number of samples N ; minimum and maximum values of the fluxes v_i^{MIN} and v_i^{MAX} ($i \in [1,n]$) obtained with Flux Variability Analysis;
Output : N samples $v(k) \in \mathbb{R}^n$ ($k \in [1,N]$);

```

1  for  $k = 1$  to  $N$  do
2      for  $i = 1$  to  $n$  do
3          generate  $v_i^*$  from a uniform distribution on  $[v_i^{MIN}, v_i^{MAX}]$ ;
4          end
5          while  $v^*$  (with elements  $v_i^*$ ,  $i \in [1,n]$ ) does not satisfy  $A v^* \leq b$  do
6              redo previous loop 2 to 4;
7          end
8           $v(k) = v^*$ ;
9  end
```

Fig. 1. Rejection algorithm.

The samples $v(k)$ which are the outputs of this rejection algorithm genuinely correspond to a uniform distribution. Indeed, the values of the probability density function are all equal to

$$p[v(k)] = \frac{w(k)}{\sum_{k=1}^N w(k)} \quad (10)$$

with

$$w(k) = \prod_{i=1}^n \frac{1}{v_i^{MAX} - v_i^{MIN}} \quad \forall k \quad (11)$$

hence leading to the well-known formula for the mean of uniform samples

$$\bar{v} = \sum_{k=1}^N p[v(k)] v(k) = \frac{1}{N} \sum_{k=1}^N v(k) \quad (12)$$

As mentioned in Section 1, this algorithm cannot be used for polytopes in high dimensional space given the fraction of rejected samples which rapidly explodes the computational load.

3.2 The DISCOPOLIS Algorithm

The DISCOPOLIS (DIcrete Sampling of CONvex POLYtopes via Linear program Iterative Sequences) algorithm is presented in Fig. 2. For each sample $v(k)$, a first flux index i is randomly selected in $[1,n]$ (line 6) and its value v_i is randomly chosen from a uniform distribution between its minimum and maximum values (lines 8 and 9). Rather than a continuous sampling on $[v_i^{MIN}, v_i^{MAX}]$, a discretized uniform sampling is implemented by defining S equally spaced grid points in this interval and choosing randomly one grid point from a discrete uniform distribution on $[1,S]$. This approach is a key feature of the algorithm, which is indispensable when the number of fluxes is large and will be justified below. After fixing this first flux v_i , the loop from lines 10 to 24 iteratively fixes the values of all the other fluxes. After having randomly chosen a new index i (line 12), a new interval of solutions $[v_i^{MINnew}, v_i^{MAXnew}]$ is computed (lines 14 and 15) based on 2 LPs subject to the inequality constraints (2) defining the solution polytope and the set of equality constraints (line 11) which takes into account all the flux values fixed in the previous iterations. The number S^{new} of grid points remaining in the reduced interval is then computed in line 16 and one grid point is randomly chosen from a uniform distribution on $[1,S^{new}]$ (line 18). The corresponding value v_i is computed on line 19. An exception is considered if the reduced interval $[v_i^{MINnew}, v_i^{MAXnew}]$ only contains a single grid point, in which case v_i is set equal to the center of the interval (line 21).

The crucial update of the weights on line 23 takes into account that the uniform sampling on the reduced interval $[v_i^{MINnew}, v_i^{MAXnew}]$ (due to the additional constraints in line 11 corresponding to the flux values fixed at the previous iterations) artificially increases the probability of selecting one of the new grid points.

```

Input : solution polytope defined by  $A$  and  $b$ ; number of samples  $N$ ;
number of grid points  $S$ ; minimum and maximum values of the fluxes
 $v_i^{MIN}$  and  $v_i^{MAX}$  ( $i \in [1, n]$ ) obtained with Flux Variability Analysis
Output :  $N$  samples  $v(k) \in \mathbb{R}^n$  ( $k \in [1, N]$ ) with their weights  $w(k)$ 

1   $A_{eq} = \emptyset$ ;  $b_{eq} = \emptyset$ ; /* initialize empty matrices for equality constraints
2   $L_i = (v_i^{MAX} - v_i^{MIN}) / (S - 1)$ ; /* compute for each flux  $v_i$  the interval
   between 2 grid points
3  for  $k = 1$  to  $N$  do
4     $w(k) = 1$ ; /* initialize weight of the  $k^{\text{th}}$  sample
5     $I = [1, n]$ ; /* set of all indexes  $i$  of all the fluxes  $v_i \in v$ 
6    Randomly select an index  $i$  in  $I$ ;
7    Remove index  $i$  from set  $I$ ;
8    Generate one index  $g$  from a uniform distribution on  $[1, S]$ ;
9     $v_i = v_i^{MIN} + (v_i^{MAX} - v_i^{MIN}) * (g - 1) / (S - 1)$ ; /* discrete uniform
   sampling of  $v_i$  corresponding to the  $g^{\text{th}}$  grid point
10   while  $I \neq \emptyset$  do
11     Augment  $A_{eq}$  and  $b_{eq}$  to account for last fixed  $v_i$ ;
12     Randomly select an index  $i$  in  $I$ ;
13     Remove index  $i$  from set  $I$ ;
14      $v_i^{MINnew} = \min_v v_i$  computed with LP subject to  $A*v \leq b$ 
       and  $A_{eq}*v = b_{eq}$ ;
15      $v_i^{MAXnew} = \max_v v_i$  computed with LP subject to  $A*v \leq b$ 
       and  $A_{eq}*v = b_{eq}$ ;
16      $S^{new} = 1 + \text{floor}((v_i^{MAXnew} - v_i^{MINnew}) / L_i)$ ; /* number of grid
       points remaining in the new constrained solution interval
17     if  $S^{new} > 1$  then
18       Generate one index  $g$  from a uniform distribution on  $[1, S^{new}]$ ;
19        $v_i = v_i^{MINnew} + (v_i^{MAXnew} - v_i^{MINnew}) * (g - 1) / (S^{new} - 1)$ ; /* discrete
       uniform sampling of  $v_i$  corresponding to the  $g^{\text{th}}$  grid point
20     else
21        $v_i = (v_i^{MAXnew} + v_i^{MINnew}) / 2$ ; /* use of the center of the new
       solution interval in case of only 1 remaining grid point
22     end
23      $w(k) = w(k) * S^{new} / S$ ; /* update weight of the  $k^{\text{th}}$  sample
24   end
25 end

```

Fig. 2. DISCOPOLIS algorithm.

Indeed, a thoughtless continuous uniform sampling on $[v_i^{MINnew}, v_i^{MAXnew}]$, without taking into account its reduction from $[v_i^{MIN}, v_i^{MAX}]$, would correspond to replace the factors $(v_i^{MAX} - v_i^{MIN})^{-1}$ in (11) with the larger values $(v_i^{MAXnew} - v_i^{MINnew})^{-1}$ given the new, narrower, constrained intervals.

The update in line 23 cancels this effect through the correcting factor $(v_i^{MAXnew} - v_i^{MINnew}) / (v_i^{MAX} - v_i^{MIN})$ or, equivalently in the case of discrete sampling using grid points, the correcting factor S^{new} / S . Our procedure can be viewed as a generalized uniform sampling inside the polytope, in which each sample has a specific weight. It allows recovering the rigorous determination of the mean value of a uniform flux distribution using

$$\bar{v} = \sum_{k=1}^N p[v(k)]v(k) \quad (13)$$

with $p[v(k)]$ given by (10) and $w(k)$ by the output of the DISCOPOLIS algorithm.

Finally, the use of grid points for a discrete sampling in the constrained intervals prevents the weights to rapidly tend to 0 as the number of constraints in line 11 accumulate. This is especially important for solution polytopes that are generally extremely constrained and irregular. The lower bound of $w(k)$ is indeed $1 / S^{(n-1)}$, corresponding to the extreme case of a sample whose fluxes are almost perfectly determined when one particular flux has been fixed at the first iteration. As will

be illustrated on case studies in Section 4, decreasing the number of grid points S leads to a higher fraction of samples $v(k)$ whose sum of weights represents a high proportion of the total sum of weights, *i.e.* a higher fraction of samples with significant probability densities for computing the mean (13). There is thus a trade-off between the increase of the number of grid points to reach good precision, and the decrease of their number to get a reasonable fraction of samples with relatively high weights.

4. CASE STUDIES

4.1 Toy Example

This toy example consists of a simple 2D polytope for which the mean of a genuine uniform sampling can easily be computed through the rejection algorithm. The first objective is to illustrate that the use of the ellipsoid with largest volume inscribed in the solution polytope (CHRR method, Haraldsdóttir et al., 2017) may lead to a biased estimation of the flux distribution mean. The second objective is to prove that the DISCOPOLIS algorithm provides accurately the mean of the genuine uniform sampling.

Let us consider the polytope defined by fluxes $v^T = [v_1 \ v_2]$ belonging to the intersection of half-lines (3) with

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \quad (14)$$

To accurately determine the mean of the flux distribution, we apply the rejection algorithm (Fig. 1) to generate 5,000 uniformly distributed random samples. The result is presented in Fig. 3, the straight line pointing to the mean of the distribution which is equal to $\bar{v}^T = [0.78 \ 0.45]$ (green diamond).

The *preprocess* routine of the COBRA toolbox computes the ellipsoid with largest volume inscribed in the solution polytope, which consists of the rounding step of the CHRR method. The center of the ellipsoid $\{(3),(14)\}$ is $\bar{v}^T = [0.75 \ 0.50]$, and thus differs significantly from the mean of the uniform distribution because of the points lost by the inscribed ellipsoid. Notably, any subsequent hit-and-run sequence of samples provided by the CHRR algorithm necessarily has its mean positioned on the center of this ellipsoid.

The results provided with the DISCOPOLIS algorithm (with different numbers of samples, numbers of grid points and seed values for the random generator) are provided in Table 1 and represented in Fig. 3. They show that the mean of the genuine uniform sampling performed with the rejection algorithm is recovered with very high accuracy by the DISCOPOLIS algorithm. Moreover, the results are robust with respect to the tuning parameters and random seeds.

Table 1. Mean of the flux distribution in the solution polytope $\{(3),(14)\}$, computed with (13) and the samples and weights obtained from the DISCOPOLIS algorithm (Fig. 2). N is the number of samples, S the number of grid points. The 2 columns for each pair (N,S) correspond to 2 different seed values ($rng(0)$ and $rng(1)$) of the random number generator ($rand$ in Matlab).

		$N = 10^3$		$N = 10^4$	
		seed 1	seed 2	seed 1	seed 2
$S = 10^2$	v_1	0.78	0.80	0.78	0.78
	v_2	0.46	0.45	0.45	0.44
$S = 10^3$	v_1	0.78	0.79	0.77	0.78
	v_2	0.46	0.44	0.45	0.44

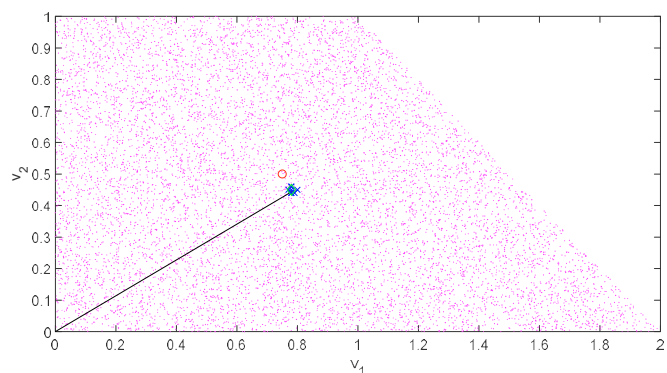


Fig. 3. 5,000 random samples (magenta dots) uniformly distributed on the solution polytope defined by $\{(3),(14)\}$, generated with the rejection algorithm (Fig. 1). The straight-line points to the mean of the flux distribution $\bar{v}^T = [0.78 \ 0.45]$ (green diamond). The 8 blue crosses correspond to the DISCOPOLIS solutions provided in Table 1 and the red circle to the center of the inscribed ellipsoid with largest volume.

4.2 Core Metabolic Network of *Escherichia coli*

This second case study consists of the core metabolic network of *Escherichia coli* (Orth et al., 2010b). The COBRA model is available in the COBRA toolbox (*Ecoli_core_model.mat*). It consists of 95 fluxes with upper and lower bounds. As proposed in the supplementary tutorial of Haraldsdóttir et al. (2017), we set the maximum glucose uptake rate to 18.5 mmol/gDW/h and we remove the cellular objective (no FBA). We only consider the aerobic model, with unlimited oxygen uptake. There are 72 intracellular metabolites which are assumed balanced, hence leading to a stoichiometric matrix $N \in \mathfrak{R}^{72 \times 95}$ with a rank equal to 67. The equality constraints (2) correspond to the mass balances (1) augmented with 9 fluxes which are fixed because the difference between their maximum and minimum values (computed with FVA through (4)) is less than 10^{-6} . The rank of $A_e \in \mathfrak{R}^{81 \times 95}$ is equal to 72.

Hence, the dimension of the reduced flux vector q (see Section 2.2) is equal to 23. The matrix $A' \in \mathfrak{R}^{172 \times 23}$ defines the solution polytope (6) and corresponds, through (7), to the lower and upper bounds on the original fluxes (except for the 9 fluxes whose values are fixed).

The mean of the flux distribution is computed with the DISCOPOLIS algorithm for different numbers of samples N (10^3 , 10^4 , 10^5), different numbers of grid points S (10 and 10^2) and 2 different seed values. The solution with the largest number of samples ($N = 10^5$) and 10 grid points is considered as the reference and is compared to the other solutions in Table 2. The linear correlation coefficients R^2 between the reference and the other DISCOPOLIS solutions are extremely close to 1 (0.99 in the worst case) for the 4 solutions of the first row which use 10 grid points. The correlation is significantly lower (between 0.59 and 0.91) when using 100 grid points. This *a priori* counterintuitive result can be explained by looking at the percentage of samples whose sum of weights represents 99.9% of the total sum of weights. The very low percentages obtained for $S = 100$ (1.3% at most) show that the mean value (13) depends on a very low number of samples, which leads to less robust results. In contrast, the mean calculated with 10 grid points involve approximately 70% of the samples. This illustrates the key importance of the discretization used in the algorithm.

Fig. 2 provides the flux mean values for the reference solution (blue crosses) and the intervals between minimum and maximum values obtained with the 4 other solutions using 10 grid points (first row in Table 2). These results are compared with the flux mean values based on the center of the largest inscribed polytope (red circles). Although both algorithms lead to very close results for many fluxes, several other fluxes have a significantly different mean, which is probably the consequence of some parts of the solution polytope lost by the largest inscribed ellipsoid.

Table 2. At the top of each cell, linear correlation coefficients R^2 for assessing the fitting between the mean of two flux distributions obtained with the DISCOPOLIS algorithm: one corresponding to N samples and S grid points, the other being the reference corresponding to 10^5 samples and 10 grid points. At the bottom, percentage of samples whose sum of weights represents 99.9% of the total sum of weights. The 2 columns for each pair (N,S) correspond to 2 different seed values ($rng(0)$ and $rng(1)$) of the random number generator ($rand$ in Matlab).

		$N = 10^3$		$N = 10^4$	
		seed 1	seed 2	seed 1	seed 2
$S = 10$	R^2	0.99	0.96	0.995	0.9996
	%	74%	73%	69%	70%
$S = 10^2$	R^2	0.87	0.59	0.91	0.75
	%	0.6%	0.9%	0.6%	1.3%

5. CONCLUSIONS AND PERSPECTIVES

The DISCOPOLIS algorithm iteratively uses linear programs for constraining the flux distribution samples of a metabolic network inside the solution polytope, taking into account all the previously estimated fluxes. The solutions are weighted to ensure sampling uniformity. A first toy example showed that the mean of a genuine uniform distribution could indeed be recovered, which was not the case with the center of the largest inscribed ellipsoid (used in the CHRR method) that provides a biased estimate due to some lost regions of the

solution polytope. The importance of the use of grid points and the robustness of the results has been illustrated on a second example about the core metabolic network of *E. coli*.

Future work will consist in optimizing the algorithm in terms of computational load and in testing its use with more complex networks. Finally, we will investigate the issue of flux rescaling. Rather than considering an equal number of discrete points along each flux, we will vary this number according to the range of original flux values and compare the results.

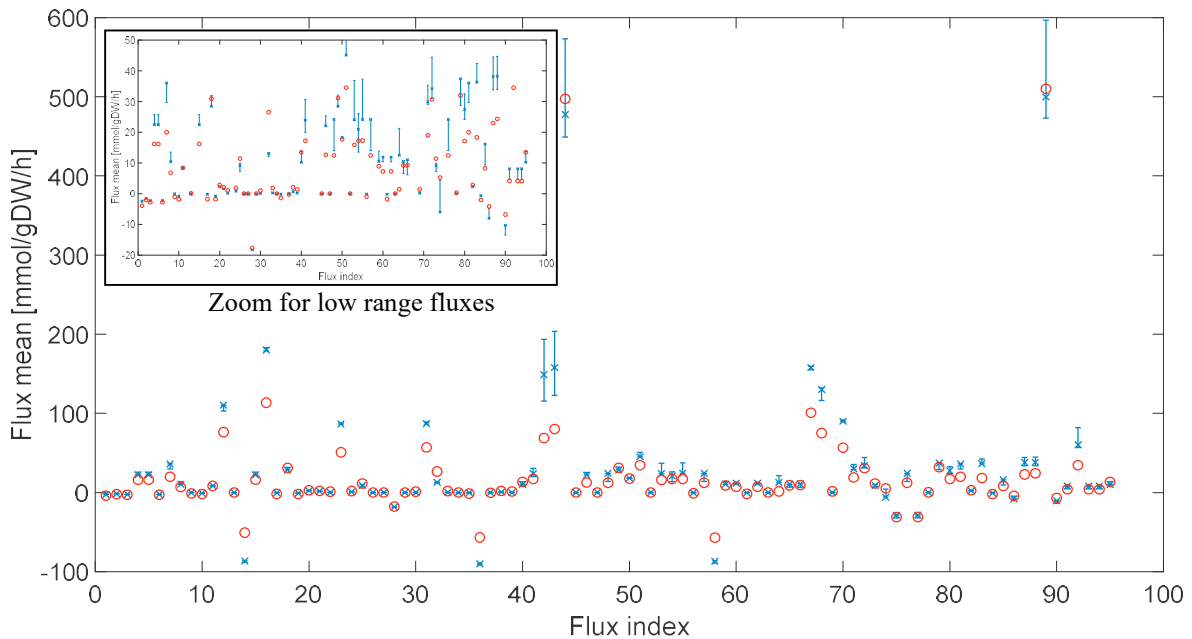


Fig. 4. Mean values of the flux distribution within the *E. coli* core metabolic network, obtained with the DISCOPOLIS algorithm (blue crosses : $N = 10^5$ samples and $S = 10$ grid points; blue bars : intervals between minimum and maximum values obtained with the 4 cases in the first row of Table 2) and with the center of the largest inscribed ellipsoid (red circles).

REFERENCES

- Bogaerts, Ph., Mhallem Gziri, K., and Richelle, A. (2017). From MFA to FBA: Defining linear constraints accounting for overflow metabolism in a macroscopic FBA-based dynamical model of cell cultures in bioreactor. *J. Process Control*, 60, 34-47.
- Haraldsdóttir, H., Cousins, B., Thiele, I., Fleming, R., Vempala, S. (2017). CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33 (11), 1741-1743.
- Kaufman, D., and Smith, R. (1998). Direction choice for accelerated convergence in hit-and-run sampling. *Oper. Res.*, 46 (1), 84-95.
- Klamt, S., and Stelling, J. (2003). Two approaches for metabolic pathway analysis? *Trends in Biotech.*, 21 (2), 64-69
- Llaneras, F., and Picó, J. (2007). An interval approach for dealing with flux distributions and elementary modes activity patterns. *J. of Theor. Biol.*, 246, 290-308.
- Mahadevan, R., and Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, 5, 264-276.
- Mhallem Gziri, K., and Bogaerts, Ph. (2019). Determining a unique solution to underdetermined metabolic networks via a systematic path through the Most Accurate Fluxes. Accepted in *IFAC-PapersOnLine, Proceedings of the 12th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems (DYCOPS 2019)*.
- Orth, J., Thiele, I., and Palsson, B. (2010a). What is flux balance analysis? *Nature Biotechnol.*, 28 (3), 245-248.
- Orth, J., Fleming, R., and Palsson, B. (2010b). Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus*, 1(10).
- Richelle, A., Mhallem Gziri, K., and Bogaerts, Ph. (2016). A methodology for building a macroscopic FBA-based dynamical simulator of cell cultures through flux variability analysis. *Biochem. Eng. J.*, 114, 50-61.
- Rubinstein, R. (1982). Generating random vectors uniformly distributed inside and on the surface of different regions. *Eur. J. Oper. Res.*, 10 (2), 205-209.
- Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J, Feist, A., Zielinski, D., Bordbar, A., Lewis, N., Rahmanian, S., Kang, J., Hyduke, D., Palsson, B. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, 6 (9), 1290-1307.
- Smith, R.. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.*, 32 (6), 1296-1308.
- Stephanopoulos, G., Aristidou, A., and Nielsen, J. (1998). *Metabolic engineering: Principles and methodologies*, chapter 8. Academic Press, San Diego.