Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Xiangning Xue Date

Comparison of Imputation Methods on Metabolomics Data with Triplet Data

By

Xiangning Xue
Master of Public Health

Biostatistics and Bioinformatics

_____

Tianwei Yu, PhD
Committee Chair

_____

Xiangqin Cui, PhD
Committee Member

Comparison of Imputation Methods on Metabolomics Data with Triplet Data

By

Xiangning Xue

B.S.
Xiamen University
2016

Thesis Committee Chair: Tianwei Yu, PhD
Committee Member: Xiangqin Cui, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2019

Abstract

Comparison of Imputation Methods on Metabolomics Data with Triplet Data

By Xiangning Xue

Missing value imputation in mass spectrometry-based metabolomics data is important for subsequent data analysis. There are many methods available for tackling the problem, most of which were initially developed for microarray or RNA sequencing data. Metabolomics data represent unique challenges in missing value imputation. Some missingness in the data are indeed missing, which we call true missings, while others may represent true non-existence of the metabolite, which can be called true zeros. It is difficult to differentiate the true missings from true zeros in the dataset. Most of the current imputation methods would impute all the missingness. In addition, assessment of imputation methods based on the knockout-impute scheme may not represent the true performance of the imputation methods on metabolomics data, as the true missingness mechanism is complicated. In this study, we utilized datasets with triplicate measures on each sample, which offers some unique advantage over the knockout-impute scheme. Taking one measurement from each sample at a time, the remaining two measurements offer information as to whether each missing location is more likely to be true missing or true zero. With this data set, we were able to evaluate the performance of different imputation methods, assessing their performance on true missing and true zeros. The result shows that SVD and LLS tend to have better performance with true missings, and scImpute performs better for the true-zeros but not as reliable for true missings.

Comparison of Imputation Methods on Metabolomics Data with Triplet Data

By

Xiangning Xue

B.S.
Xiamen University
2016

Thesis Committee Chair: Tianwei Yu, PhD
Committee Member: Xiangqin Cui, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2019

Table of Contents

1. Introduction

Metabolites are all the small molecular weight intermediate products and end products involved in the metabolic processes. For example, in the tricarboxylic acid cycle, carbohydrates, fats, and proteins break down to metabolites and are oxidized to provide energy for the human body. Metabolomics is the study of metabolite profiles of biological samples like blood, urine, and tissues. It is widely acknowledged that the health status of an individual is determined by the genomic feature, personal behavior, and environmental exposure, and metabolomics has proved its ability to capture all these features (National Academies of Sciences & Medicine, 2016). Metabolomic profiles provide a snapshot of the biological processes, some of which closely related to disease status (Rodrigues et al., 2019). These metabolomic profiles can indirectly reflect the genomic features of individuals, thus is helpful in revealing the biological pathways of the disease. Some metabolic perturbations come from personal behavior/exposure, such as eating habits, and could affect the risk of certain chronic disease (Rothwell et al., 2019). The study of exposome, i.e. environmental exposure, is a relatively new field. The metabolomic profile is a common tool for the study of exposome, such as detection of chemical compounds in the exposome (Bloszies & Fiehn, 2018).

Liquid chromatography-mass spectrometry (LC-MS) is one of the most commonly used techniques for acquiring metabolomics data. To be quantified by the detector, the metabolite molecules are first converted into ions by an ion source, then they are resolved by the mass analyzer in a time-of-flight tube or an electromagnetic field. LC-MS can efficiently separate the metabolites in biological samples based on the mass-to-charge ratio of the chemicals (Turi, Romick-Rosendale, Ryckman, & Hartert, 2018). In addition to LC-MS, GC-MS and NMR are also used for generating metabolomics data (Madji Hounoum, Blasco, Emond, & Mavel, 2016).

The quality of LC-MS data is affected by many factors, such as sample loss during metabolite extraction and pre-processing of data. One of the most common problems associated is missing values. In statistical analyses, missing values are categorized into three types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Missing completely at random (MCAR) happens when the missingness is not associated with any observable or unobservable parameters. Observations missing at random (MAR) are associated with certain variables but are independent of their value. On the other hand, missing not at random (MNAR) refers to missingness that is related to the value of the variables. LC-MS missingness is likely a combination of all three. In the LC-MS data generation process, MCAR arises from random errors such as incomplete ionization. One MAR example in LC-MS data is co-eluting compounds, which is chemical compounds that is difficult to separate and identify with chromatographic column. Increasing concentration of co-eluting compounds would suppress the signal. Limit of quantification (LOQ) is one common case of MNAR where metabolites with concentration lower than the detection threshold are found missing. Another cause of MNAR in LC-MS data is ion suppression caused by matrix effects, which is the presence of interfering components in the sample extract. The consequence of ion suppression is multiple, including affecting ion ratio, linearity, and non-detection of certain metabolites which lead to MNAR of that metabolite (Antignac et al., 2005).

Imputation is commonly used to fill the missing data before down-stream analyses. Common strategies for imputing MCAR/MAR data include local similarity approaches like KNN (K nearest neighbors) and global-structure approaches like BPCA (Bayesian principal component analysis). For MNAR, we expect the missing values to be lower than a certain threshold if the reason for missing is LOQ. Usually, the MNAR data are imputed with single-value approaches that use a constant or a random abundance to replace the missing value. Common single-value approaches include replacing all the missing values with the global limit of detection (LOD) -

minimum of the observed abundance (LOD1), half of the minimum value (LOD2), a randomly generated value from the left tail of the proposed distribution of the dataset (RT1), zeros, mean of the global data set, or median of the global data set.

Different imputation methods suit different missing types. However, in practice, it is difficult to identify the nature of the missingness in a dataset. As a result, MCAR/MAR imputation methods are often used to fill all the missing values in data analysis. Numerous studies have indeed shown that MCAR/MAR imputation methods applied to all missing data tends to generate better results than using only MNAR method for all the missing data (Armitage, Godzien, Alonso-Herranz, López-Gonzálvez, & Barbas, 2015; Lazar, Gatto, Ferro, Bruley, & Burger, 2016; Webb-Robertson et al., 2015). This result makes sense because MNAR is usually small values. The imputation methods for MNAR would perform badly because of their focus on left-censored data (Lazar et al., 2016). An added difficulty in LC/MS data is that MNAR can be caused by ion suppression, which means the true value may not be below the detection threshold (Antignac et al., 2005).

One potential solution to the issue of mixed types of missingness is to identify the MNAR from all the missing values and assign the imputed value as zero while imputing the rest of the missing values with common MCAR/MAR imputation methods. One such method is scImpute (Li & Li, 2018). scImpute was proposed to address the "dropout" phenomenon in single-cell RNA sequencing (scRNA-seq) data, where a gene is observed in one cell but undetected in another cell. This method could be applied to metabolomics data because of the similarity of the situation. However given the differences in data characteristics, as well as the different mechanisms of MNAR, we do not yet know whether the MNAR estimation in scImpute works in metabolomics data. In this study, scImpute was evaluated along with common MCAR/MAR methods, i.e. K-nearest neighbors (KNN), Bayesian principal component analysis (BPCA), Singular value decomposition (SVD) and (Local least squares) LLS.

One major difficulty of generating MNAR data for the evaluation of imputation methods is the lack of knowledge of the true mechanism of missingness. A knockout-impute scheme cannot faithfully represent MNAR situations in real data. In this study, we tried to answer the question using a triplicate metabolomics dataset. In this dataset, each subject has three measurements of abundance for each metabolic feature. Thus, the abundance of the feature might be missing in one, two, or all three measurements. Since the three measurements were taken from the same sample, their true measures of abundance should be close, and the probability for them to be missing should be the same, assuming the pre-processing of the data gave a constant performance in the three measurements. It is unlikely that three measurements would be all be missing due to MCAR and MAR, then we can presume that the values with one or two missing measurements, especially one missing, are MCAR/MAR, while those with three missing measurements are more likely a combination of true zero and MNAR.

2. Method

2.1 The Data Set

This data set comes from an untargeted study measured with liquid chromatography-mass spectrometry. The data set was generated from the Emory-Georgia Tech Predictive Health Initiative, Cohort of the Center for Health Discovery and Well Being (CHDWB). It contains both positive- and negative- ion mode data, each of which contains triplet measures of 498 individuals, i.e. the sample from each person were measured three times. The two matrices, one from positive ion mode and one from negative ion mode, were analyzed separately.

The distribution of the original feature abundance is heavily right-skewed. We applied log-transformation, $y_{ij} = \log(x_{ij} + 1)$ (where x_{ij} is the original data and y_{ij} is the transformed data), to the data before imputation. The log transformation is already implemented in scImpute, so we passed the original data to the scImpute function.

2.2 Creating the Reference Matrices

For both positive ion mode data and negative ion mode data, a reference matrix, which contains the true value of abundance for all metabolic features, was computed from the triplet data. Since there were three measures for each metabolic feature, the strategy for computing was as follows: (1) if all the three measures were missing, then the reference value was set to 0 according to the assumptions claimed in the introduction; (2) if not all three measures were missing, the reference value was the mean of the non-missing values.

2.3 Imputation Method

2.3.1 scImpute (Li & Li, 2018)

R package scImpute implements this new method that tries to tackle the missing value problem in single-cell RNA sequencing data. Single-cell RNA sequencing (scRNA-seq) is a technology that quantifies the RNA at the cellular level. scRNA-seq data contains the count values of different RNAs which displays some similar distribution properties with the abundance levels in metabolomics data. It also has the problem of MNAR that comprise the true zero caused by limit of detection (LOD). In the case of scRNA-seq data, the true zero represents the non-expression of certain genes in the subject cell.

2.3.2 K-Nearest Neighbors (KNN) (Botstein et al., 2001)

KNN finds the k nearest neighbors of the metabolite basing on the Euclidean metric of the abundance of the metabolite in samples where it is not missing. The distance between the metabolite and all other metabolites is defined as the average distance of all the non-missing abundance across the samples. The missing abundance is then imputed as the average abundance of the k nearest neighbors. This method is included in the R package *impute*.

2.3.3 Bayesian Principal Component Analysis (BPCA) (Oba et al., 2003)

The extraction of Principal Components (PCs) reduces the dimension of the data. The values in the original matrix can be expressed as a linear combination of the PCs with an error term. BPCA assumes that the PCs and error terms obey normal distributions (Tipping & Bishop, 1999). The coefficients for the PCs and the missing values are estimated with an EM-like repetitive algorithm called variation Bayes (VB) algorithm. This method is included in the R package *pcaMethod*.

2.3.4 Singular Value Decomposition (SVD) (Botstein et al., 2001)

SVD imputation approximates the missing value as a linear combination of a set of mutually orthogonal eigenvalues that are derived from the principle components of the data matrix. The algorithm estimates the missing values iteratively until convergence.

2.3.5 Local Least Squares (LLS) (Golub, Park, & Kim, 2004)

This method selects several metabolic features that are most informative with regard to the specific feature to be imputed based on Pearson, Spearman or Kendall correlation coefficients. Then the imputation is conducted using linear regression.

2.4 Imputation Scheme and Evaluation Criteria.

2.4.1 Imputation Scheme

For each of the two data sets, we split the data matrix into three sub-matrices, with each sub-matrix containing a single measurement from each subject. We then run the imputation methods on the three sub-matrices respectively. The imputation results are compared with the corresponding reference matrix using the two criteria in the following sub-sections. The results from the three sub-matrices are averaged.

2.4.2 Normalized Root Mean Squared Error (NRMSE)

NRMSE is a common scale-free method to evaluate the accuracy of imputation with the following formula:

$$NRMSE_{\{i,j:\, y_{ij}\ missing\}} = \sqrt{\frac{mean((y_{ij} - y_{ij}^{impute})^2)}{var(y_{ij})}},$$

where y_{ij} represents the reference data and y_{ij}^{impute} represents the imputed data. We use NRMSE to select the optimal parameters for different algorithms.

2.4.3 Log-transformed root mean squared error (LRMSE)

For the true zero/MNAR group, i.e. the group with reference value of zero, NRMSE is not applicable. As a result, LRMSE is used to compare the imputation accuracy of metabolic features with different degrees of missingness (Brock, Shaffer, Blakesley, Lotz, & Tseng, 2008). LRMSE is calculated as:

$$LRMSE_{\{i,j:\, y_{ij}\ missing\}} = \sqrt{\frac{\Sigma(y_{ij}^{impute} - y_{ij})^2}{\#\{y_{ij}\ missing\}}}.$$

3. Result

3.1 1. Relationship between the number of missings and metabolic feature abundance

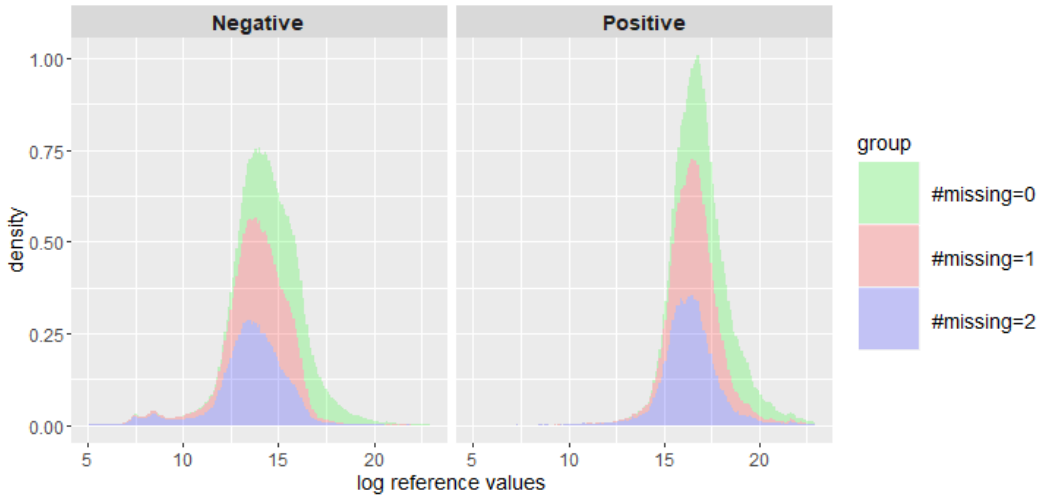

Figure 1: Relationship of number of missings in the triplets and the metabolic feature abundance

In this study we assumed that missingness is random for those with one or two missing measures in the triplets. However, there is evidence against such an assumption. We have observed a difference in the log-scale abundance of those with different numbers of missing values (Figure 1). Although not a clear distinction in the plot, the p-value of the t-tests shows that the three distributions have the relationship $\mu_{\#missing=2} < \mu_{\#missing=1} < \mu_{\#missing=0}$ for both data sets (Table 1), i.e. the smaller the abundance is, the greater the probability that the value would be missing. This conflicts with the missing at random assumption. On the other hand, the scale of the difference is not substantial.

Table 1: Mean Log Abundance vs. Number of Missing Values

	$\mu_{\#missing=0}$	$\mu_{\#missing=1}$	$\mu_{\#missing=2}$
Positive ion mode	17.65	16.76	16.38
Negative ion mode	15.36	14.06	13.52

3.2 2. Optimal parameters for the imputation methods

Parameter setting might influence the performance of different methods. For the comparison to be more reasonable, it needs to be done with the optimal parameters of each method. In LLS, KNN and scImpute, we need to prespecify the number of clusters of the metabolic features, while for SVD and BPCA, the number of PCs need to be assigned. For scImpute, we need to set an extra parameter *t*, which is the threshold of dropout probability to determine if the function would impute the missing value or leave it as zero. The optimal parameters were selected with the least average NRMSE using grid search for the positive ion mode data set and the negative ion mode data set separately.

For BPCA, we tested the number of PCs = 2, 4, …, 30 and selected *nPCs* = 14 for the positive ion mode and *nPCs* = 16 for the negative ion mode; for KNN, we tested the number of neighbors = 2, 4, …, 30 and selected k = 12 for the positive ion mode and k = 28 for the negative ion mode; for LLS, we tested the number of clusters = 2, 4, …, 30 and selected k = 10 for both positive ion mode and negative ion mode; for SVD, we tested the number of PCs = 2, 4, …, 30 and selected *nPCs* = 22 for the positive ion mode and *nPCs* = 6 for the negative ion mode; for scImpute, we tested k = 2, 4, 6, …30, t = 0.1, 0.2, …0.9, 0.95, 0.99, and selected k = 12, t = 0.8 for the positive ion mode data and k = 18, t = 0.7 for the negative mode data. The selected parameters are also annotated in Figure 3.

3.3 3. Correlation between imputed value and true value

Overall, BPCA, KNN and LLS showed similar patterns of correlation between imputed values and reference values. The scatterplot of SVD showed greater variance of imputed values with the

positive ion mode data set. The scatterplot of scImpute shows that it gives imputed values that are a lot smaller than the reference values for a good portion of the data.

We can also see that while other imputation methods fail to distinguish true zeros by giving imputed values to true zeros along the range of the other missing values, scImpute would impute many missing values as zero. The proportion of true zeros correctly identified is 17.1% for the positive data set and is 51.9% for the negative data set. However, we can see that scImpute wrongly identifies many missing values with non-zero reference value as true zeros. The false discovery rate is 66.3% for the positive data set and 60.0% for the negative dataset. Unlike single cell RNA-seq data, the clusters in the metabolomics data tend to be less compact, causing scImpute to make errors when attempting to identify the true-zeros.

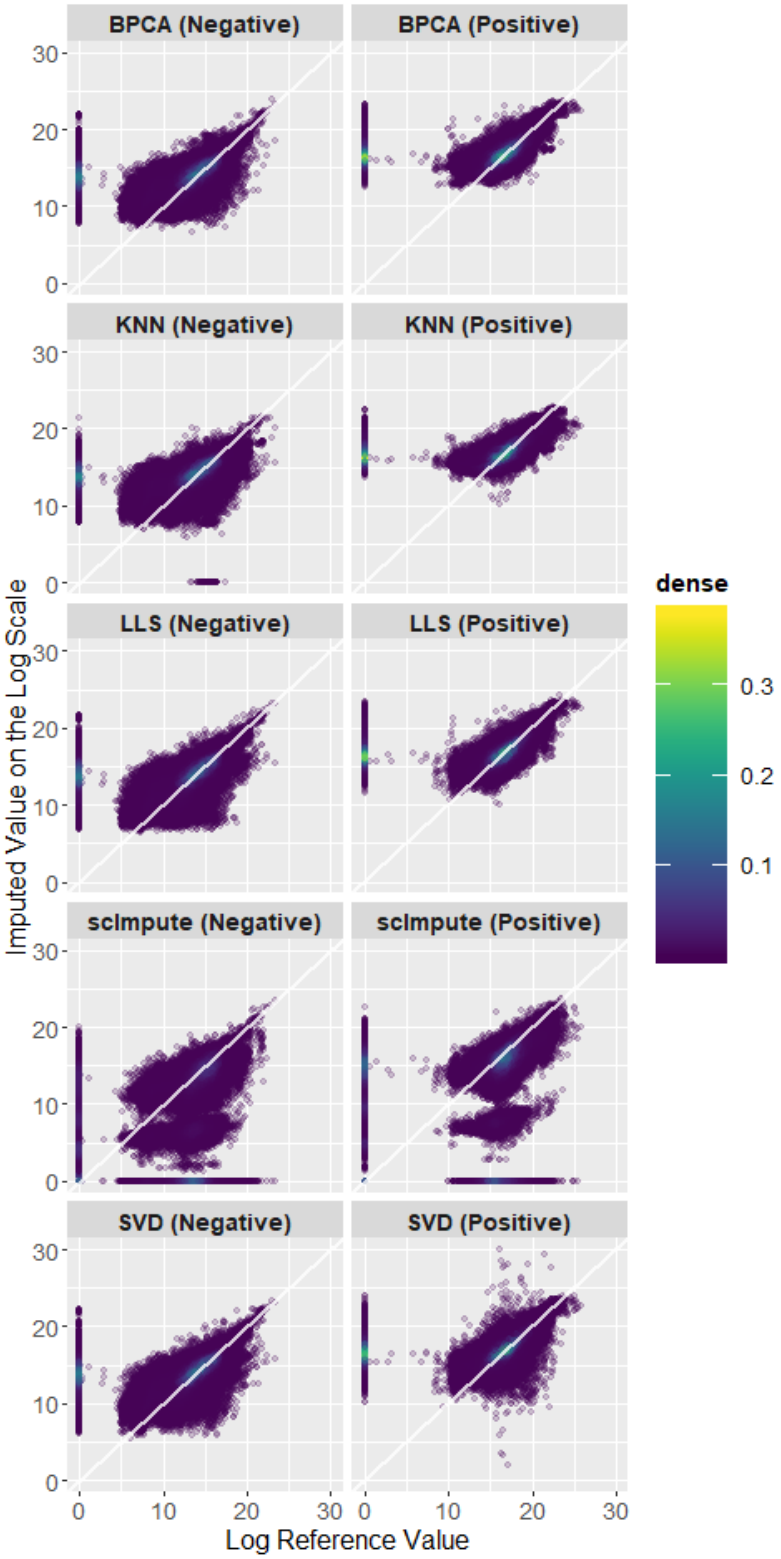

Figure 2. Scatter plot of imputed values against reference values.

3.3.4. Imputation efficiency with different numbers of missing

We expect that with more values missing in the triplicate measurements, we would have less information of the true value, thus less accurate evaluation of the imputation method and larger LRMSE. The trend is clear comparing the first and second rows of Figure 3.

It is also noticeable that the LRMSE of scImpute is larger than other imputation methods for the single and double missing cases (Figure 3; first and second row) and is smaller than other imputation methods for those triple missing cases (Figure 3; third row). This makes sense if we consider its property of identifying true zeros that we saw in the scatterplot in Figure 2. Since the accuracy of such identification is limited, scImpute performs not so well when the reference value is non-zero but would perform better when the reference value is zero.

We can see distinctively that the LRMSE is a lot larger for those metabolic features that are missing in all three triplets. This is because we assume the true value is zero, however, our current imputation methods do not perform well in terms of distinguishing true zeros from the missing values, including scImpute. The cases with three missings are a mixture of true zeros and MNAR cases caused by ion suppression. Thus, the results cannot fully reflect the true performance of the methods on the true zeros.

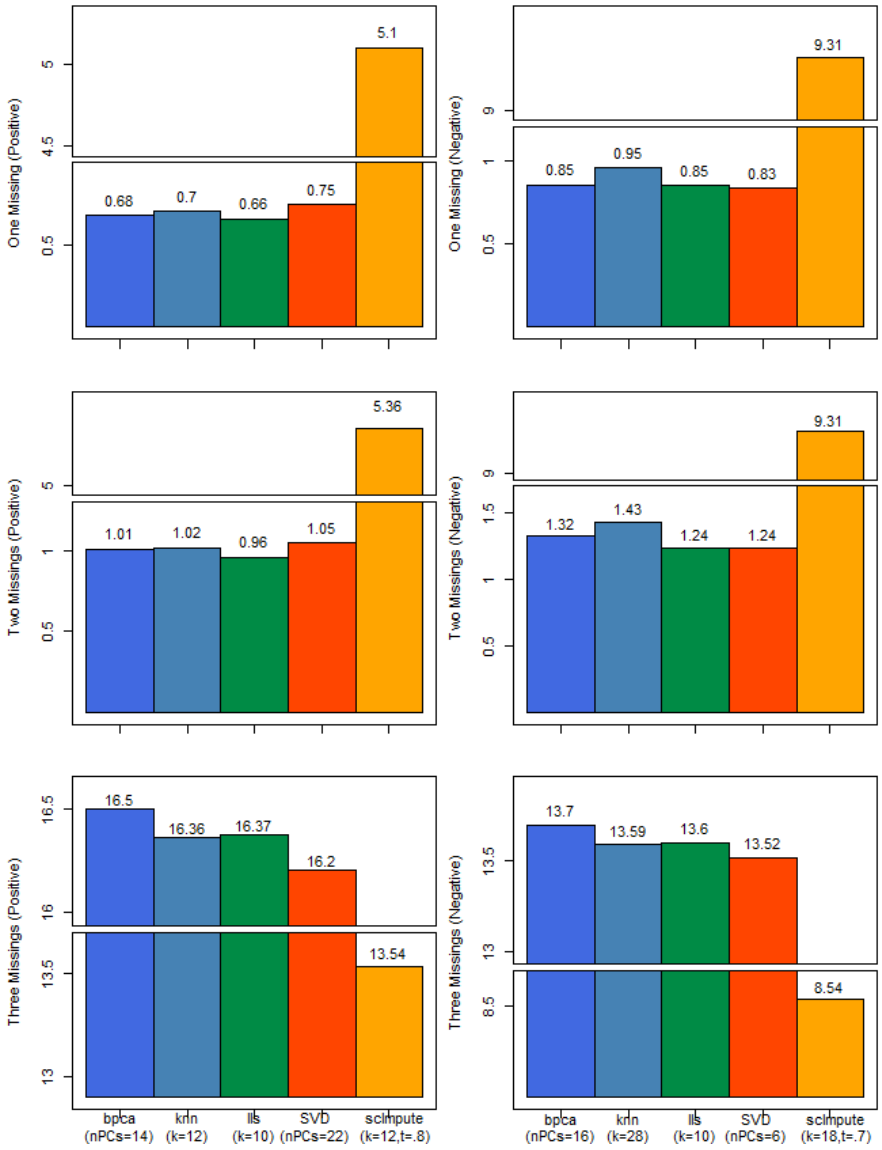

Figure 3. Imputation error vs. number of missings.

4. Conclusion and Discussion

Overall, the imputation accuracy of SVD, LLS, KNN and BPCA are close, shown by both LRMSE and scatterplots between imputed value and reference value on the log scale. The performance of scImpute varies between observations with different numbers of missingness due

to its feature of identifying true zeros in the dataset. The performance of scImpute is better than the other methods on true zeros. However, the accuracy of scImpute to identify the true zeros isn't ideal, with the false discovery rate 66.3% for the positive data set and 60% for the negative data set.

While scImpute has the strength of distinguishing some true zeros from missing values, the imputation method does not give satisfying imputation results with regard to those observations where neither the reference value nor the imputed value is zero. In such cases, scImpute does attempt to impute the values, yet the imputed values tend to be smaller than reference values. One possible solution is for scImpute to keep its feature of finding true zeros while imputing the rest of the missingness using other well-developed methods discussed in this thesis, e.g. KNN.

In this study, we used a data set with triplet measures for each sample. We assumed that the probability of one metabolic feature to be missing in all three measures are low, thus missing in all three measures is non-random and should have been induced by the true absence of that metabolomic feature, or some other mechanism such as ion suppression. On the other hand, we found that for those metabolic features that are not missing in all three measurements, their chance of missing is associated with their value rather than totally random. Thus, further development for identification of true zeros, as well as the relation between missingness and metabolic feature abundance is needed.

5. Reference

Antignac, J.-P., de Wasch, K., Monteau, F., De Brabander, H., Andre, F., & Le Bizec, B. (2005). The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. *Analytica Chimica Acta, 529*(1), 129-136. Retrieved from http://www.sciencedirect.com/science/article/pii/S0003267004011158. doi:https://doi.org/10.1016/j.aca.2004.08.055

Armitage, E. G., Godzien, J., Alonso-Herranz, V., López-Gonzálvez, Á., & Barbas, C. (2015). Missing value imputation strategies for metabolomics data. *36*(24), 3050-3060. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/elps.201500352. doi:doi:10.1002/elps.201500352

Bloszies, C. S., & Fiehn, O. (2018). Using untargeted metabolomics for detecting exposome compounds. *Current Opinion in Toxicology, 8*, 87-92. Retrieved from http://www.sciencedirect.com/science/article/pii/S2468202017301389. doi:https://doi.org/10.1016/j.cotox.2018.03.002

Botstein, D., Sherlock, G., Cantor, M., Troyanskaya, O., Brown, P., Tibshirani, R., . . . Hastie, T. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520-525. Retrieved from https://dx.doi.org/10.1093/bioinformatics/17.6.520. doi:10.1093/bioinformatics/17.6.520 %J Bioinformatics

Brock, G. N., Shaffer, J. R., Blakesley, R. E., Lotz, M. J., & Tseng, G. C. (2008). Which missing value imputation method to use in expression profiles: a comparative study and two selection schemes. *BMC Bioinformatics, 9*(1), 12. Retrieved from https://doi.org/10.1186/1471-2105-9-12. doi:10.1186/1471-2105-9-12

Golub, G. H., Park, H., & Kim, H. (2004). Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics, 21*(2), 187-198. Retrieved

from https://dx.doi.org/10.1093/bioinformatics/bth499. doi:10.1093/bioinformatics/bth499 %J Bioinformatics

Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research, 15*(4), 1116-1125. Retrieved from https://doi.org/10.1021/acs.jproteome.5b00981. doi:10.1021/acs.jproteome.5b00981

Li, W. V., & Li, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications, 9*(1), 997. Retrieved from https://doi.org/10.1038/s41467-018-03405-7. doi:10.1038/s41467-018-03405-7

Madji Hounoum, B., Blasco, H., Emond, P., & Mavel, S. (2016). Liquid chromatography–high-resolution mass spectrometry-based cell metabolomics: Experimental design, recommendations, and applications. *TrAC Trends in Analytical Chemistry, 75*, 118-128. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165993615002575. doi:https://doi.org/10.1016/j.trac.2015.08.003

National Academies of Sciences, E., & Medicine. (2016). *Use of Metabolomics to Advance Research on Environmental Exposures and the Human Exposome: Workshop in Brief*. Washington, DC: The National Academies Press.

Oba, S., Sato, M.-a., Takemasa, I., Monden, M., Matsubara, K.-i., & Ishii, S. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics, 19*(16), 2088-2096. Retrieved from https://dx.doi.org/10.1093/bioinformatics/btg287. doi:10.1093/bioinformatics/btg287 %J Bioinformatics

Rodrigues, D., Pinto, J., Araújo, A. M., Jerónimo, C., Henrique, R., Bastos, M. d. L., . . . Carvalho, M. (2019). GC-MS Metabolomics Reveals Distinct Profiles of Low- and High-Grade Bladder Cancer Cultured Cells. *9*(1), 18. Retrieved from http://www.mdpi.com/2218-1989/9/1/18.

Rothwell, J. A., Loftfield, E., Wedekind, R., Freedman, N., Kambanis, C., Scalbert, A., & Sinha, R. (2019). A Metabolomic Study of the Variability of the Chemical Composition of Commonly Consumed Coffee Brews. *9*(1), 17. Retrieved from http://www.mdpi.com/2218-1989/9/1/17.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic Principal Component Analysis. *61*(3), 611-622. Retrieved from https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00196. doi:doi:10.1111/1467-9868.00196

Turi, K. N., Romick-Rosendale, L., Ryckman, K. K., & Hartert, T. V. (2018). A review of metabolomics approaches and their application in identifying causal pathways of childhood asthma. *Journal of Allergy and Clinical Immunology, 141*(4), 1191-1201. Retrieved from https://doi.org/10.1016/j.jaci.2017.04.021. doi:10.1016/j.jaci.2017.04.021

Webb-Robertson, B.-J. M., Wiberg, H. K., Matzke, M. M., Brown, J. N., Wang, J., McDermott, J. E., . . . Waters, K. M. (2015). Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *Journal of Proteome Research, 14*(5), 1993-2001. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25855118. doi:10.1021/pr501138h