COMPOSITIONAL METABOLIC FLUX ANALYSIS

A PREPRINT

Teddy Groves [1], **Te Chen** [1], **Sergi Muyo Abad** [1], **Nicholas Luke Cowie** [1], **Daria Volkova** [1], **Christian Brinch** [2], **and Lars Keld Nielsen** [1,3]
[1]The Novo Nordisk Center for Biosustainability, DTU, Kongens Lyngby, Denmark
[2]National Food Institute, DTU, Kongens Lyngby, Denmark
[3]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 Introduction

Labelling-based Metabolic flux analysis is the study of

1.1 Previous work

This section briefly reviews previous work in labelling-based metabolic flux analysis. For more detailed review papers see XXXXX

1.1.1 Experimental methods

1.1.2 The forward problem

- Cumomers
- EMU

1.1.3 Software

- OpenFlux
- Freeflux
- INCA
- 13CFlux
- ...

1.1.4 Bayesian 13C MFA

1.2 Problem statement

The topic of how to statistically model isotope labelling pattern measurements has received relatively little attention in the development of metabolic flux analysis. Most presentations and software applications quantify the discrepancy between a species's measured and predicted isotope labelling distribution using a Euclidean distance, and advocate

choosing a flux configuration whose labelling pattern minimises this distance, possibly with per-species and/or per-isotope-equivalence-class weights. This is equivalent to using maximum likelihood estimation, where the likelihood is given by an indepedent normal distribution centered on the predictedc labelling distribution, with error standard deviations determined by the weights, i.e., for each species s,

$$y_s \sim N(\hat{y_s}, \sigma_s)$$

where y_s is the observed labelling distribution for species s, $\hat{y_s}$ is the predicted labelling distribution and $sigma_s$ is a vector of standard deviations.

There are two key reasons why this approach is flawed in the case where y_s and $\hat{y_s}$ are compositions. First, the Euclidean distance is inappropriate for measuring discrepancies between compositions. Second, the use of an independent error model neglects the fact that composition components are intrinsically correlated. This issue is especially pronounced in the case where there are relatively few composition components.

1.3 Isotopes, isotopologues and mass isotopologues

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

Isotopologues are forms of a compound that differ only by substitution of isotopes. For example, [1-13C] glucose, [U-13C] glucose and [2-13C] glucose are isotopologues that differ only in the isotopes of the carbon atoms in positions 1 and 2. In general, for a compound with A occurrences of an atom with I isotopes, there are I^A corresponding isotopologues. For example, glucose has six carbon atoms: assuming only 12C and 13C isotopes are present, there are 2^6 carbon isotopologues.

A mass isotopologue is an equivalence class of isotopologues that share the same atomic mass. For example, [1-13C] glucose and [2-13C] glucose each have five 12C atoms and one 13C atom and therefore belong to the glucose mass isotopologue M_1 with atomic mass 181.15 g/mol. Mass isotopologues are important because measurements can often distinguish between mass isotopologues, but not between isotopologues with the same atomic mass.

1.4 13C labelling experiments

1.5 13C Metabolic Flux Analysis

13C 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.

The remaining input for 13C MFA is as follows:

- Known isotope proportions for some compounds, typically the feed.
- Measured fluxes for some reactions, possibly with known measurement error.
- Measured mass isotopologue proportions for some compounds, possibly with known measurement error.

The task of inferring the label pattern corresponding to a known flux assignment is known as the "forward problem". [REFERENCE] shows how, assuming that the network is in a metabolic and isotopic steady state, so that neither the concentrations of the compounds nor the distributions of isotopologues are changing, it is possible to calculate the isotopologue distribution for each compound given a known flux; in this way one can calculate the labelling pattern $r(v)$ corresponding to any flux assignment v.

Unfortunately, solving the forward problem in terms of isotopologue is of limited use for real applications due to the prohibitively large number of isotopomers that need to be considered. As a result of this difficulty there has been considerable interest in more concise representations of the forward problem [REFERNECES]. Below [INTERNAL REFERENCE] we consider in detail the "elementary metabolite unit" representation introduced in [Antoniewicz et al., 2007].

The inverse problem of inferring steady state fluxes from measured 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 a linear model, or equivalently optimising v by least squares, yields the following relationship:

$$r_{obs} \sim N(r(v), \Sigma)$$

1.6 The Elementary Metabolite Unit representation

1.7 Compositional Regression

Compositional data is data that is subject to a unit-sum constraint. For example, a compositional dataset might record the amount of fat, protein and other ingredients in some blocks of butter as proportions of the total mass of each block. These proportions are constrained to sum to exactly one.

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].

Compositional regression methods employ constrained measurement distributions to analyse compositional data, allowing induced correlations to be accounted for naturally. Examples of such distributions include the logistic-normal and Dirichlet distributions [Aitchison, Ch. 3] among others.

Compositional regression methods are appropriate for 13C MFA because mass isotopologue distribution vectors are compositional. We therefore considered it likely that the standard practice of applying non-compositional statistical analysis to such data would produce incorrect results.

1.8 Existing solutions

Existing implementations of 13C MFA include:

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

See [Dai and Locasale, 2017], [de Falco et al., 2022]for reviews of available software implementing 13C MFA. We wish to note several limitations of the currently available software:

- There is no previous implementation of compositional regression analysis in the context of 13C 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 least-squares optimisation.
- The only software implementing Bayesian 13C MFA is proprietary.

References

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/`.

J Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall Ltd. ISBN 978-94-010-8324-9.

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.

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.

Axel Theorell, Samuel Leweke, Wolfgang Wiechert, and Katharina Nöh. To be certain about the uncertainty: Bayesian statistics for 13C 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.