

Systems biology

IsoCor: isotope correction for high-resolution MS labeling experiments

Pierre Millard ^{1,*}, Baudoin Delépine ^{1,†}, Matthieu Guionnet ^{1,2},
Maud Heuillet ^{1,2}, Floriant Bellvert ^{1,2} and Fabien Létisse ¹

¹LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse 31077, France and ²MetaToul-MetaboHUB, National Infrastructure of Metabolomics and Fluxomics, Toulouse 31077, France

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Jonathan Wren

Received on October 17, 2018; revised on February 22, 2019; editorial decision on March 18, 2019; accepted on March 20, 2019

Abstract

Summary: Mass spectrometry (MS) is widely used for isotopic studies of metabolism and other (bio)chemical processes. Quantitative applications in systems and synthetic biology require to correct the raw MS data for the contribution of naturally occurring isotopes. Several tools are available to correct low-resolution MS data, and recent developments made substantial improvements by introducing resolution-dependent correction methods, hence opening the way to the correction of high-resolution MS (HRMS) data. Nevertheless, current HRMS correction methods partly fail to determine which isotopic species are resolved from the tracer isotopologues and should thus be corrected. We present an updated version of our isotope correction software (IsoCor) with a novel correction algorithm which ensures to accurately exploit any chemical species with any isotopic tracer, at any MS resolution. IsoCor v2 also includes a novel graphical user interface for intuitive use by end-users and a command-line interface to streamline integration into existing pipelines.

Availability and implementation: IsoCor v2 is implemented in Python 3 and was tested on Windows, Unix and MacOS platforms. The source code and the documentation are freely distributed under GPL3 license at <https://github.com/MetaSys-LISBP/IsoCor/> and <https://isocor.readthedocs.io/>.

Contact: pierre.millard@insa-toulouse.fr

1 Introduction

Mass spectrometry (MS) is extensively used for isotopic studies of metabolism and other (bio)chemical processes. During the last decades, quantitative approaches based on isotope labeling experiments (ILE)—e.g. ¹³C-metabolic flux analysis—have been increasingly used in systems and synthetic biology to infer the topology and operation of biological networks. In such experiments, different labeled forms—isotopologues—of metabolites are quantified from their isotopic clusters in MS spectra. These clusters also contain information on all other isotopes that occur at natural abundance in the molecules. To extract meaningful labeling information—isotopologue

distribution—the contribution of naturally-occurring isotopes has first to be removed (Midani *et al.*, 2017). Several tools were developed to correct isotopic clusters measured with low resolution instruments ($R < 1000$) (Midani *et al.*, 2017; Millard *et al.*, 2012). Carreer *et al.* (2013) proposed a method for the correction of ultra-high resolution ($R > 400\,000$) MS data. However, it is applicable to the correction of only a few elements with no more than two isotopes (C, N or H) and is restricted to ultra-high resolution instruments. Similarly, the correction method implemented in IsoCorrectoR (Heinrich *et al.*, 2018) is restricted to the correction of either low resolution or ultra-high resolution data. AccuCor

(Su *et al.*, 2017) made possible the correction of MS data at any resolution in ^{13}C , ^2H or ^{15}N -ILEs. However, AccuCor is also limited to a few elements and the proposed correction method fails in particular situations (see below). Hence, there was no method to ensure in all situations accurate correction of HRMS data collected with modern MS instruments.

Here, we present a major update of IsoCor to correct MS data, which ensures accurate correction in all cases since it can handle any chemical species, for any isotopic tracer, at any MS resolution.

2 Materials and Methods

State-of-the-art correction methods are based on a correction matrix that relates isotopologue distribution to the measured mass fractions. As detailed in Midani *et al.* (2017), each element (i, j) of the correction matrix must contain the abundance of all isotopic species of the molecule containing j tracer atoms which are unresolved from the mass fraction corresponding to the i th tracer isotopologue. van Winden *et al.* (2002) proposed an elegant formalism to construct low resolution correction matrices by multiplying correction matrices of individual elements. Su *et al.* (2017) extended this formalism for HRMS by accounting for isotopic mass defects, which denote the difference between the exact mass of an atom and the nearest integer (Thurman and Ferrer, 2010). When constructing the matrices for individual elements, Su *et al.* (2017) determine which isotopologues of a given element are resolved from the tracer isotopologues and should thus be corrected. However, information on the total isotopic mass defects of each isotopic species of the molecule—and thus information on the ability of the MS analyzer to distinguish them—is lost when elements are combined during matrix multiplication. This may result in overcorrecting the data when individual mass defects cumulate (e.g. $^{13}\text{C} + ^{15}\text{N}$), as illustrated below and in Figure 1, or in under correcting the data when individual mass defects cancel each other out (e.g. $^{18}\text{O} + ^{34}\text{S}$).

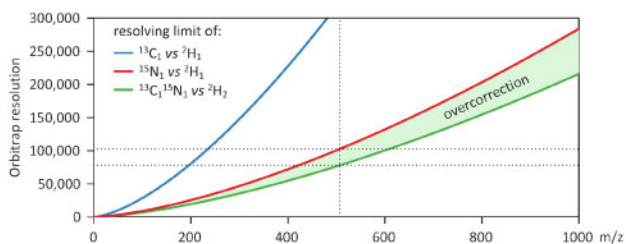


Fig. 1. Example of overcorrection of isotopic data measured at high-resolution on an Orbitrap. For ^2H labeling experiments, the resolving limits of $^{13}\text{C}_1$ and $^{15}\text{N}_1$ versus $^2\text{H}_1$ isotopologues are shown in blue and red, respectively, as detailed in Su *et al.* (2017). Taking ATP as an example (at $m/z = 506$, vertical dotted line), mass fractions must be corrected for the contribution of individual ^{13}C and ^{15}N isotopes to M_1 at resolutions below 323 000 and 102 000, respectively (i.e. below the resolving limits of ^{13}C and ^{15}N isotopes). State-of-the-art algorithms also correct for the combined contribution of all individually-unresolved isotopes, and will thus correct for the contribution of $^1\text{H}_2^13\text{C}_1^15\text{N}_1$ to M_2 . However, the resolving limit of $^1\text{H}_2^13\text{C}_1^15\text{N}_1$ from M_2 is even lower than the resolving limits of individual isotopes (78 000, green line). $^1\text{H}_2^13\text{C}_1^15\text{N}_1$ is thus resolved at resolutions between 78 000 and 102 000, and its contribution should not be corrected. State-of-the-art algorithms thus systematically overcorrect ATP mass fractions measured in this range of resolutions (green area), which unfortunately typically stands in the high-resolution area. Note that there is no such problem at ultra-high resolution or at low-resolution (Color version of this figure is available at *Bioinformatics* online.)

To illustrate this point, let us consider correcting for natural isotope abundance a hypothetical molecule with formula CNH_2 , analyzed at a resolution of 2500, where ^2H is the isotopic tracer. The labeling fractions [$^2\text{H}_0$, $^2\text{H}_1$, $^2\text{H}_2$] of the hydrogen isotopologue distribution are related to the measured mass fractions [M_0 , M_1 , M_2] by the equation

$$\begin{bmatrix} ^2\text{H}_0 \\ ^2\text{H}_1 \\ ^2\text{H}_2 \end{bmatrix} = \begin{bmatrix} \text{Isotopic} \\ \text{correction} \\ \text{matrix} \end{bmatrix}^{-1} \cdot \begin{bmatrix} M_0 \\ M_1 \\ M_2 \end{bmatrix}. \quad (1)$$

Considering the natural abundance of non-tracer elements, we have $^{15}\text{N}_1$ and $^{13}\text{C}_1$ isotopologues. At a resolution of 2500, the minimal difference to resolve two MS peaks is $0.011207 m/z$. $^{15}\text{N}_1$ and $^{13}\text{C}_1$ isotopologues have a mass difference of 0.009242 and $0.002922 m/z$ compared to the $^2\text{H}_1$ labeling fraction. These isotopic species are thus unresolved and their contribution has to be removed from the measured mass fractions. Using the method proposed by Su *et al.* (2017)—and neglecting correction of natural abundance and purity of the tracer element—leads to the following correction matrix:

$$\begin{aligned} \begin{bmatrix} \text{Isotopic} \\ \text{correction} \\ \text{matrix} \end{bmatrix} &= \begin{bmatrix} \text{Nitrogen} \\ \text{matrix} \end{bmatrix} \cdot \begin{bmatrix} \text{Carbon} \\ \text{matrix} \end{bmatrix} \\ &= \begin{bmatrix} \alpha & 0 & 0 \\ (1-\alpha) & \alpha & 0 \\ 0 & (1-\alpha) & \alpha \end{bmatrix} \cdot \begin{bmatrix} \beta & 0 & 0 \\ (1-\beta) & \beta & 0 \\ 0 & (1-\beta) & \beta \end{bmatrix} \\ &= \begin{bmatrix} \alpha \cdot \beta & 0 & 0 \\ \alpha + \beta - 2 \cdot \alpha \cdot \beta & \alpha \cdot \beta & 0 \\ (1-\alpha) \cdot (1-\beta) & \alpha + \beta - 2 \cdot \alpha \cdot \beta & \alpha \cdot \beta \end{bmatrix}, \quad (2) \end{aligned}$$

where α and β are the natural abundance of ^{14}N and ^{12}C isotopes, respectively. A side-effect of this correction method is the automatic (and implicit) correction for the combined contribution of all individually-unresolved isotopes, even if their combination results in well-resolved species (Fig. 1). In this example, the contribution of $^1\text{H}_2^13\text{C}_1^15\text{N}_1$ into M_2 is corrected via the term $(1-\alpha) \cdot (1-\beta)$. However, because mass defects of the individual ^{15}N and ^{13}C isotopes cumulate, $^2\text{H}_0^13\text{C}_1^15\text{N}_1$ is, in fact, resolved from M_2 (with a mass difference of $0.012163 m/z$) and should *not* be corrected. This remark is the essence of the improvement we made from Su *et al.* (2017). At the operating resolution, the correct correction matrix is thus

$$\begin{bmatrix} \text{Isotopic} \\ \text{correction} \\ \text{matrix} \end{bmatrix} = \begin{bmatrix} \alpha \cdot \beta & 0 & 0 \\ \alpha + \beta - 2 \cdot \alpha \cdot \beta & \alpha \cdot \beta & 0 \\ 0 & \alpha + \beta - 2 \cdot \alpha \cdot \beta & \alpha \cdot \beta \end{bmatrix}. \quad (3)$$

IsoCor v2 implements a novel algorithm to account for these effects (Fig. 2). Briefly, each element (i, j) of the correction matrix is obtained by (i) calculating the exact masses and corresponding abundances of *all* the isotopic species of the molecule containing j tracer atoms—using combinatorial probabilities (van Winden *et al.*, 2002)—and (ii) summing the abundance of the species which cannot be distinguished from the mass fraction of the i th tracer isotopologue, according to the operating resolution of the instrument. This algorithm takes into account the ion's charge when constructing the correction matrix, hence allowing correction of singly and multiply-charged ions. Elements (i, j) may also include (or not) the isotopic impurity of the tracer and natural abundance of the tracer element depending whether their contribution has to be corrected.

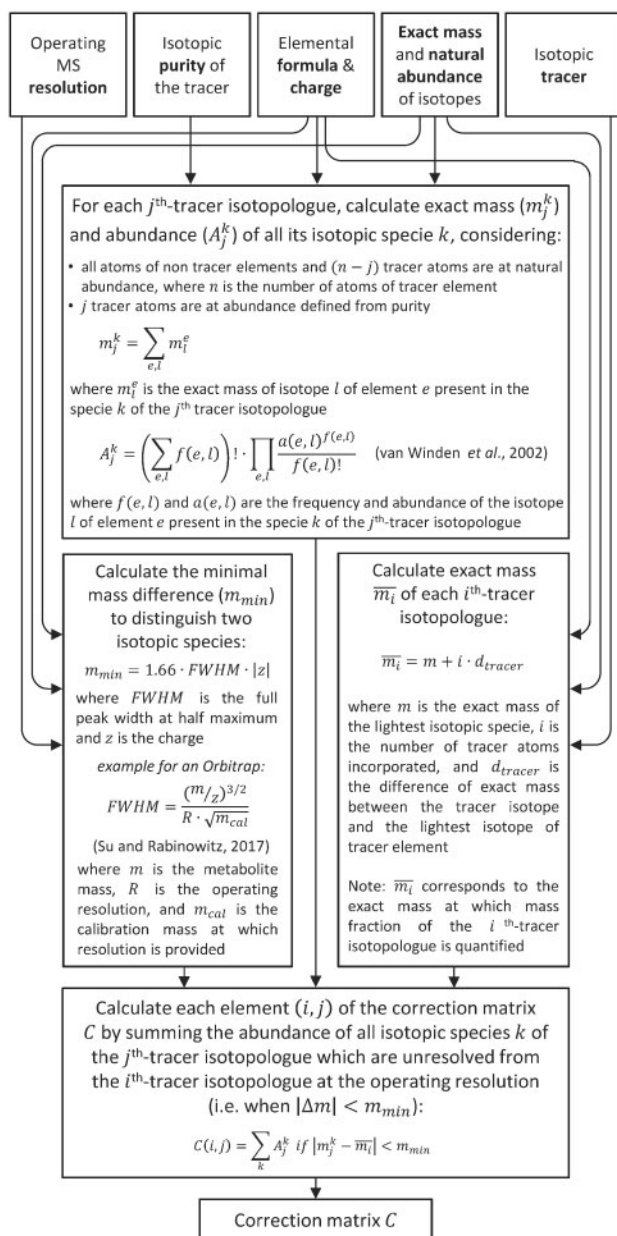


Fig. 2. Algorithm implemented in IsoCor v2 to construct accurate correction matrices at any MS resolution (option 'High-resolution')

Finally, Equation (1) is solved using the L-BFGS-B algorithm (Byrd et al., 1995), with the lower boundary of corrected mass fractions constrained to 0 to avoid negative values. The cost function is defined as the sum of the squared weighted errors, and the stopping criterion is fixed to 10^{-10} . The isotopologue distribution is obtained by normalizing the vector of corrected mass fractions to 1.

3 Results

As an illustrative example, we used IsoCor and AccuCor to correct 1880 ^{15}N -mass fractions (94 mass fractions from 22 metabolites in 20 samples) quantified on an Orbitrap at a resolution of 140 000 (dataset provided as example with AccuCor). Differences between the two correction methods are shown in Figure 3.

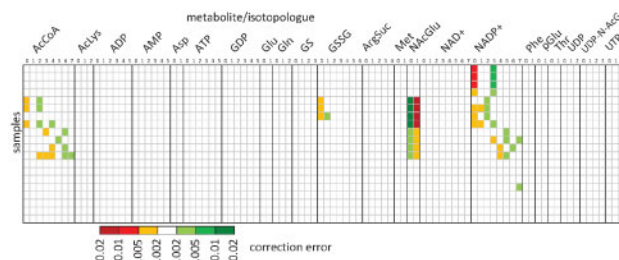


Fig. 3. Comparison of AccuCor and IsoCor correction results. Differences between ^{15}N -isotopologue distributions calculated by AccuCor and IsoCor from 94 mass fractions of 22 metabolites (columns) in 20 samples (rows)

The improvement brought by IsoCor v2 is highly dependent on the metabolite and on the sample, consistently with theoretical considerations. AccuCor is accurate in most of the tested cases (with errors < 0.002), but the correction error increased up to 0.014—i.e. similar to the accuracy and precision that can be reached on an Orbitrap (Heuillet et al., 2018; Su et al., 2017)—for four metabolites (acetylCoA, glutathione disulfide, N-acetyl-glutamate and NADP⁺). Even low yet systematic isotopic biases may lead to significant flux errors or biased interpretations (e.g. when evaluating the goodness of fit), as previously reported (Antoniewicz et al., 2006; Mairinger et al., 2018; Su et al., 2017). Importantly, the discrepancy observed in this illustrative example should not be considered as a higher bound since it is directly related to the isotopic tracer, metabolite, or resolution. The algorithm improvement made in IsoCor ensures to prevent such biases in all situations.

4 Conclusion

We present a major update of IsoCor to process low- and high-resolution mass spectrometry data. IsoCor v2 allows correction to be applied to any chemical species and can be used to exploit any isotopic tracer at any resolution. IsoCor v2 includes a graphical user interface for intuitive use and a command-line interface to streamline its integration into existing pipelines. Moreover, IsoCor v2 implements a novel algorithm which ensures accurate correction of HRMS data compared to state-of-the-art tools. We discussed current correction methods and showed that they fail to fully account for the additive nature of isotopic mass defects when constructing the correction matrix. The resulting biases depend on each situation (in terms of metabolite, isotopic tracer, MS resolution and labeling pattern) but should be prevented to fully exploit the accuracy of modern mass spectrometers. Such biases may turn out to be critical for quantitative downstream applications such as ^{13}C -fluxomics. With this contribution, we aim at strengthening the reliability of isotopic data and thereby the biological value of isotope labeling experiments.

Acknowledgement

MetaboHub (ANR-11-INBS-0010, metabohub.fr) is gratefully acknowledged.

Funding

This work was supported by the French National Research Agency (project ENZINVIVO, ANR-16-CE11-0022) and by the European Union (ERA-CoBioTech's project CIPRO, ANR-17-COBI-0003-05).

Conflict of Interest: none declared.

References

- Antoniewicz, M.R. *et al.* (2006) Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.*, **8**, 324–337.
- Byrd, R.H. *et al.* (1995) A limited memory algorithm for bound constrained optimization. *Siam J. Sci. Comput.*, **16**, 1190–1208.
- Carreer, W.J. *et al.* (2013) A computational framework for high-throughput isotopic natural abundance correction of omics-level ultra-high resolution FT-MS datasets. *Metabolites*, **3**, 853–866.
- Heinrich, P. *et al.* (2018) Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS- and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrector. *Sci. Rep.*, **8**, 17910.
- Heuillet, M. *et al.* (2018) Methodology for the validation of isotopic analyses by mass spectrometry in stable-isotope labeling experiments. *Anal. Chem.*, **90**, 1852–1860.
- Mairinger, T. *et al.* (2018) Comprehensive assessment of measurement uncertainty in ^{13}C -based metabolic flux experiments. *Anal. Bioanal. Chem.*, **410**, 3337–3348.
- Midani, F.S. *et al.* (2017) The importance of accurately correcting for the natural abundance of stable isotopes. *Anal. Biochem.*, **520**, 27–43.
- Millard, P. *et al.* (2012) IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics*, **28**, 1294–1296.
- Su, X. *et al.* (2017) Metabolite spectral accuracy on orbitraps. *Anal. Chem.*, **89**, 5940–5948.
- Thurman, E.M. and Ferrer, I. (2010) The isotopic mass defect: a tool for limiting molecular formulas by accurate mass. *Anal. Bioanal. Chem.*, **397**, 2807–2816.
- van Winden, W.A. *et al.* (2002) Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol. Bioeng.*, **80**, 477–479.