

---

# COMPOSITIONAL METABOLIC FLUX ANALYSIS

---

A PREPRINT

✉ **Teddy Groves**<sup>1</sup>, ✉ **Te Chen**<sup>1</sup>, ✉ **Sergi Muyo Abad**<sup>1</sup>, ✉ **Nicholas Luke Cowie**<sup>1</sup>, ✉ **Daria Volkova**<sup>1</sup>, ✉ **Christian Brinch**<sup>2</sup>, and ✉ **Lars Keld Nielsen**<sup>1,3</sup>

<sup>1</sup>The Novo Nordisk Center for Biosustainability, DTU, Kongens Lyngby, Denmark

<sup>2</sup>National Food Institute, DTU, Kongens Lyngby, Denmark

<sup>3</sup>Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, St Lucia 4067, Australia

## ABSTRACT

Metabolic Flux Analysis aims to infer the values of metabolic fluxes from measurements of isotope labelling distributions. Since these distributions are positive, sum-constrained and relatively low-dimensional, we argue that they should be analysed using specialised methods that target compositional data. We illustrate our argument using a simple pedagogical example, then show how compositional analysis leads to improved results on a typical dataset.

## 1 Note to Christian

Metabolism refers to all chemical reactions that occur within living cells, allowing organisms to grow, reproduce, and respond to their environment. These reactions form interconnected networks called metabolic pathways, where each step converts one molecule to another through enzymatic reactions. Just as understanding traffic flow in a city requires knowing not just the road layout but also how many vehicles travel each route, understanding cellular function requires knowing both the structure of metabolic pathways and the rates at which molecules flow through them. These rates are called “metabolic fluxes.” While measuring the concentrations of molecules (metabolites) in a cell is relatively straightforward with modern techniques, determining how quickly these molecules are being produced, converted, or consumed inside the cell remains challenging. Scientists have developed approaches using isotopes—variants of elements with different atomic weights—as tracers to track the flow of atoms through metabolic pathways. These isotopes act like molecular GPS trackers, allowing researchers to follow the journey of specific atoms through the complex network of cellular metabolism. The following section explains how these isotope-based approaches work and introduces a new perspective on analyzing the resulting data.

## 2 Introduction

The rates, or “fluxes”, of metabolic reactions are key to understanding how cells behave. The rates of internal metabolic reactions, which create and destroy chemicals inside of cells, are particularly interesting, but difficult to measure directly. Isotope labelling experiments make it possible to infer these internal fluxes indirectly, thanks to the fact that metabolic reactions split and join molecules in predictable ways.

Specifically, given a known “atom mapping” describing how each reaction in a network moves potentially-labelled atoms, a specification of fluxes, and some known isotope distributions (typically the feed), one can calculate expected isotope distributions for internal metabolites. These internal isotope distributions can also be measured, for example using liquid chromatography and mass spectrometry. Comparing measured and expected isotope distributions makes it possible to say infer flux specifications.

This technique has existed in its modern form for a while, leading to impressive results, but we think that a crucial step is missing. Measured and expected isotope distributions are compositions, but they have not previously been analysed using compositional methods.

## 2.1 Isotopes, isotopomers and isotopologues

Isotopes are atoms whose nuclei have the same number of protons but different numbers of neutrons. Isotopes of the same element have similar chemical properties, but have different atomic masses and physical properties. For example, carbon has three naturally occurring isotopes:  $^{12}\text{C}$ ,  $^{13}\text{C}$  and  $^{14}\text{C}$ , with respective atomic masses 12, 13 and 14.  $^{14}\text{C}$  occurs in negligible quantities, and the natural ratio of  $^{12}\text{C}$  to  $^{13}\text{C}$  is known, making carbon suitable for isotope labelling experiments where  $^{12}\text{C}$  is artificially replaced by  $^{13}\text{C}$ .

Isotopomers are forms of a compound that differ only by substitution of isotopes. For example,  $[1-^{13}\text{C}]$  glucose and  $[2-^{13}\text{C}]$  glucose are isotopomers of glucose that differ only in the isotopes of the carbon atoms in positions 1 and 2 (see figure below). In general, for every  $a$  occurrences of an atom with  $I$  isotopes in a compound, there are  $I^a$  corresponding isotopomers. For example, glucose has six carbon atoms: assuming only  $^{12}\text{C}$  and  $^{13}\text{C}$  isotopes are present, there are  $2^6 = 64$  carbon isotopomers.

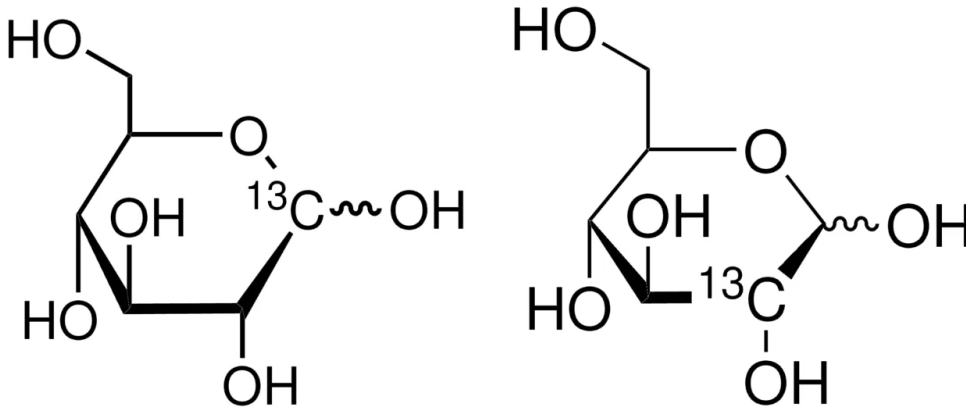


Figure 1: Left:  $[1-^{13}\text{C}]$  glucose Right:  $[2-^{13}\text{C}]$  glucose

Isotopologues are forms of a compound that differ in atomic mass. For example, the isotopomers  $[1-^{13}\text{C}]$  glucose and  $[2-^{13}\text{C}]$  glucose each have five  $^{12}\text{C}$  atoms and one  $^{13}\text{C}$  atom and therefore belong to the same isotopologue  $M_1$  with atomic mass 181.15 g/mol. Isotopologues are important because measurements can often distinguish between isotopologues, but not between isotopomers with the same atomic mass.

## 2.2 Isotope labelling based Metabolic Flux Analysis

In more detail, MFA considers a known metabolic network consisting of  $M$  compounds and  $N$  reactions with stoichiometric coefficients  $S \in \mathbb{R}^{M \times N}$  representing the amount of each compound consumed and produced by each reaction, plus an atom transition map for each reaction. The atom transition map for a reaction specifies in what order the potentially-labelled atoms occur in each of the reaction’s substrates and products.

In this study we focus on the most straightforward case, where it is safe to assume that the network is in isotopic and metabolic steady state. This means that the concentrations of internal metabolites and isotopomers are constant.

The remaining input for MFA is as follows:

- Known isotopomer proportions for some compounds, typically external ones like the feed.
- Measured fluxes for some reactions.
- Measured isotopologue proportions for some compounds, possibly with known measurement error.

The “forward problem” of inferring the label pattern given a flux specification can be solved by writing down a balance equation describing the rate of change of each isotopomer in the network. These equations can be found by combining the atom map and stoichiometric matrix to produce a matrix  $S_i \in \mathbb{R}^{I \times N}$  of stoichiometric coefficients for isotopomers. At isotopic and metabolic steady state we have  $Sv = 0$ , giving  $N$  balance equations. Given some known isotopomer proportions, other isotopomer proportions can be calculated by solving these equations.

While conceptually simple, this approach to solving the forward problem is typically unfeasible due to the prohibitively large number of isotopomers. Practical applications therefore require a method to solve the forward problem more

simply: see [Dai and Locasale, 2017, §3.1] for an overview of approaches to this problem. The most widely used of these is the “elementary metabolite unit” representation introduced in [Antoniewicz et al., 2007].

The inverse problem of inferring steady state fluxes from measured isotopomer or isotopologue distributions can be solved using a statistical model that links these measurements with latent parameters representing flux configurations. In general, such a model specifies the probability  $p(r_{obs} \mid r(v))$  of observing labelling pattern  $r_{obs}$  given a true flux assignment  $v$  and true labelling pattern  $r(v)$ . For example, assuming an independent linear model with measurement standard deviation vector  $\sigma$  yields the following relationship:

$$p(r_{obs} \mid r(v)) = pdf_{normal}(r_{obs} \mid r(v), \sigma)$$

Sometimes a statistical model is not mentioned explicitly, but  $v$  is optimised by minimising a scaled Euclidean distance from  $r(v)$  to  $r_{obs}$ , which produces the same result as optimising by maximising the likelihood  $p(r_{obs} \mid r(v))$ .

## 2.3 Previous work

This section briefly reviews previous work in labelling-based metabolic flux analysis. For more detailed review papers see [Dai and Locasale, 2017], [de Falco et al., 2022].

### 2.3.1 Experimental methods

See the following references:

- [Long and Antoniewicz, 2019]
- [de Falco et al., 2022]

### 2.3.2 Bayesian methods

See the following references:

- [Theorell et al., 2017]
- [Theorell and Nöh, 2020]

### 2.3.3 Software

Existing implementations of MFA include:

- INCA
- 13CFLUX2
- Metran
- OpenFlux(2)
- FluxPyt
- mfapy
- Sysmetab
- iso2flux
- Flux-P
- WUFlux
- OpenMebius
- influx\_s

See [REFERENCE] for reviews of available software implementing MFA.

There is no previous implementation of compositional regression analysis in the context of MFA; all previous implementations apply a linear model either explicitly as in [Theorell et al., 2017, Eq. 3] or more commonly implicitly through the use of Euclidean distance based optimisation.

### 3 Problem statement

Discrepancies between observed and expected isotopologue distributions have previously been modelled using linear regression models, despite the fact that both are compositions. It is well known that, in general, applying non-compositional data analysis methods to compositional data is dangerous because these methods can easily misinterpret constraint-induced correlations [Aitchison, Ch. 3]. This issue is especially pronounced in the case where there are relatively few composition components.

We therefore propose modelling augmenting existing labelling-based MFA methods with specialised compositional data analysis. In this section we draw on previous literature where compositional data analysis to discuss common arguments against the use of specialised compositional methods, explaining why we think they don't apply in the case of labelling based MFA.

#### 3.1 Arguments against using compositional methods

### 4 Simple pedagogical example

### References

- Ziwei Dai and Jason W. Locasale. Understanding metabolism with flux analysis: From theory to application. *Metabolic Engineering*, 43:94–102, September 2017. ISSN 10967176. doi:10.1016/j.ymben.2016.09.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S1096717616301380>.
- Maciek R. Antoniewicz, Joanne K. Kelleher, and Gregory Stephanopoulos. Elementary Metabolite Units (EMU): A novel framework for modeling isotopic distributions. *Metabolic engineering*, 9(1):68–86, January 2007. ISSN 1096-7176. doi:10.1016/j.ymben.2006.09.001. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1994654/>.
- Bruna de Falco, Francesco Giannino, Fabrizio Carteni, Stefano Mazzoleni, and Dong-Hyun Kim. Metabolic flux analysis: A comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas. *RSC Advances*, 12(39):25528–25548, 2022. doi:10.1039/D2RA03326G. URL <https://pubs.rsc.org/en/content/articlelanding/2022/ra/d2ra03326g>.
- Christopher P. Long and Maciek R. Antoniewicz. High-resolution  $^{13}\text{C}$  metabolic flux analysis. *Nature Protocols*, 14(10):2856–2877, October 2019. ISSN 1750-2799. doi:10.1038/s41596-019-0204-0. URL <https://www.nature.com/articles/s41596-019-0204-0>.
- Axel Theorell, Samuel Leweke, Wolfgang Wiechert, and Katharina Nöh. To be certain about the uncertainty: Bayesian statistics for  $^{13}\text{C}$  metabolic flux analysis. *Biotechnology and Bioengineering*, 114(11):2668–2684, 2017. ISSN 1097-0290. doi:10.1002/bit.26379. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.26379>.
- Axel Theorell and Katharina Nöh. Reversible jump MCMC for multi-model inference in Metabolic Flux Analysis. *Bioinformatics*, 36(1):232–240, January 2020. ISSN 1367-4803. doi:10.1093/bioinformatics/btz500. URL <https://doi.org/10.1093/bioinformatics/btz500>.
- J Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall Ltd. ISBN 978-94-010-8324-9.