bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESEARCH

Probabilistic quotient's work & pharmacokinetics' contribution: countering size effect in metabolic time series measurements

Mathias Gotsmy [1†], Julia Brunmair [1†], Christoph Büschl [1], Christopher Gerner [1,2] and Jürgen Zanghellini [1*]

*Correspondence: juergen.zanghellini@univie.ac.at
[1] Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria
Full list of author information is available at the end of the article
†Equal contributor

Abstract

Metabolomic time course analyses of biofluids are highly relevant for clinical diagnostics. However, some sampling methods suffer from unknown sample sizes commonly known as size effects, which prevents absolute quantification of biomarkers. Recently, studies have developed mathematical post acquisition normalization methods to overcome these problems either by exploiting already known pharmacokinetic information or with statistics.

Here we present an improved normalization method, MIX, that unifies the advantages of both approaches. It combines two normalization terms, one representing pharmacokinetic normalization (PKM) and one representing a popular statistical approach (PQN).

To test the performance of MIX, we generated synthetic data closely resembling real finger sweat metabolome measurements. MIX was subsequently used for normalization of the synthetic data and we were able to show that it overcomes weaknesses of the two normalization strategies applied separately. Moreover, we validate our results by using real finger sweat metabolome data from literature. There, we were able to demonstrate that MIX is more robust than the normalization strategy originally used for the data set.

In conclusion, the MIX method improves the reliability and robustness of biomarker measurements in finger sweat and other biofluids. Moreover, it has potential to pave the way for quantitative biomarker discovery and hypothesis generation from metabolic time course data.

Keywords: Metabolomics; Finger Sweat; PKM; PQN

1 Introduction

In recent years, the analysis of the sweat metabolome has recieved increased attention from several fields of study [1, 2, 3]. For example, sweat has been in the focus of forensic scientists since it is possible to analyze metabolomic profiles of finger prints that have been found (e.g. at a crime scene) [4]. Also, drug testing can easily be performed on sweat samples. One advantage of this method is to not only identify already illegal substances but their metabolic degradation products as well, thereby allowing to distinguish between drug consumption and mere contact [1]. Another application of sweat metabolomics is in diagnostics for personalized medicine, where the focus is put on discerning metabolic states of the body and trying to optimize nutrition and treatment based upon information of biomarkers in sweat [5, 6, 7].

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sweat metabolomics offers several technical advantages. Firstly, sweat is a rich source of biomolecules and thus offers great potential for biomarker discovery [8, 9]. Secondly, sweat sampling is easy compared to sampling of other biofluids (e.g. blood or urine). Moreover, it is non-invasive and can in principle be rapidly repeated.

Several sampling methods have been developed [2, 3, 9, 10]. However, most of them work in a very similar manner: a water absorbing material is put onto the skin's surface to collect sweat for some (short) time. Sweat metabolites are subsequently extracted from this material and analyzed [3, 10]. Methods differ, however, in if and how they induce sweating. Some methods induce increased sweating by physical exercise [9] or chemical stimulation [2], whereas in other studies no sweat induction is performed and the natural sweat rate is sufficient for metabolomic analysis [3, 11].

Regardless of the exact sampling method, most of the above mentioned studies suffer from one major drawback. The sweat flux is highly variable, depending not only on interindividual differences, but also on body location, temperature, humidity, exercise and further factors that may change multiple times over the course of one day [12]. For example, even with conservative estimates a variability of sweat flux q_{sweat} on the finger tips between 0.05 and $0.62\,\text{mg}\,\text{cm}^{-2}\,\text{min}^{-1}$ needs to be accounted for [13, 14, 15]. This is a major challenge for comparative or quantitative studies, which has been acknowledged by many, e.g. [1, 4, 8, 16, 17, 18], however only actively approached by few – most notably [9].

The key problem is associated to the fact that often one is interested in the true metabolite concentrations, $\mathbf{C} \in \mathbb{R}^{n_{\text{metabolites}}}$, of $n_{\text{metabolites}}$ metabolites, which is obscured by an unknown and time-dependent sweat flux. Thus, the measured metabolites' intensities are not proportional to $\mathbf{C}$ but to the metabolite mass vector, $\widetilde{\mathbf{M}} \in \mathbb{R}^{n_{\text{metabolites}}}$,

$$\widetilde{\mathbf{M}}(t) = a_{\text{sample}} \int_{t-\tau}^{t} \mathbf{C}(t')\; q_{\text{sweat}}(t')\; \mathrm{d}t'. \tag{1}$$

Here a_{sample} and τ denote the surface area of skin that is sampled, and the time it takes to collect one sample, respectively. We assume that $\mathbf{C}$ changes little over the integration time τ. Thus (1) simplifies to

$$\widetilde{\mathbf{M}}(t) \approx \mathbf{C}(t)\; V_{\text{sweat}}(t), \tag{2a}$$

with an unknown sweat volume during sampling

$$V_{\text{sweat}}(t) := a_{\text{sample}} \int_{t-\tau}^{t} q_{\text{sweat}}(t')\; \mathrm{d}t', \tag{2b}$$

and the problems reads: given $\widetilde{\mathbf{M}}$, how can we compute $\mathbf{C}$ if we don't know V_{sweat}?

We emphasize that throughout the manuscript the mass of a metabolite is defined as the measured abundance of the metabolite in a measured sample, and neither as the molar mass or mass to charge ratio. Moreover, we acknowledge that without a calibration curve the measured abundances have a relative peak-area unit and are thus strictly neither absolute masses nor concentrations. The proportionality constant that scales measured intensities to mass units is determined by the calibration

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

curve. The proper calibration curve is not further discussed here but assumed to be available when applicable.

The need to calculate absolute metabolite concentrations from small biological samples of unknown volume is not unique to sweat metabolomics, but known throughout untargeted metabolomics. The problem is commonly referred to as size effects [19].

Three strategies have been developed to tackle it:

Direct Sweat Volume Measurement. Measuring V_{sweat}, for instance via microfluidics [9, 20, 21], is the most straight forward method to solve (2) and typically very accurate with minimally required volumes in the rang of ~ 5 to $100\,\mu\text{L}$ [9, 20, 21]. However, in case of sweat sampling it may take quite some time, large sample areas or increased (i.e. induced) sweating to collect enough sweat for robust volume quantification. Another alternative is the volume estimation via paired standards [22], however, such method increases complexity of analysis. Either option would impede fast and easy sample collection and analysis.

Indirect Sweat Volume Computation. If the chemical kinetics of targeted metabolite concentrations are known, then kinetic parameters and the sweat volume at each time point can be simultaneously determined by fitting the measured mass vector to (2). Recently, we used this strategy to computationally resolve not only sample volumes in the nL to single digit µL-range but also accurately quantify personalized metabolic response patterns upon caffeine ingestion [23]. Albeit feasible for determination of individual differences with knowledge of reaction kinetics, this method quickly becomes unconstrained when too little prior information is available. Therefore, it is not suited for the discovery of unknown reaction kinetics. Moreover, this method requires several sampling time points to allow modeling the kinetics of different metabolites thereby decreasing simplicity of sampling.

Statistical Normalization. With this approach the aim is to normalize the mass vector by the apparent mass of a marker that scales proportionally to the sample volume, so that the ratio becomes (at least approximately) independent of the sample volume. Various strategies have been developed for untargeted metabolomics; for example, normalization by total measured signal [24], and singular value decomposition-based normalization [25]. However, one of the best performing methods–referred to as probabilistic quotient normalization (PQN) – simply assumes that the median of the ratio of two apparent mass vectors is proportional to the sample volume [19, 26, 27, 28]. Although PQN does not allow one to compute sample volumes *per se*, it enables one to assess differential changes [26].

In this manuscript we explore the performance of combining targeted and untargeted metabolomics normalization strategies in a MIX normalization model. We show that MIX can significantly outperform other indirect sweat volume estimation methods. To validate the results we characterize caffeine metabolization in the human body analyzed in the finger sweat.

2 Theory

2.1 Probabilistic Quotient Normalization

Definition Probabilistic quotient normalization (PQN) assumes that for a large, untargeted set of metabolites the median metabolite concentration fold change between two samples (e.g. two measured time points, t_r and t_s) is approximately 1,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$Q^{\mathbf{C}} = \text{median}\left\{\frac{C_j(t_r)}{C_j(t_s)}\right\} \approx 1, \quad j \in [1, n_{\text{metabolites}}]. \tag{3a}$$

Consequently, fold changes calculated from $\widetilde{\mathbf{M}}$ instead of $\mathbf{C}$ are proportional to the ratio of V_{sweat},

$$Q^{\mathbf{M}} = Q^{\mathbf{C}}\ \frac{V_{\text{sweat}}(t_r)}{V_{\text{sweat}}(t_s)} \approx \frac{V_{\text{sweat}}(t_r)}{V_{\text{sweat}}(t_s)} \tag{3b}$$

with

$$Q^{\mathbf{M}} = \text{median}\left\{\frac{M_j(t_r)}{M_j(t_s)}\right\}, \quad j \in [1, n_{\text{metabolites}}]. \tag{3c}$$

In order to minimize the influence of experimental errors

$$M_j^{\text{ref}} = \text{median}\left\{\widetilde{M}_j(t_i)\right\}, \quad i \in [1, n_{\text{time points}}] \tag{4}$$

often replaces the dedicated sample in $M_i(t_s)$ in the denominator of (3c) [26]. Therefore, the normalization quotient by PQN is calculated as

$$Q^{\text{PQN}}(t) = \text{median}\left\{\frac{\widetilde{M}_1(t)}{M_1^{\text{ref}}}\right\}, \quad j \in [1, n_{\text{metabolites}}]. \tag{5}$$

Q^{PQN} is a relative measure and distributes around 1. In analogy to (3b), we define its relation to the sweat volume $V_{\text{sweat}}^{\text{PQN}}$ as

$$Q^{\text{PQN}}(t) = \frac{V_{\text{sweat}}^{\text{PQN}}(t)}{V_{\text{sweat}}^{\text{ref}}}, \tag{6}$$

where $V_{\text{sweat}}^{\text{ref}}$ denotes some unknown, time-independent reference (sweat) volume. Note that with real data only $Q^{\text{PQN}}(t)$ values can be calculated, but $V_{\text{sweat}}^{\text{PQN}}(t)$ as well as $V_{\text{sweat}}^{\text{ref}}$ remain unknown.

Discussion The biggest advantage of this method is that no calibration curves and prior knowledge about changes over time of measured metabolites is required. Moreover, PQN is independent from the number of sample points measured in a time series. However, its major drawback is that the normalization quotient is not an absolute quantification and only shows relative changes. I.e. it does not quantify V_{sweat} as given in Eq. 2 directly with an absolute value, but instead normalizes relative abundances between samples and time points.

2.2 Pharmacokinetic Normalization

Definition In the pharmacokinetic model (PKM) we assume that we know at least the functional dependence, i.e. the pharmacokinetics, but not necessarily the value

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the k (pharmaco-)kinetic parameters $\boldsymbol{\theta} \in \mathbb{R}^k$ for $2 \leq \ell \leq n_{\text{metabolites}}$ metabolites. Without loss of generality we (re-)sort $\widetilde{\mathbf{M}}$ such that the first ℓ elements (collected in the vector $\widetilde{\mathbf{M}}_\ell$) correspond to metabolites with known pharmacokinetic dependence, while the remaining $n_{\text{metabolites}} - \ell$ elements (collected in the vector $\widetilde{\mathbf{M}}_{\ell+}$) correspond to metabolites with unknown kinetics. Then (2) takes the form of

$$\begin{pmatrix} \widetilde{\mathbf{M}}_\ell \quad (t) \\ \widetilde{\mathbf{M}}_{\ell+}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{C}_\ell \quad (t; \boldsymbol{\theta}) \\ \mathbf{C}_{\ell+}(t) \end{pmatrix} V_{\text{sweat}}^{\text{PKM}}(t) \tag{7a}$$

with physically meaningful bounds;

$$V_{\text{lower bound}} \leq V_{\text{sweat}}^{\text{PKM}}(t) \leq V_{\text{upper bound}}, \tag{7b}$$

$$\boldsymbol{\theta}_{\text{lower bound}} \leq \quad \boldsymbol{\theta} \quad \leq \boldsymbol{\theta}_{\text{upper bound}}. \tag{7c}$$

$V_{\text{sweat}}^{\text{PKM}}(t)$ as well as $\boldsymbol{\theta}$ can be obtained by parametric fitting of $\widetilde{\mathbf{M}}_\ell^{\text{PKM}}(t)$. Note that this allows not only to compute absolute values of $\mathbf{C}_\ell^{\text{PKM}}(t; \boldsymbol{\theta})$ but – with $V_{\text{sweat}}^{\text{PKM}}(t)$ – also of all other concentrations via $\mathbf{C}_{\ell+}(t) = \widetilde{\mathbf{M}}_{\ell+}(t)/V_{\text{sweat}}^{\text{PKM}}(t)$.

As $V_{\text{sweat}}^{\text{PKM}}(t_i)$ may be different at every time step t_i, we need to know the (pharmaco-)kinetics of at least two metabolites, otherwise the number of parameters is larger than the number of data points.

Discussion The biggest advantage of this method is that is can implicitly estimate absolute values of V_{sweat} without the need of direct measurements. Therefore, sweat volumes can become smaller than the minimum required in volumetric methods and shorter sampling times also become possible. A drawback of this method is the fact that it is only feasible if one has prior knowledge on relevant pharmacological parameters (i.e. ingested dose of metabolites of interest, volume of distribution, body mass of specimen, range of expected kinetic constants), which is limiting the approach to studies where at least two metabolites together with their pharmacokinetics are well known . Moreover, calibration curves of metabolites of interest, and sufficiently many samples in a time series are required for robustly fitting the equation system. In a previously performed sensitivity analysis, an increase in the quality of fit was observed as the number of samples increased from 15 to 20 time points per measured time series [23].

2.3 Mixed Normalization

Definition The mixed normalization model (MIX) is a combination of PQN and PKM. It is designed to incorporate robust statistics of untargeted metabolomics via its PQN term as well as an absolute estimation of V_{sweat} via its PKM term.

Optimal parameters of MIX are found via optimization of two equations,

$$\begin{pmatrix} \widetilde{\mathbf{M}}_\ell \quad (t) \\ \widetilde{\mathbf{M}}_{\ell+}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{C}_\ell \quad (t; \boldsymbol{\theta}) \\ \mathbf{C}_{\ell+}(t) \end{pmatrix} V_{\text{sweat}}^{\text{MIX}}(t) \tag{8a}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and

$$Q^{\mathrm{PQN}}(t) = \frac{V^{\mathrm{MIX}}_{\mathrm{sweat}}(t)}{V^{\mathrm{ref}}_{\mathrm{sweat}}} \tag{8b}$$

where $V^{\mathrm{ref}}_{\mathrm{sweat}}$ is an additional, time-independent parameter of the model that scales the absolute values of $V^{\mathrm{MIX}}_{\mathrm{sweat}}$ to the relative abundances of Q^{PQN}. $V^{\mathrm{MIX}}_{\mathrm{sweat}}$ and $\boldsymbol{\theta}$ are constrained between physically meaningful bounds,

$$V_{\text{lower bound}} \leq V^{\mathrm{MIX}}_{\mathrm{sweat}}(t) \leq V_{\text{upper bound}}, \tag{8c}$$

$$\boldsymbol{\theta}_{\text{lower bound}} \leq \quad \boldsymbol{\theta} \quad \leq \boldsymbol{\theta}_{\text{upper bound}}. \tag{8d}$$

Bounds for V_{sweat} are calculated by Equation 2a and minimal and maximal sweat rates from literature. The weighting of error residuals of both equations (8a, 8b) is implemented over a hyperparameter, λ, as explained in Section 3.1.2.

Discussion We hypothesize that MIX model can combine the advantages of PQN and PKM normalization models. Firstly, in many metabolomics studies a combination of targeted and untargeted measurements is used and thus the data is already available resulting in less to little additional effort. Moreover, we believe that MIX inherits the statistical robustness of PQN while simultaneously estimating absolute values as fitted by PKM.

3 Methods

3.1 Implementation

A generalized version of PKM and MIX was implemented as a Python class. As input it requires the number of metabolites measured ($n_{\mathrm{metabolites}}$), a vector of time points as well as the measured mass data ($\widetilde{\mathbf{M}}(t)$, matrix with samples in the rows and metabolites in the columns). MIX additionally takes a $\mathbf{Q}^{\mathrm{PQN}}$ vector (calculated with the PQN method from untargeted data) for all time points of a time series. Upon optimization (carried out with `self.optimize_monte_carlo`, which is a wrapper for SciPy's `optimize.curve_fit` [29]) the kinetic constants and sweat volume parameters are optimized to the measured data by minimizing the functions listed in Equation 9b and 9a for PKM and MIX respectively:

$$\mathcal{L}^{\mathrm{MIX}} = \mathcal{L}^{\mathrm{PKM}} + \mathcal{L}^{\mathrm{PQN}} \tag{9a}$$

where

$$\mathcal{L}^{\mathrm{PKM}} = \sum_{i=1}^{n_{\text{time points}}} \sum_{j=1}^{n_{\mathrm{metabolites}}} L\left[\lambda\left(\widetilde{M}_{ij} - C_{ij}\ V^{\mathrm{MIX}}_{\text{sweat }i}\right)^2\right] \tag{9b}$$

$$\mathcal{L}^{\mathrm{PQN}} = \sum_{i=1}^{n_{\text{time points}}} L\left[(1-\lambda)\left(V^{\mathrm{MIX}}_{\text{sweat }i}/V^{\mathrm{ref}}_{\mathrm{sweat}} - Q^{\mathrm{PQN}}_{\text{sweat }i}\right)^2\right] \tag{9c}$$

and L is the loss function. The key difference between PKM and MIX is that the fitted V_{sweat} in MIX are biased towards relative abundances as calculated by PQN.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

An important additional hyperparameter of the MIX model is λ, which weights the error residuals of $\mathcal{L}^{\mathrm{PKM}}$ and $\mathcal{L}^{\mathrm{PQN}}$. Its calculation is discussed in Section 3.1.2. If $\lambda = 1$, the MIX model simplifies again to a pure PKM model. If $\lambda = 0$ then the MIX model simplifies to the PQN within the constraints defined in Equation 8c.

To summarize, an overview of the differences of PKM and MIX model is given in Supplementary Table 1.

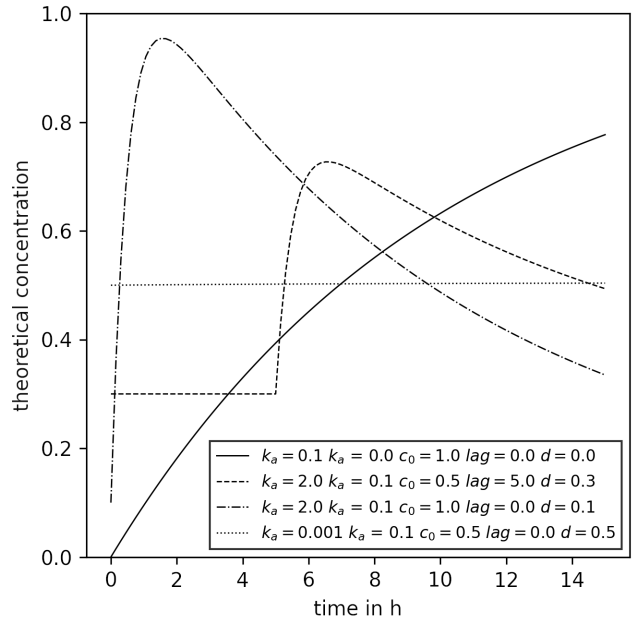

Figure 1 Examples of concentration time series that can be modeled with the modified Bateman equation used. The legend shows the kinetic parameters used to create the respective curves. All parameters are within the bounds that were used for kinetic parameter fitting.

3.1.1 Hyperparameters

Several hyperparameters can be set for the PKM and MIX Python classes.

Firstly, it is possible to choose the kinetic function used to calculate $\mathbf{C}$. In this study we focused on a modified Bateman function ($F(t)$) with 5 kinetic parameters ($k_a,\ k_e,\ c_0,\ lag,\ d$):

$$F(t) = \begin{cases} b(t) + d & \text{if } b(t) \geq 0 \\ d & \text{if } b(t) < 0 \end{cases} \tag{10}$$

with

$$b(t) = c_0 \, \frac{k_a}{k_e - k_a} \left(e^{-k_a(t-lag)} - e^{-k_e(t-lag)} \right). \tag{11}$$

This function was designed to be flexible and able to represent several different metabolite consumption and production kinetics as exemplified by Figure 1. Intu-

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

itively, k_a and k_e correspond to kinetic constants of absorption and elimination of a metabolite of interest with the unit h^{-1}. c_0 is the total amount of a metabolite absorbed over the volume of distribution with the unit $\mathrm{mol\,L}^{-1}$. The *lag* term with the unit h shifts the function along the X-axis, intuitively defining the starting time point of absorption of a metabolite of interest, whereas the d term with the unit $\mathrm{mol\,L}^{-1}$ shifts the function along the Y-axis.

Secondly, the loss function can be set to either any loss function implemented by SciPy's `curve_fit` [29] or one of two separately implemented functions (`max_linear_loss`, `max_cauchy_loss`). The reason for the addition of custom loss functions was that a good performance can be achieved when the relative and absolute error residuals were calculated and their maximum was used for loss estimation (thus the word `max` is expressed in their names) [23]. In this study we used the `max_cauchy_loss` loss function, which first calculates said maximum and then estimates its loss with the Cauchy loss function implemented according to SciPy [29].

Thirdly, a weighting constant for every measured data point can be used by the model (via the function `self.set_sigma`) and is required for its optimization. The optimization of both, PKM and MIX models, is done with a Monte Carlo strategy where the initial parameter guesses are randomly sampled from an uniform distribution between their bounds. Performing a sensitivity analysis, we previously showed that this method is preferable to a single fitting procedure [23]. In this study the number of Monte Carlo replicates for model fitting was set to 100.

3.1.2 Weighting of MIX Loss Terms

Different arguments for the extend of λ corresponding to more or less importance of the PKM or PQN in the parameter optimization can be used. In the following paragraphs we present two diametrical approaches to the estimation of λ (metabolite-dependent weighting and equal weighting), which span the space of reasonable solutions to the problem. In reality, we expect the optimal size of λ to be between these extreme points.

Metabolite-Depending Weighting On one hand, one could argue that the loss terms should be weighted twofold, firstly, according to the number of metabolites used for their calculation respectively and, secondly, the number of data points in each loss term (Supplementary Equations 15). The solution for λ in such a case is given in Equation 12. This solution is the most reasonable if the confidence of measured data is exactly the same for all metabolites.

$$\lambda = \frac{1}{n_{\text{metabolites}}^{\text{PQN}} + 1} \tag{12}$$

Equal Weighting On the other hand, however, one can argue that targeted metabolites have higher confidence (i.e. smaller experimental error) than the average metabolite measured in an untargeted fashion. In the most extreme case one would propose a method where the loss terms are just weighted by the number of data

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

points, but not by the number of metabolites used in the calculation of each term (Supplementary Equations 16). For such a method the solution for λ is given by Equation 13.

$$\lambda = \frac{1}{n_{\text{metabolites}}^{\text{PKM}} + 1} \tag{13}$$

3.1.3 Full and Minimal Models

In this study we differentiate between full and minimal models. With full models we refer to pharmacokinetic normalization models (PKM or MIX) where all metabolites of a given data set are used for the pharmacokinetic normalization. This means that, for example, if $n_{\text{metabolites}} = 20$ all 20 metabolites were modeled with the modified Bateman function and thus in Equations (7a) and (8a) $\ell = n_{\text{metabolites}}$ and $\widetilde{\mathbf{M}}_{\ell+}$ is an empty vector. On the other hand, minimal models are models where only the few, known, better constrained metabolites were modeled with a kinetic function. This means that the information used for $\text{PKM}_{\text{minimal}}$ does not change upon addition of synthetic metabolites. Therefore, its goodness of fit measure should stay constant within statistical variability upon change of $n_{\text{metabolites}}$. This behaviour was used to verify if the simulations worked as intended and no biases in the random number generation exist. On the other hand $\text{MIX}_{\text{minimal}}$ model still gained information from the increase of $n_{\text{metabolites}}$ as the PQN part of this model was calculated with all $n_{\text{metabolites}}$. Therefore, changes in the goodness of fit measures for $\text{MIX}_{\text{minimal}}$ are expected. We emphasize that the definition of full and minimal models is specific to this particular study. Here we explicitly set $\ell = 4$, which originates from previous work where 4 targeted metabolites (caffeine, paraxanthine, theobromine, theophylline) with known kinetics were measured [23].

3.2 Synthetic Data Creation

Three different types of synthetic data were investigated. In all three cases data creation started with a simple toy model closely resembling the concentration time series of caffeine and its degradation products (paraxanthine, theobromine, and theophylline) in the finger sweat as described elsewhere [23]. The respective parameters are listed in Supplementary Table 2. With them the concentration of metabolites #1 to #4 were calculated for 20 time points (between 0 and 15 h in equidistant intervals, Figure 2). Subsequently, new synthetic metabolite concentration time series were sampled and appended to the toy model (i.e. to the concentration vector, $\mathbf{C}(t)$). Three different synthetic data sampling strategies were tested and their specific details are explained in the following sections. Next, sweat volumes (V_{sweat}) were sampled from a log-normal distribution truncated at ($0.05 \leq V_{\text{sweat}} \leq 4\,\mu\text{L}$) closely resembling the distribution of sweat volumes estimated in our previous publication [23], Supplementary Figure S2. Finally, an experimental error ($\boldsymbol{\epsilon}$) was sampled for every metabolite and time point from a normal distribution with a coefficient of variation of 20% and the synthetic data was calculated as

$$\widetilde{\mathbf{M}}(t) = \text{diag}(\mathbf{C}(t))\ V_{\text{sweat}}(t)\ \boldsymbol{\epsilon}(t). \tag{14}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

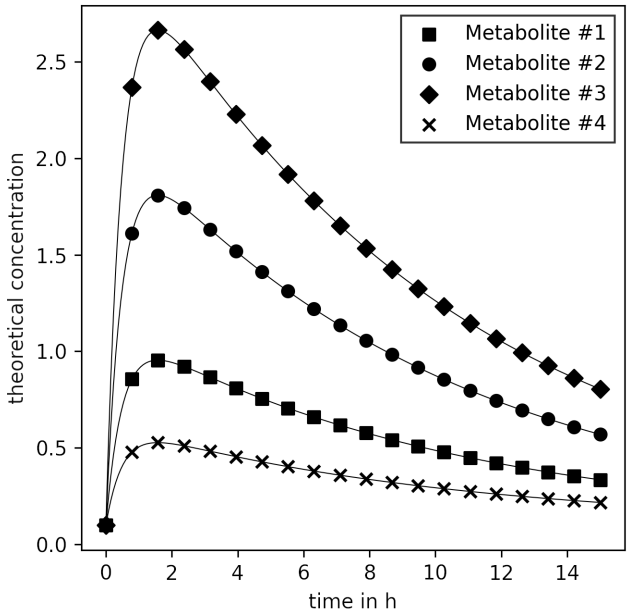

Figure 2 C for the first four metabolites of the synthetic data. Kinetic parameters used for calculation are listed in Supplementary Table 2.

For every tested condition 100 synthetic data replicates were generated and the normalization models were fitted.

3.2.1 Sampled Kinetics

In simulation v1, data was generated by sampling kinetic parameters for new metabolites from an uniform distribution. The distribution was constrained by the same bounds also used for the PKM and MIX model fitting: $(0,0,0,0)^{\mathrm{T}} \leq (k_a, k_e, c_0, lag, d)^{\mathrm{T}} \leq (3,3,5,15,3)^{\mathrm{T}}$. Subsequently the concentration time series of the synthetic metabolites were calculated according to the modified Bateman function (Equation 10).

3.2.2 Sampled Mean and Standard Deviation

Means and standard deviations of the concentration time series of metabolites were calculated from untargeted real finger sweat data (for details see Section 3.4). The PDF of both can be described by a log-normal distribution (Supplementary Figure S5). For the data generation of simulation v2, per added metabolite one mean and one standard deviation were sampled from the fitted distribution and used as an input for another log-normal distribution from which a random concentration time series was subsequently sampled. This results in synthetic concentration values that behave randomly and, therefore, cannot be easily described by our pharmacokinetic models.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.2.3 Sampled from Real Data

To get an even better approximation to real data, in simulation v3 concentration time series were directly sampled from untargeted real finger sweat data (for details see Section 3.4). To do so, the untargeted metabolite $\widetilde{\mathbf{M}}$ time series data set was normalized with PQN. As the number of metabolites in this data set was comparably large ($n_{\text{metabolites}} = 3446$) we could assume that the relative error (or rSD, for more explanation see Section 4.1) was negligibly small. Therefore, the PQ normalized data set could be used as ground truth for concentration time series sampling. Subsequently, a subset of the original ground truth data was sampled for synthetic data generation.

3.3 Normalization Model Optimization

Normalizing for the sweat volume by fitting kinetics through the measured values only has a clear advantage over PQN if it is possible to infer absolute sweat volumes and concentration data. In order to be able to do that, some information about the kinetics and the starting concentrations of metabolites of interest need to be known. For example, when modeling the caffeine network in our previous publication [23] we knew that the *lag* parameter of all metabolites was 0 and that the total amount of caffeine ingested (which corresponds to c_0) was 200 mg. Moreover, we knew that caffeine and its metabolites are not synthesized by humans and implemented the same strategy into our toy model (corresponding to d). As the toy model was designed to resemble such a metabolism we translated these information to the current study. Therefore, we assumed that the first 4 metabolites in our toy model had known c_0, lag, and d parameters. For their corresponding k_a and k_e and the parameters of all other metabolites the bounds were set to the same $(0,0,0,0)^{\mathrm{T}} \leq (k_a, k_e, c_0, lag, d)^{\mathrm{T}} \leq (3,3,5,15,3)^{\mathrm{T}}$ used in kinetic data generation. Figure 1 shows examples of concentration time series that can be described with the modified Bateman function and parameters within the fitting bounds.

3.4 Real Finger Sweat Metabolome Data

The real world finger sweat data was extracted from Study C of ref. [23]. It was downloaded from MetaboLights (MTBLS2772 and MTBLS2776).

Preprocessing The metabolome data set was split into two parts: targeted and untargeted. The targeted data (i.e. the mass time series data for caffeine, paraxanthine, theobromine, and theophylline) was directly adopted from the mathematical model developed by [30]. This data is available on GitHub (`https://github.com/Gotsmy/finger_sweat`).

For the untargeted metabolomics part, the raw data was converted to the mzML format with the msConvert tool of ProteoWizard (version 3.0.19228-a2fc6eda4)[31]. Subsequently, the untargeted detection of metabolites and compounds in the samples was carried out with MS-DIAL (version 4.70)[32]. A manual retention time correction was first applied with several compounds present in the majority (more than 90%) of the samples. These compounds were single chromatographic peaks with no isomeric compounds present at earlier or later retention times (m/z 697.755 at 5.57 min, m/z 564.359 at 5.10 min, m/z 520.33 at 4.85 min, m/z 476.307 at 4.58

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

min, m/z 415.253 at 4.28 min, m/z 371.227 at 3.95 min, m/z 327.201 at 3.56 min, m/z 283.175 at 3.13 min, m/z 239.149 at 2.633 min, m/z 166.144 at 1.69 min, m/z 159.113 at 1.19). After this, untargeted peak detection and automated alignment (after the manual alignment) were carried out with the following settings: Mass accuracy MS1 tolerance: 0.005 Da, Mass accuracy MS2 tolerance: 0.025 Da, Retention time begin: 0.5 min, Retention time end: 6 min, Execute retention time correction: yes, Minimum peak height: 1E5, Mass slice width: 0.01 Da, Smoothing method: Linear weighted moving average, Smoothing level: 3 scans, Minimum peak width: 5 scans, Alignment reference file: `C_D1_I_o_pos_ms1_1.mzML`, Retention time tolerance: 0.3 min, MS1 tolerance: 0.015 Da, Blank removal factor: 5 fold change). No blank-subtraction was carried out as the internal standard caffeine was spiked into each sample including the blanks. Peak abundances and meta-information were exported with the `Alignment results export` functionality.

Features with retention times after 5.5 min as well as features with sample abundances of $< 5 \times$ blank average (except for the internal standard, caffeine-D9) were excluded from the data set. Additionally, we excluded isomers within a m/z difference of less than 0.001 Da and a retention time difference of less than 0.5 min. Moreover, features that were present in less than 80% of finger sweat measurements were excluded. Data imputation for both targeted and untargeted was done with quantile regression imputation of left-censored data (QRILC) performed in R 4.0.2 using the `ImputeLCMD` package [33, 34].

Real Data Normalization In this finger sweat data set, time series of targeted as well as untargeted metabolomics are listed. The kinetics of the four targeted metabolites (caffeine, paraxanthine, theobromine, and theophylline) are known. A reaction network of the metabolites is shown in Supplementary Figure S3. Briefly, caffeine is first absorbed and then converted into three degradation metabolites. Additionally, all four metabolites are eliminated from the body. All kinetics can be described with first order mass action kinetics [35, 36].

In order to assess the performance of the sweat volume normalization methods the full network was split up into three subnetworks that all contained caffeine and one degradation metabolite each (Supplementary Figure S4). The solution of the first order differential equations describing such network is given in Supplementary Equations 17a and 17b. Moreover, the 3446 untargeted metabolite time series were randomly split up into three (almost) equally sized batches and each batch was assigned to one subnetwork. All three networks were subsequently separately normalized with $\text{PKM}_{\text{minimal}}$ and $\text{MIX}_{\text{minimal}}$ methods with kinetic parameters that were adjusted to the specific reaction network (Supplementary Figure S4). Subsequently, the kinetic constants (k'_1, k'_2, k'_3, k'_4) were estimated for 37 measured concentration time series. As all three subnetwork data sets originate from the same finger sweat masurements, the underlying kinetic constants should be exactly identical. As the kinetic constants of absorption ($= k'_1$) and elimination ($= k'_2 + k'_3$) of caffeine are estimated in all three subnetworks we used their standard deviation to test the robustness of the tested normalization methods.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

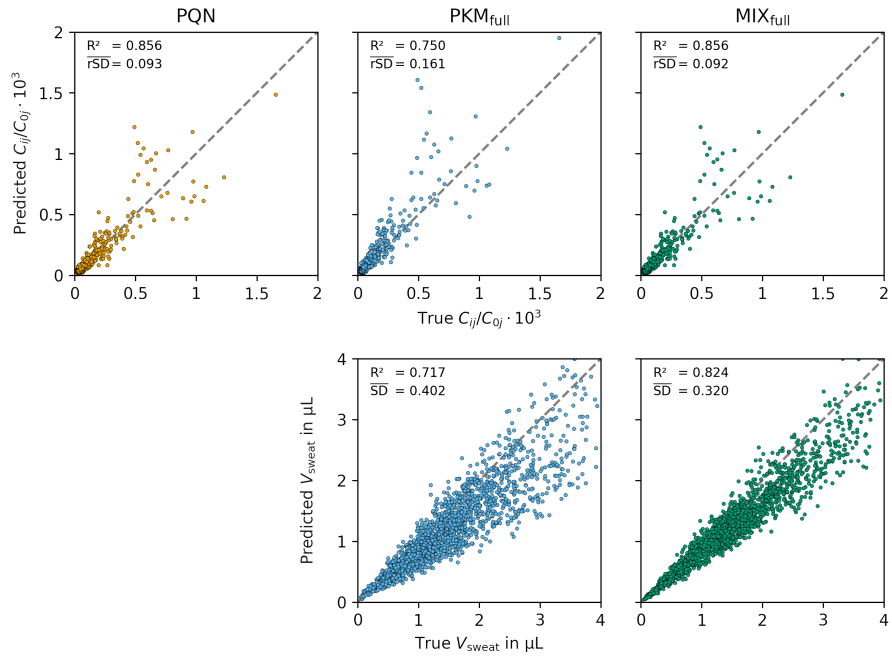

Figure 3 Relative and absolute normalization performance. In the top row the predicted C_{ij}/C_{0j} ($i \in \{1, ..., n_{\text{time points}}\}$, $j \in \{1, ..., n_{\text{metabolites}}\}$) are plotted as a function of the true, underlying C_{ij}/C_{0j}. The bottom row shows the predicted V_{sweat} as a function of the true, underlying V_{sweat}. The columns represent different normalization models (PQN, PKM_{full}, and MIX_{full} from left to right). As no absolute V_{sweat} can be calculated from PQN the bottom left plot is omitted. To illustrate the effect of different SD and rSD sizes (which both are calculated from V_{sweat}), we show their mean over 100 replicates in comparison to the R^2 values calculated from the points plotted. Intuitively rSD is a measure of good correlation on the top row whereas SD is a measured of good correlation on the bottom row (high R^2, low rSD/SD respectively).

3.5 Data Analysis

Goodness of Normalization Two goodness of fit measures are calculated to analyze the performance of the tested methods. SD is the standard deviation of the residuals of a sampled sweat volume time series vector ($\mathbf{V}^{\text{true}}_{\text{sweat}}$) minus the fitted sweat volume vector ($\mathbf{V}^{\text{fit}}_{\text{sweat}}$), while rSD is the standard deviation of the ratio of sampled and fitted $\mathbf{V}_{\text{sweat}}$ vectors normalized by its mean. Intuitively, SD is a measure of how much absolute difference there is between the fit and a true value, rSD on the other hand gives an estimate on how good the fitted sweat volumes are relative to each other. A visual depiction of SD and rSD is shown in Supplementary Figure S1 and their exact definition is given in the equations in 3.3.

Statistical Analysis The significant differences in the mean of goodness of fit measures were investigated by calculating p values with the non-parametric pairwise Wilcoxon signed-rank test [37] (SciPy's `stats.wilcoxon` function [29]). Significance levels are indicated by *, **, and *** for $p \leq 0.05$, 0.01, and 0.001 respectively.

4 Results

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

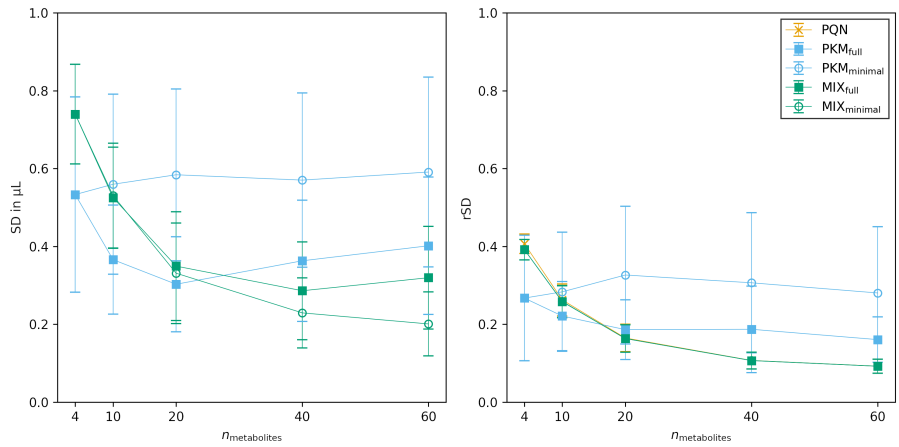

Figure 4 Goodness of normalization measures of simulation v3. The mean for 100 replicates for different sweat volume normalization models is given for SD (left panel) and rSD (right panel). The error bars represent standard deviations of the replicates. For the PQN method no SD was calculated.

4.1 Synthetic Data Simulations

In order to test the performance of different normalization models we generated 100 synthetic data sets with three different methods (simulations v1, v2, v3) and five different $n_{\text{metabolites}}$ (4, 10, 20, 40, 60) each, where the underlying $\mathbf{C}$, V_{sweat}, and $\boldsymbol{\epsilon}$ values were known. Simulations v1, v2, and v3 differ in the way how $\mathbf{C}$ was generated (kinetic, random, sampled from real data set, respectively). In order to quantify the the normalization model performance two measures of goodness of normalization were used for the analysis of the results: SD and rSD.

To visualize the obtained normalization performances we plotted the results for simulation v3 and $n_{\text{metabolites}} = 60$ in Figure 3 for three normalization models (from left to right column, PQN, PKM_{full}, and MIX_{full}). The top row shows the predicted $C_j(t_i;\boldsymbol{\theta})/C_j(0;\boldsymbol{\theta})$ (i.e. the concentration of each metabolite j at each time point i divided by its concentration at time 0) as a function of the true $C_j(t_i)/C_j(0)$ values. It illustrates the correlation of the relative abundances of one metabolite across all time points. Good correlations (i.e. high R^2) as seen for PQN and MIX_{full} result in a low rSD measure. On the bottom row of Figure 3 the absolute values of predicted V_{sweat} are plotted as a function of the true V_{sweat}. There it becomes evident that good correlations of absolute values result in low SD measures.

In the following sections we will focus on the size of SD and rSD respectively as they are both calculated from the predicted V_{sweat} directly. Note that for PQN no absolute V_{sweat} can be estimated and, therefore, no SD is calculated.

Dependence on the number of metabolites In a first step, we tracked SD and rSD of normalization methods for different numbers of metabolites ($n_{\text{metabolites}}$) to investigate how the methods behave with different amounts of available information. An overview of their goodness of normalization measures as a function of $n_{\text{metabolites}}$ on sampled kinetic data and completely random data is given in the Supplementary Figures S6 and S7 respectively. Most importantly, their performance of sampled subsets of real data is given in Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Models that fit a kinetic function through all possible metabolites (PKM_{full}, MIX_{full}, $\ell = 4$) perform well (low SD, low rSD) when the $\mathbf{C}$ data originates from a kinetic function (simulation v1). However, when the underlying data does not originate from kinetic time series their performance is reduced drastically. For PKM_{full} this is resembled in an increase of SD (from 0.19 ± 0.08 µL to 0.29 ± 0.13 µL for $n_{\text{metabolites}} = 60$) as well as of rSD (from 0.08 ± 0.02 to 0.11 ± 0.05 for $n_{\text{metabolites}} = 60$). Conversely, the rSD from MIX_{full} performs similarly well as the best methods on the completely random data (0.05 ± 0.01 for $n_{\text{metabolites}} = 60$). This property can be explained by the PQN term incorporated in the objective function of MIX_{full} which biases the results to better rSD values.

Another observation is the behaviour of PQN. Its rSD approaches a value close to 0 with increasing $n_{\text{metabolites}}$, indifferently on how the underlying data was generated (compare Figures 4, S6, S7). Conversely, no SD values can be calculated for PQN. The reason for this is that with PQN no absolute V_{sweat} values can be estimated.

Interestingly, the results from simulation v3 lie between the results from simulation v1 and v2. This gets especially evident when comparing the performance of MIX_{full} and PKM_{full} between Figure 4 and Supplementary Figures S6, S7. Such a result suggests that not all of the untargeted metabolites measured are completely random, but some can be described with the modified Bateman function. This leads to the hypothesis that after sweat volume normalization, the real finger sweat data (from which values for v3 were sampled) has high potential for the discovery of unknown kinetics.

Exact numbers for SD and rSD for all normalization methods and $n_{\text{metabolites}}$ are given in Supplementary Tables 3 and 4 respectively. Moreover, pairwise comparisons of SD and rSD of normalization methods relative to the results from $\text{PKM}_{\text{minimal}}$ are plotted in Supplementary Figures S10 and S11.

Statistical Testing As at $n_{\text{metabolites}} = 60$ the goodness of normalization measures start to flatten out we further investigated this condition for statistical significance. We used the non-parameteric Wilcoxon signed-rank test to compare pairwise differences in mean of SD and rSD between the tested models. p-values for all combinations are given in Supplementary Tables 5 and 6.

As Figure 4 already indicated, the overall best performance in SD as well as rSD is observed for the $\text{MIX}_{\text{minimal}}$ model. For $n_{\text{metabolites}} = 60$ it significantly outperforms every other method's SD (Figure 5). There is only one exception; in simulation v1, between $\text{MIX}_{\text{minimal}}$ and MIX_{full} no significant difference was found (Supplementary Figure S8). Moreover, $\text{MIX}_{\text{minimal}}$'s performance in rSD is at least equal to or better than all other tested methods (Supplementary Table 6). Compared to the previously used $\text{PKM}_{\text{minimal}}$ [23], the SD of $\text{MIX}_{\text{minimal}}$ improves by $-59 \pm 24\,\%$, the rSD by $-56 \pm 23\,\%$ (Supplementary Figures S10, S11).

We, therefore, conclude that normalizing the sweat volume by the $\text{MIX}_{\text{minimal}}$ method reduces the error for the estimated V_{sweat} compared to other tested methods. Compared to PKM, $\text{MIX}_{\text{minimal}}$ has the advantage that its performance does not vary if metabolites' concentration time series can be described with a modified Bateman function (i.e. simulations v1, v2 v3 have little influence on its performance). Therefore, it is especially advantageous if this property cannot be guaranteed.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

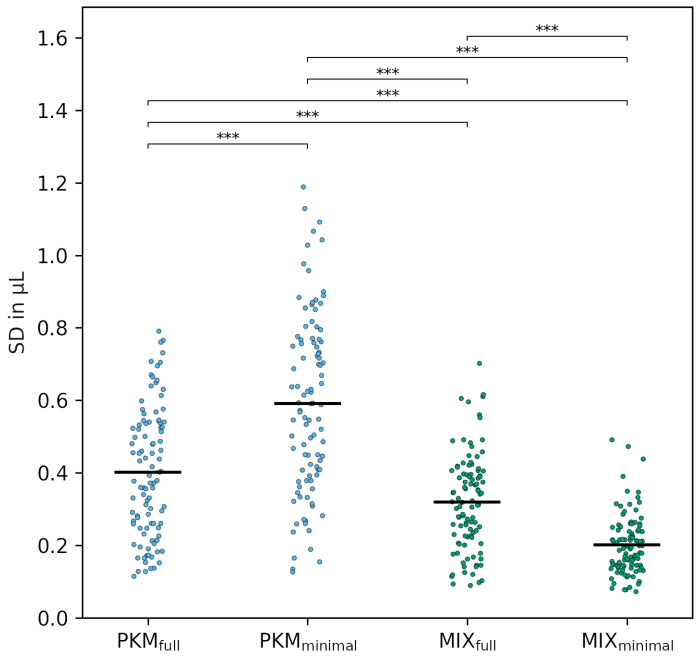

Figure 5 SD measures of simulation v3 with $n_{\text{metabolites}} = 60$. The significance between the methods was calculated on 100 paired replicates with the Wilcoxon signed-rank test.

4.2 Computational Performance

Analysis of metabolomics data sets is usually a computationally exhaustive process. There are several steps in (pre-)processing that need to be executed, many of them lasting for hours. Therefore, computational time can quickly stack to large numbers. Normalization models are no exception to this general rule. As $n_{\text{metabolites}}$ in a pharmacokinetic model increases, the time for optimization of pharmacokinetic models may become limiting. Therefore, we investigated the average time for one time series normalization for different methods and different numbers of metabolites.

The computational time spent for one optimization step as a function of $n_{\text{metabolites}}$ is given in Figure 6 for simulation v3. It increased for some normalization models, however not for all of them and not equally. Within the investigated range, PQN stays well under 1 second per normalization, whereas with PKM_{full} the normalization time increases drastically from $2 \pm 1\,\text{s}$ for a model with 4 metabolites to $72 \pm 48\,\text{s}$ for 60 metabolites. Similar, although even longer, normalization times were observed for MIX_{full} maxing out at $170 \pm 22\,\text{s}$ for $n_{\text{metabolites}} = 60$. In stark contrast to the exponential increase in computational power needed for full models are the minimal models. Their time to optimize stays nearly constant within the investigated metabolite range (Supplementary Table 7).

Here we demonstrate that $\text{MIX}_{\text{minimal}}$ is not only superior to other tested models in terms of its normalization performance, but also in terms of computational feasibility. We hypothesize that even data sets with thousands of untargeted metabolites will have a minor impact on its speed.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

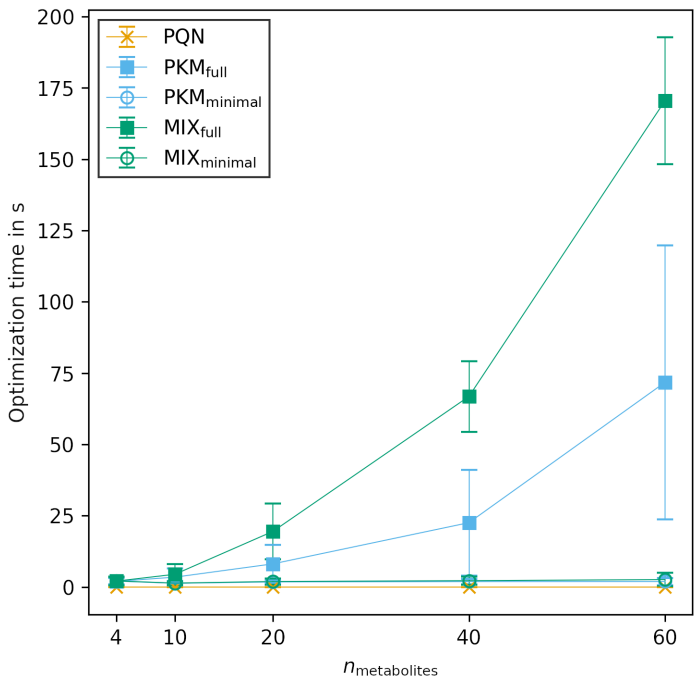

Figure 6 Time in seconds for optimization of one normalization model in simulation v3. The error bars represent the standard deviation of normalization times between 100 replicates.

4.3 Application on a Caffeine Network

In the original study, the authors identified and quantified four metabolites (caffeine, paraxanthine, theobromine, and theophylline) in a time series after the ingestion of a single dose of caffeine [23]. To investigate the performance of normalization models on a real finger sweat data set, we split all 3446 measured $\widetilde{\mathbf{M}}$ time series into three parts that contained two targeted metabolites each, only one shared by all, namely caffeine. Subsequently we fitted a $\text{PKM}_{\text{minimal}}$ and $\text{MIX}_{\text{minimal}}$ model ($\ell = 2$) with specialized kinetics (for details see 3.2) through the three sub data sets. Due to the nature of the metabolite subnetworks (Supplementary Figure S4) it is possible to calculate two kinetic constants describing the absorption and elimination of caffeine (k_a^{caf} and k_e^{caf}) in all three cases. As the data for all three subnetworks was measured in the same experiment we can assume that the underlying ground truth of these constants has to be the same. Therefore, by comparing the variance of k_a^{caf} and k_e^{caf} it is possible to infer the performance of normalization methods.

In Figure 7 the precisions of the $\text{PKM}_{\text{minimal}}$ (top panel) and $\text{MIX}_{\text{minimal}}$ (bottom panel) normalization methods are illustrated. It is evident that using $\text{MIX}_{\text{minimal}}$ improves the precision of estimated k_a^{caf} as well as k_e^{caf}. The standard deviation over all data points decreases from 32 % to 8 % for k_a^{caf} and from 42 % to 19 % for k_e^{caf} when switching to $\text{MIX}_{\text{minimal}}$ normalization.

In contrast to Figure 7 where the standard deviation was calculated over all 37 measured time series, in Figure 8 the standard deviation within one measured $\widetilde{\mathbf{M}}$ time series is illustrated. In the top panel of Figure 8 the standard deviations of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

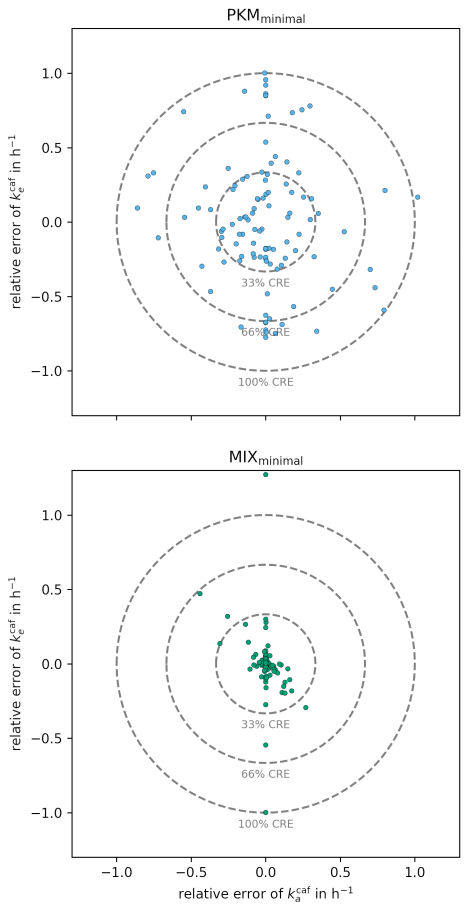

Figure 7 Cumulative relative error (CRE) of PKM$_{minimal}$ (top panel) and MIX$_{minimal}$ (bottom panel) on a real data set. Per measured $\widetilde{\mathbf{M}}$ time series the mean value of $k_a^{caf} = k_1'$ and $k_e^{caf} = k_2' + k_3'$ for the three subnetworks were assumed as ground truth. Relative errors of kinetic constants were calculated between subnetwork fits and assumed ground truth for all 37 measured time series. Therefore, the number of points per panel corresponds to the number of concentrations time series present in the data set times three. On the X-axis the relative error of $k_a^{caf} = k_1'$ and on the Y-axis the relative error $k_e^{caf} = k_2' + k_3'$ are plotted.

absorption constant of caffeine, k_a^{caf}, are shown for different normalization models. Their mean measured mass time series for PKM$_{minimal}$ is significantly larger than for both versions of the MIX$_{minimal}$ model ($p = 2.8 \times 10^{-5}$ and $p = 2.2 \times 10^{-5}$ for equal and metabolite-depending weighting respectively). The difference between two MIX$_{minimal}$ versions is due to the loss term weighting and discussed in the following section. On the bottom panel of Figure 8, the standard deviation of the elimination constant of caffeine is shown. There again, a significant decrease of the mean standard deviation for MIX$_{minimal}$ models was found ($p = 4.5\ 10^{-7}$ and $p = 4.2 \times 10^{-7}$ equal and metabolite-depending weighting respectively) compared to the previously used PKM$_{minimal}$ model.

We emphasize that the original study introducing pharmacokinetic normalization was especially focused on the absorption, conversion and degradation of caffeine [23]. Therefore, the kinetic network resembled specific kinetics of that metabolic

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

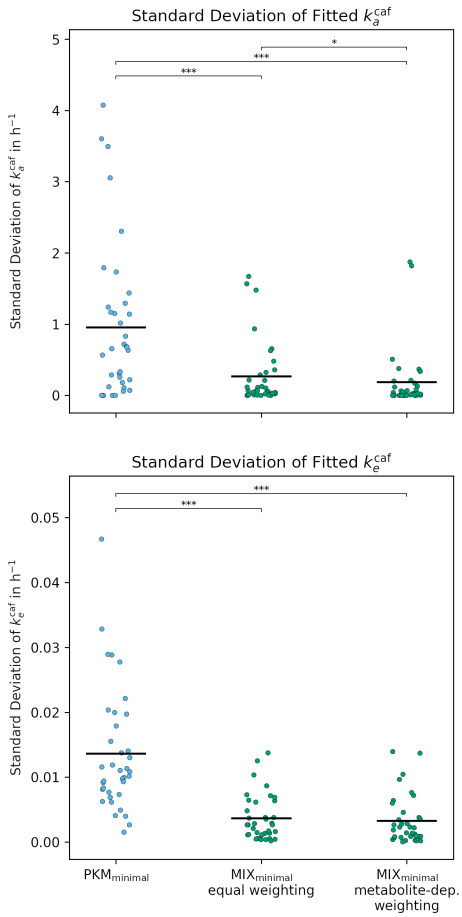

Figure 8 Performance of PKM$_{minimal}$ and MIX$_{minimal}$ on a real data set. On the top panel the standard deviation of $k_a^{caf} = k_1'$ fitted with three different subnetworks is plotted. On the bottom panel the standard deviation of $k_e^{caf} = k_2' + k_3'$ fitted with three different subnetworks is plotted.Two different strategies for the weighting of the loss terms were investigated; equal weighting in the middle panel, metabolite-depending weighting in the bottom panel. The number of points per method corresponds to the number of concentrations time series present in the real data set. The Wilcoxon signed-rank test was used to test for significant differences.

pathway. In contrast, PKM and MIX proposed in this study are generalized forms relying on the same principle, but the modified Bateman equation used here is a lot less constrained than the kinetic functions used for the description of the original kinetic network. However in this section, we demonstrate that the fundamental improvement found by switching from PKM to a MIX model can be also translated back again to a more specific metabolic network. In order to support this argument, we show the applicability of the $\text{MIX}_{\text{minimal}}$ normalization method on a real finger sweat data set. The results with real data emphasize the validity of the simulations done on synthetic data sets. They show that especially, when known metabolic networks are small, the $\text{MIX}_{\text{minimal}}$ model significantly improves the robustness of normalization and thus kinetic constants inferred from finger sweat time series measurements.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.4 Weighting MIX of Loss Terms

In the previous evaluations, we weighted the loss terms (defined in Equation 9a) of the MIX model according to the number of metabolites used for the calculation of each term (i.e. metabolite-depending weighting). However, in real world experiments typically the analytical method is optimized for targeted metabolites, which results in high confidence. Similar level of confidence may not be expected for all other metabolites. This differences in confidence can be accounted for in the loss weighting term λ (Equations 9b, 9c). Here we tested two diametric principles to calculate λ; metabolite-depending and equal weighting.

The results of the test are shown in Figure 8 (compare $\text{MIX}_{\text{minimal}}$ equal weighting and $\text{MIX}_{\text{minimal}}$ metabolite-depending weighting). In determining k_a^{caf} metabolite-dependent weighting outperforms equal weighting ($p = 1.7 \times 10^{-2}$). However, no significant differences between the weighting strategies are found for k_e^{caf}.

The reason for this investigation was that one would expect different confidences of measured metabolites. As the finger sweat data was analysed with optimizing the precision of caffeine and its degradation products in mind, it is reasonable to assume that their measured mass time series are more precise (i.e. have higher confidence) than ones from untargeted measurements. Another argument for the use of equal weighting is the fact that untargeted data may be highly correlated. However, it is *a priori* not feasible to differentiate between real correlations already present in $\mathbf{C}$ and correlations in $\widetilde{\mathbf{M}}$ that were introduced by V_{sweat} [38].

The difference in weighting methods for k_a^{caf} has a low significance level. As several pairwise comparisons were performed on the results and the p values are given without correction for multiple testing, we argue that this significance is coincidental. This is further supported by the fact that no significance was found for k_e^{caf}. Overall we hypothesize the value for λ would lie within the two extremes tested here, however, if no information about the confidence and correlation of different features is available, we argue that it is preferable to use equal weighted loss terms as it emphasises the better controlled targeted measurements (known calibration curve, easier quality control).

5 Discussion

In this study we present a generalized framework for the PKM normalization model, first introduced in reference [23]. Moreover, we extend the existing model to incorporate untargeted metabolite information, dubbed as MIX model. Both models are implemented in Python and are openly available at GitHub `https://github.com/Gotsmy/sweat_normalization`.

The quality of normalization methods was tested on synthetic data sets. Synthetic data sets are necessary as it is impossible to obtain validation data without fundamentally changing the (finger) sweat sampling method as described above [23]. However, three different synthetic data generation methods (v1, v2, v3) were employed to ensure that synthetic data sets are close to real data as possible. We found that, when $n_{\text{metabolites}} \geq 60$, $\text{MIX}_{\text{minimal}}$ performs equally well or better than all other tested normalization methods.

Despite true V_{sweat} values remaining unknown, the real finger sweat data can be used as validation for relative robustness of normalization methods. There,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\text{MIX}_{\text{minimal}}$ significantly outperforms $\text{PKM}_{\text{minimal}}$ independently from how λ is calculated. The decreased variance of kinetic constants estimated by $\text{MIX}_{\text{minimal}}$ likely originates from the fact that Q^{PQN} does not differ much for three subsets as long as $n_{\text{metabolites}} = 60$ is exceeded in each subset (which is the case here). On the other hand, as only few data points are used for $\text{PKM}_{\text{minimal}}$ optimization, small errors in one of the two targeted metabolites measured mass have a high potential to change the normalization result.

To recapitulate, the proposed $\text{MIX}_{\text{minimal}}$ model has several crucial advantages over other tested methods.

- $\text{MIX}_{\text{minimal}}$ significantly outperforms $\text{PKM}_{\text{minimal}}$ in relative (rSD, $-56 \pm 23\,\%$) and absolute (SD, $-59 \pm 24\,\%$) errors with as little as 60 untargeted metabolites used as additional information (Figure 5).
- $\text{MIX}_{\text{minimal}}$ is invariant to whether untargeted metabolites follow an easily describable kinetic concentration curve (compare Supplementary Figure S6 to Supplementary Figure S7).
- $\text{MIX}_{\text{minimal}}$ performs equally well as PQN for relative abundances, but additionally it estimates absolute values of V_{sweat}, similar to pharmacokinetic (PKM) models (Figure 4).
- Moreover, pharmacokinetic (PKM) models need at least two metabolite time courses with known kinetics to be feasible [23], for $\text{MIX}_{\text{minimal}}$, however, strictly speaking only one metabolite in the PKM part is necessary as the size effect changes can be accounted for in the PQN part of the model.
- $\text{MIX}_{\text{minimal}}$ performs well in this proof of principle study, moreover, it is a basis for further improvements. Firstly, different, more sophisticated statistical normalization methods (e.g. EigenMS [25]) can be used as input for the PQN part of the model. Secondly, Bayesian priors describing uncertainties of different metabolites can be implemented over the λ parameter in a similar fashion as discussed in reference [39].
- Strikingly, the results showed that for all normalization methods tested the SD and rSD values flattened once 60 metabolites were present in the original information. This suggested that the presented normalization models, especially $\text{MIX}_{\text{minimal}}$ can be applied even for biomatrices or analytical methods with as few as 60 compounds measured.

6 Conclusion

In this study we described and define two metabolomics time series normalization models: PKM and MIX. Subsequently, we elaborated several advantages of the $\text{MIX}_{\text{minimal}}$ model over PKM and previously published normalization methods. We are confident that this will further improve the reliability of metabolomic studies done on finger sweat and other non-conventional biofluids. However, we acknowledge that a more thorough investigation with data sets of several more quantified metabolites and determined sweat volumes need to be carried out to assess the full potential of the proposed method.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendix

Funding

This research received no external funding. Open Access Funding by the University of Vienna.

Abbreviations

Symbol	Name
a_{sample}	sampling skin area
b	part of modified Bateman function
$C, \mathbf{C}$	underlying concentration (vector)
c_0	kinetic parameter
d	kinetic parameter
F	modified Bateman function
$V_{\text{sweat}}^{\text{ref}}$	PQN correction factor
i	time point index
j	metabolite index
k	kinetic parameter
ℓ	metabolites used for kinetic fitting
$\ell+$	metabolites not used for kinetic fitting
$\mathcal{L}$	loss
L	loss function
lag	kinetic parameter
$\widetilde{M}, \widetilde{\mathbf{M}}$	measured mass (vector)
M^{ref}	reference mass for PQN
m/z	mass over charge ratio
$n_{\text{metabolites}}$	number of metabolites
$n_{\text{time points}}$	number of time points
p	p-value
q_{sweat}	sweat rate
$Q^{\mathbf{C}}$	median concentration fold change of two samples
$Q^{\mathbf{M}}$	median mass fold change of two samples
Q^{PQN}	normalization quotient calculated by PQN
R^2	coefficient of determination
rSD	relative measure of goodness of normalization
SD	absolute measure of goodness of normalization
t	time
V_{sweat}	collected sweat volume
v1, v2, v3	synthetic data sets
$\boldsymbol{\epsilon}$	experimental error vector
$\boldsymbol{\theta}$	kinetic parameter vector for fitting
λ	loss weighting parameter
τ	time to collect one sample

Availability of data and materials

All analysis (except stated otherwise) was performed in Python 3.7 heavily relying on NumPy [40], Pandas [41], and SciPy [29]. Code and generated data is available on GitHub `https://github.com/Gotsmy/sweat_normalization` under the GNU GPL version 3.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

	MG‡	JB‡	CB	CG	JZ
Conceptualization	■				■
Funding acquisition				■	■
Methodology	■	■			
Software	■				
Formal analysis	■		■		
Investigation	■	■			
Resources				■	■
Validation	■	■	■		
Supervision				■	■
Visualization	■				
Writing – original draft	■				■
Writing – review & editing	■	■	■	■	■

Author details

[1] Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria. [2] Joint Metabolome Facility, University and Medical University of Vienna, Vienna, Austria.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Jang, M., Costa, C., Bunch, J., Gibson, B., Ismail, M., Palitsin, V., Webb, R., Hudson, M., Bailey, M.: On the relevance of cocaine detection in a fingerprint. Scientific reports **10**(1), 1–7 (2020)
2. Delgado-Povedano, M., Calderón-Santiago, M., de Castro, M.L., Priego-Capote, F.: Metabolomics analysis of human sweat collected after moderate exercise. Talanta **177**, 47–65 (2018)
3. Brunmair, J., Bileck, A., Stimpfl, T., Raible, F., Del Favero, G., Meier-Menches, S.M., Gerner, C.: Metabo-tip: a metabolomics platform for lifestyle monitoring supporting the development of novel strategies in predictive, preventive and personalised medicine. EPMA Journal, 1–13 (2021)
4. Czerwinska, J., Jang, M., Costa, C., Parkin, M.C., George, C., Kicman, A.T., Bailey, M.J., Dargan, P.I., Abbate, V.: Detection of mephedrone and its metabolites in fingerprints from a controlled human administration study by liquid chromatography-tandem mass spectrometry and paper spray-mass spectrometry. Analyst **145**(8), 3038–3048 (2020)
5. Calderón-Santiago, M., Priego-Capote, F., Turck, N., Robin, X., Jurado-Gámez, B., Sanchez, J.C., De Castro, M.D.L.: Human sweat metabolomics for lung cancer screening. Analytical and bioanalytical chemistry **407**(18), 5381–5392 (2015)
6. Cui, X., Zhang, L., Su, G., Kijlstra, A., Yang, P.: Specific sweat metabolite profile in ocular behcet's disease. International Immunopharmacology **97**, 107812 (2021)
7. Harshman, S.W., Browder, A.B., Davidson, C.N., Pitsch, R.L., Strayer, K.E., Schaeublin, N.M., Phelps, M.S., O'Connor, M.L., Mackowski, N.S., Barrett, K.N., *et al.*: The impact of nutritional supplementation on sweat metabolomic content: a proof-of-concept study. Frontiers in chemistry **9**, 255 (2021)
8. Hussain, J.N., Mantri, N., Cohen, M.M.: Working up a good sweat–the challenges of standardising sweat collection for metabolomics analysis. The Clinical Biochemist Reviews **38**(1), 13 (2017)
9. Harshman, S.W., Strayer, K.E., Davidson, C.N., Pitsch, R.L., Narayanan, L., Scott, A.M., Schaeublin, N.M., Wiens, T.L., Phelps, M.S., O'Connor, M.L., *et al.*: Rate normalization for sweat metabolomics biomarker discovery. Talanta **223**, 121797 (2021)
10. Kuwayama, K., Tsujikawa, K., Miyaguchi, H., Kanamori, T., Iwata, Y.T., Inoue, H.: Time-course measurements of caffeine and its metabolites extracted from fingertips after coffee intake: a preliminary study for the detection of drugs from fingerprints. Analytical and bioanalytical chemistry **405**(12), 3945–3952 (2013)
11. Kuwayama, K., Yamamuro, T., Tsujikawa, K., Miyaguchi, H., Kanamori, T., Iwata, Y.T., Inoue, H.: Time-course measurements of drugs and metabolites transferred from fingertips after drug administration: usefulness of fingerprints for drug testing. Forensic Toxicology **32**(2), 235–242 (2014)
12. Baker, L.B.: Physiology of sweat gland function: The roles of sweating and sweat composition in human health. Temperature **6**(3), 211–259 (2019)
13. Taylor, N.A., Machado-Moreira, C.A.: Regional variations in transepidermal water loss, eccrine sweat gland density, sweat secretion rates and electrolyte composition in resting and exercising humans. Extreme physiology & medicine **2**(1), 4 (2013)
14. Ando, H., Noguchi, R.: Dependence of palmar sweating response and central nervous system activity on the frequency of whole-body vibration. Scandinavian journal of work, environment & health, 216–219 (2003)
15. Zhong, B., Jiang, K., Wang, L., Shen, G.: Wearable sweat loss measuring devices: From the role of sweat loss to advanced mechanisms and designs. Advanced Science, 2103257 (2021)
16. Harshman, S.W., Pitsch, R.L., Smith, Z.K., O'Connor, M.L., Geier, B.A., Qualley, A.V., Schaeublin, N.M., Fischer, M.V., Eckerle, J.J., Strang, A.J., *et al.*: The proteomic and metabolomic characterization of exercise-induced sweat for human performance monitoring: A pilot investigation. PLoS One **13**(11), 0203133 (2018)
17. Sonner, Z., Wilder, E., Heikenfeld, J., Kasting, G., Beyette, F., Swaile, D., Sherman, F., Joyce, J., Hagen, J., Kelley-Loughnane, N., *et al.*: The microfluidics of the eccrine sweat gland, including biomarker partitioning, transport, and biosensing implications. Biomicrofluidics **9**(3), 031301 (2015)
18. Du, Q., Zhang, Y., Wang, J., Chang, J., Wang, A., Ren, X., Liu, B.: Quantitative analysis of 17 hypoglycemic drugs in fingerprints using ultra-high-performance liquid chromatography/tandem hybrid triple quadrupole linear ion trap mass spectrometry. Rapid Communications in Mass Spectrometry **36**(1), 9199 (2022)
19. Filzmoser, P., Walczak, B.: What can go wrong at the data normalization step for identification of biomarkers? Journal of Chromatography A **1362**, 194–205 (2014)
20. Choi, J., Bandodkar, A.J., Reeder, J.T., Ray, T.R., Turnquist, A., Kim, S.B., Nyberg, N., Hourlier-Fargette, A., Model, J.B., Aranyosi, A.J., *et al.*: Soft, skin-integrated multifunctional microfluidic systems for accurate colorimetric analysis of sweat biomarkers and temperature. ACS sensors **4**(2), 379–388 (2019)
21. Kim, S.B., Koo, J., Yoon, J., Hourlier-Fargette, A., Lee, B., Chen, S., Jo, S., Choi, J., Oh, Y.S., Lee, G., *et al.*: Soft, skin-interfaced microfluidic systems with integrated enzymatic assays for measuring the concentration of ammonia and ethanol in sweat. Lab on a Chip **20**(1), 84–92 (2020)
22. Ragan, T.J., Bailey, A.P., Gould, A.P., Driscoll, P.C.: Volume determination with two standards allows absolute quantification and improved chemometric analysis of metabolites by nmr from submicroliter samples. Analytical chemistry **85**(24), 12046–12054 (2013)
23. Brunmair, J., Gotsmy, M., Niederstaetter, L., Neuditschko, B., Bileck, A., Slany, A., Feuerstein, M.L., Langbauer, C., Janker, L., Zanghellini, J., *et al.*: Finger sweat analysis enables short interval metabolic biomonitoring in humans. Nature Communications **12**(1), 1–13 (2021)
24. Warrack, B.M., Hnatyshyn, S., Ott, K.-H., Reily, M.D., Sanders, M., Zhang, H., Drexler, D.M.: Normalization strategies for metabonomic analysis of urine samples. Journal of Chromatography B **877**(5-6), 547–552 (2009)
25. Karpievitch, Y.V., Nikolic, S.B., Wilson, R., Sharman, J.E., Edwards, L.M.: Metabolomics data normalization with eigenms. PloS one **9**(12), 116221 (2014)
26. Dieterle, F., Ross, A., Schlotterbeck, G., Senn, H.: Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. application in 1h nmr metabonomics. Analytical chemistry **78**(13), 4281–4290 (2006)

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.17.476591; this version posted January 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., Cao, Q., Xue, W., Chen, N., Zhu, F.: Performance evaluation and online realization of data-driven normalization methods used in lc/ms based untargeted metabolomics analysis. Scientific reports **6**(1), 1–13 (2016)
28. Di Guida, R., Engel, J., Allwood, J.W., Weber, R.J., Jones, M.R., Sommer, U., Viant, M.R., Dunn, W.B.: Non-targeted uhplc-ms metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. Metabolomics **12**(5), 93 (2016)
29. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., *et al.*: Scipy 1.0: fundamental algorithms for scientific computing in python. Nature methods **17**(3), 261–272 (2020)
30. Brunmair, J., Gotsmy, M., Niederstaetter, L., Neuditschko, B., Bileck, A., Slany, A., Feuerstein, M.L., Langbauer, C., Janker, L., Zanghellini, J., et al.: Finger Sweat Analysis Enables Short Interval Metabolic Biomonitoring in Humans. doi:10.5281/zenodo.5222967. `https://doi.org/10.5281/zenodo.5222967`
31. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., *et al.*: A cross-platform toolkit for mass spectrometry and proteomics. Nature biotechnology **30**(10), 918–920 (2012)
32. Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., *et al.*: A lipidome atlas in ms-dial 4. Nature biotechnology **38**(10), 1159–1163 (2020)
33. Lazar, C.: imputeLCMD: A Collection of Methods for Left-censored Missing Data Imputation. (2015). R package version 2.0. `https://CRAN.R-project.org/package=imputeLCMD`
34. Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., Ni, Y.: Missing value imputation approach for mass spectrometry-based metabolomics data. Scientific reports **8**(1), 1–10 (2018)
35. Csajka, C., Haller, C., Benowitz, N., Verotta, D.: Mechanistic pharmacokinetic modelling of ephedrine, norephedrine and caffeine in healthy subjects. British journal of clinical pharmacology **59**(3), 335–345 (2005)
36. Kamimori, G.H., Karyekar, C.S., Otterstetter, R., Cox, D.S., Balkin, T.J., Belenky, G.L., Eddington, N.D.: The rate of absorption and relative bioavailability of caffeine administered in chewing gum versus capsules to normal healthy volunteers. International journal of pharmaceutics **234**(1-2), 159–167 (2002)
37. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin **1**(6), 80–83 (1945)
38. Saccenti, E.: Correlation patterns in experimental data are affected by normalization procedures: consequences for data analysis and network inference. Journal of Proteome Research **16**(2), 619–634 (2017)
39. Sheiner, L.B., Beal, S.L.: Bayesian individualization of pharmacokinetics: simple implementation and comparison with non-bayesian methods. Journal of pharmaceutical sciences **71**(12), 1344–1348 (1982)
40. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with NumPy. Nature **585**(7825), 357–362 (2020). doi:10.1038/s41586-020-2649-2
41. pandas development team, T.: Pandas-dev/pandas: Pandas. doi:10.5281/zenodo.3509134. `https://doi.org/10.5281/zenodo.3509134`