



Recommended strategies for spectral processing and post-processing of 1D ¹H-NMR data of biofluids with a particular focus on urine

Abdul-Hamid Emwas¹ · Edoardo Saccenti² · Xin Gao³ · Ryan T. McKay⁴ · Vitor A. P. Martins dos Santos² · Raja Roy⁵ · David S. Wishart⁶

Received: 12 August 2017 / Accepted: 9 January 2018 / Published online: 12 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

¹H NMR spectra from urine can yield information-rich data sets that offer important insights into many biological and biochemical phenomena. However, the quality and utility of these insights can be profoundly affected by how the NMR spectra are processed and interpreted. For instance, if the NMR spectra are incorrectly referenced or inconsistently aligned, the identification of many compounds will be incorrect. If the NMR spectra are mis-phased or if the baseline correction is flawed, the estimated concentrations of many compounds will be systematically biased. Furthermore, because NMR permits the measurement of concentrations spanning up to five orders of magnitude, several problems can arise with data analysis. For instance, signals originating from the most abundant metabolites may prove to be the least biologically relevant while signals arising from the least abundant metabolites may prove to be the most important but hardest to accurately and precisely measure. As a result, a number of data processing techniques such as scaling, transformation and normalization are often required to address these issues. Therefore, proper processing of NMR data is a critical step to correctly extract useful information in any NMR-based metabolomic study. In this review we highlight the significance, advantages and disadvantages of different NMR spectral processing steps that are common to most NMR-based metabolomic studies of urine. These include: chemical shift referencing, phase and baseline correction, spectral alignment, spectral binning, scaling and normalization. We also provide a set of recommendations for best practices regarding spectral and data processing for NMR-based metabolomic studies of biofluids, with a particular focus on urine.

Keywords Spectral processing · NMR spectroscopy · Metabolomics · Data post-processing · Baseline correction · Spectral alignment · Spectral binning · Normalization · Scaling · Urine

✉ Abdul-Hamid Emwas
abdelhamid.emwas@kaust.edu.sa

Edoardo Saccenti
esaccenti@gmail.com

Xin Gao
xin.gao@kaust.edu.sa

Ryan T. McKay
ryan.mckay@ualberta.ca

Vitor A. P. Martins dos Santos
vitor.martinsdosantos@wur.nl

Raja Roy
roy@cbmr.res.in

David S. Wishart
dwishart@ualberta.ca

² Laboratory of Systems and Synthetic Biology,
Wageningen University and Research, Stippeneng 4,
6708 WE Wageningen, The Netherlands

³ Computer, Electrical and Mathematical Sciences
and Engineering Division, Computational Bioscience
Research Center, King Abdullah University
of Science and Technology (KAUST), Thuwal 23955,
Kingdom of Saudi Arabia

⁴ Department of Chemistry, University of Alberta, Edmonton,
Canada

⁵ Centre of Biomedical Research, Formerly, Centre
of Biomedical Magnetic Resonance, Sanjay Gandhi
Post-Graduate Institute of Medical Sciences Campus,
Lucknow, India

⁶ Department of Biological Sciences, University of Alberta,
Edmonton, Canada

¹ Imaging and Characterization Core Lab, KAUST,
Thuwal 23955-6900, Kingdom of Saudi Arabia

1 Introduction

NMR has played an important role in the development and the continuing advances in metabolomics over the past two decades. Indeed, the very first metabolomics papers were based on NMR spectral analysis of biofluids, such as urine (Serkova et al. 2005; Bertram et al. 2006; Gibney et al. 2005; Beckonert et al. 2007b; Bales et al. 1986). Even today there are more than 600 papers published each year that describe the use of NMR in metabolomics studies. Continuing improvements in NMR technology, such as increased magnet field strength (> 1 GHz) (Cousin et al. 2016; Tkac et al. 2009; Abdul-Hamid M.; Emwas et al. 2013), cryogenically cooled probe technology (Keun et al. 2002), microprobe design advances (Miao et al. 2015; Nagato et al. 2015; Grimes and O'Connell 2011) and dynamic nuclear polarization (Emwas et al. 2008; Ludwig et al. 2010) have significantly improved the sensitivity of NMR for metabolomics applications. Now samples as small as 50 μL are being handled and nanomolar concentrations are now detectable. Despite not being quite as sensitive as MS-based metabolomics (Grison et al. 2016; Zhao et al. 2016; Emwas and Kharbatia 2015; Emwas 2015), NMR spectroscopy has several advantages. In particular, NMR requires: (1) little sample preparation; (2) no prior chromatographic separation and (3) no chemical derivatization. Furthermore, as an analytical technique NMR is robust and highly reproducible, it can be absolutely quantitative, it can be used in the precise structural determination of unknown metabolites, and it can be almost fully automated (Emwas 2015; Gonzalez-Gil et al. 2015; Li et al. 2016).

On the other hand, NMR spectroscopy itself and the analysis of complex biological mixtures by NMR is not trivial (Tiziani et al. 2008; Hajjar et al. 2017). In particular, the ^1H NMR spectra of samples such as urine are very complex, typically consisting of > 1000 detectable and often overlapping peaks. The position, intensity and spectral width of these peaks is highly dependent on the number and types of chemicals in the mixture, the corresponding spin-coupling patterns of those chemicals and a wide variety of sample parameters. These parameters include: sample pH, sample salt type and salt concentrations, dissolved oxygen content, the presence of paramagnetic ions, the choice of solvent(s), temperature, temperature gradients, spectrometer field homogeneity, and primary magnetic field strength (to name just a few). In addition to the sample characteristics, NMR setup and processing parameters can also have a significant impact on the quality of NMR spectra and their subsequent interpretation. The choice of the pulse sequence for data acquisition, the selection of an appropriate solvent suppression technique,

the level of decoupling power, the type of chemical shift reference(s), the length of the 90° pulse, the number of data points collected, the repetition time, receiver gain, the quality of shimming, the quality of tuning, and the number of acquisitions will all have a significant impact on the quality of NMR spectra and the presence of peak distortions or anomalies. Similarly, spectral processing choices concerning the extent of zero filling, choice of digital filters, selection of apodization functions, precision of the chemical shift referencing protocol, accuracy of the phasing, and the quality of baseline correction will also affect the results. Detailed suggestions and recommendations for handling many of these parameters, especially for NMR-based studies of urine, have been given in several recent reviews (Emwas 2015; Emwas et al. 2016).

Using these consensus recommendations, it should now be possible for almost anyone with a high-field NMR instrument to collect and generate (automatically or semi-automatically) high quality 1D ^1H spectral data from complex biofluids. However, there is still relatively little consensus in the community regarding what to do after the NMR spectra are collected—i.e. the post-processing steps. Two “camps” have emerged in the field of NMR-based metabolomics. One camp tends to use spectral deconvolution software to identify and quantify compounds in individual NMR spectra. In this approach, each NMR spectrum is analysed individually and the resulting compound IDs and concentrations from multiple spectra are compiled to create a data matrix for multivariate statistical analysis. A variety of software tools for NMR spectral deconvolution have been developed including the Chenomx NMR Suite (Mercier et al. 2011), Bruker's AMIX (Czaplicki and Ponthus 1998), Bruker's JuiceScreener (Monakhova et al. 2014) and WineScreener (Spraul et al. 2015), Batman (Hao et al. 2014), and Bayesil (Ravanbakhsh et al. 2015).

The second camp uses statistical approaches to initially align multiple NMR spectra, to scale or normalize the aligned spectra, and then to identify interesting spectral regions (e.g. binning) or peaks that differentiate cases from controls (Smith et al. 2009; Barton et al. 2008; Lindon et al. 2007; Beckonert et al. 2007a). This approach, which is often called statistical spectroscopy, performs compound identification or quantification only after the most interesting peaks have been identified. This final identification step may use spectral deconvolution, compound spike-in methods or peak look-up tables (Martinez-Arranz et al. 2015). A variety of software packages for NMR statistical spectroscopy have been developed including, MetAssimulo (Muncey et al. 2010), Automics (Wang et al. 2009), Statistical total correlation spectroscopy (Cloarec et al. 2005a, b), and MVAPACK (Worley and Powers 2014).

For relatively simple biofluids such as serum, plasma, cerebrospinal fluid (CSF), fecal water, juice, wine or beer, NMR

spectral deconvolution approaches appear to work very well (Ravanbakhsh et al. 2015). Extensive spectral libraries now exist for many of these biofluids and a number of the deconvolution software tools are becoming almost fully automated. Indeed, some software packages can be extremely fast and robust with compound coverage easily exceeding 90% and compound quantification errors often below 10% (Worley and Powers 2014; Zheng et al. 2011; Hao et al. 2014; Mercier et al. 2011; Ravanbakhsh et al. 2015). On the other hand, for very complex biofluids such as cell growth media, cell lysates and urine, the corresponding NMR spectra are often too complex for spectral deconvolution (manual or automated). The compound coverage rarely exceeds 50% and the level/quality is highly dependent on the skill and/or experience of the operator. There are also several reports showing considerable discrepancies between different laboratories (Sokolenko et al. 2013) or different users when spectral deconvolution is applied to very complex biofluids. As a general rule, for the routine analysis of urine 1D ¹H NMR spectra, statistical spectroscopy techniques presently appear to be the best option. These approaches are robust and they allow useful results to be obtained with relatively little manual effort. They also facilitate the identification and quantification of key compounds or features in NMR-based urine metabolomic studies.

The purpose of this review is to assess and provide consensus recommendations for the processing of NMR data of biofluids with a particular focus on urine. NMR data processing refers to both spectral processing and data processing, as summarized in Fig. 1. In particular, we will review and discuss consensus recommendations for spectral processing, namely chemical shift referencing, phasing and baseline correction. These steps are critical for generating high quality NMR data. The remainder of this review will focus on providing recommendations for “post processing” of NMR data, including the determination of interesting spectral regions (alignment and binning) as well as spectral normalization, scaling and transformation. These are critical steps to statistical spectroscopy and their correct implementation is essential to the successful NMR analysis of urine (and other biofluid) samples.

2 Spectral processing

2.1 Chemical shift referencing

As any good NMR spectroscopist knows, NMR spectra must always be properly referenced using an internal chemical shift standard (Emwas 2015; Emwas et al. 2016; Harris et al. 2008a, b; Nowick et al. 2003). Chemical shift referencing is important for compound identification, for peak alignment and any multivariate statistical analyses that

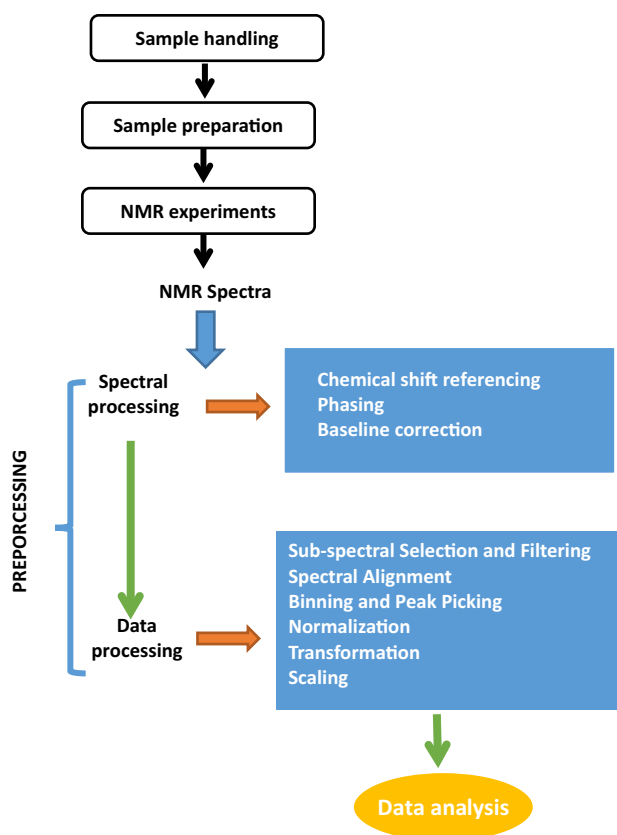


Fig. 1 Summary of spectral processing and post-processing steps on urinary NMR-data

may follow Fig. 2. Within the metabolomics community both 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (TSP) are widely used as chemical shift reference standards (Donaa et al. 2016). However, it is important to note that TSP is actually quite pH sensitive (Wishart et al. 1995).

This pH sensitivity can wreak havoc with spectral alignment, especially if samples have not been well buffered and/or carefully pH corrected. Therefore, we strongly recommend the use of DSS (especially deuterated DSS) as the chemical shift reference standard for biofluid (esp. urinary) NMR spectroscopy. We note that DSS is the chemical shift standard recommended by the IUPAC, IUPAB and IUBMB for biomolecular NMR (Markley et al. 1998). Chemical shift standards, such as DSS, can also be used for quantification, especially if the reference compound concentration is known precisely (Mercier et al. 2011). However, in biofluids such as plasma or serum, where DSS or TSP may become bound to macromolecules (proteins or lipoproteins), random variations in the reference intensity may occur, leading to inaccurate concentration estimates (Pearce et al. 2008). In these cases, an alternative internal standard for

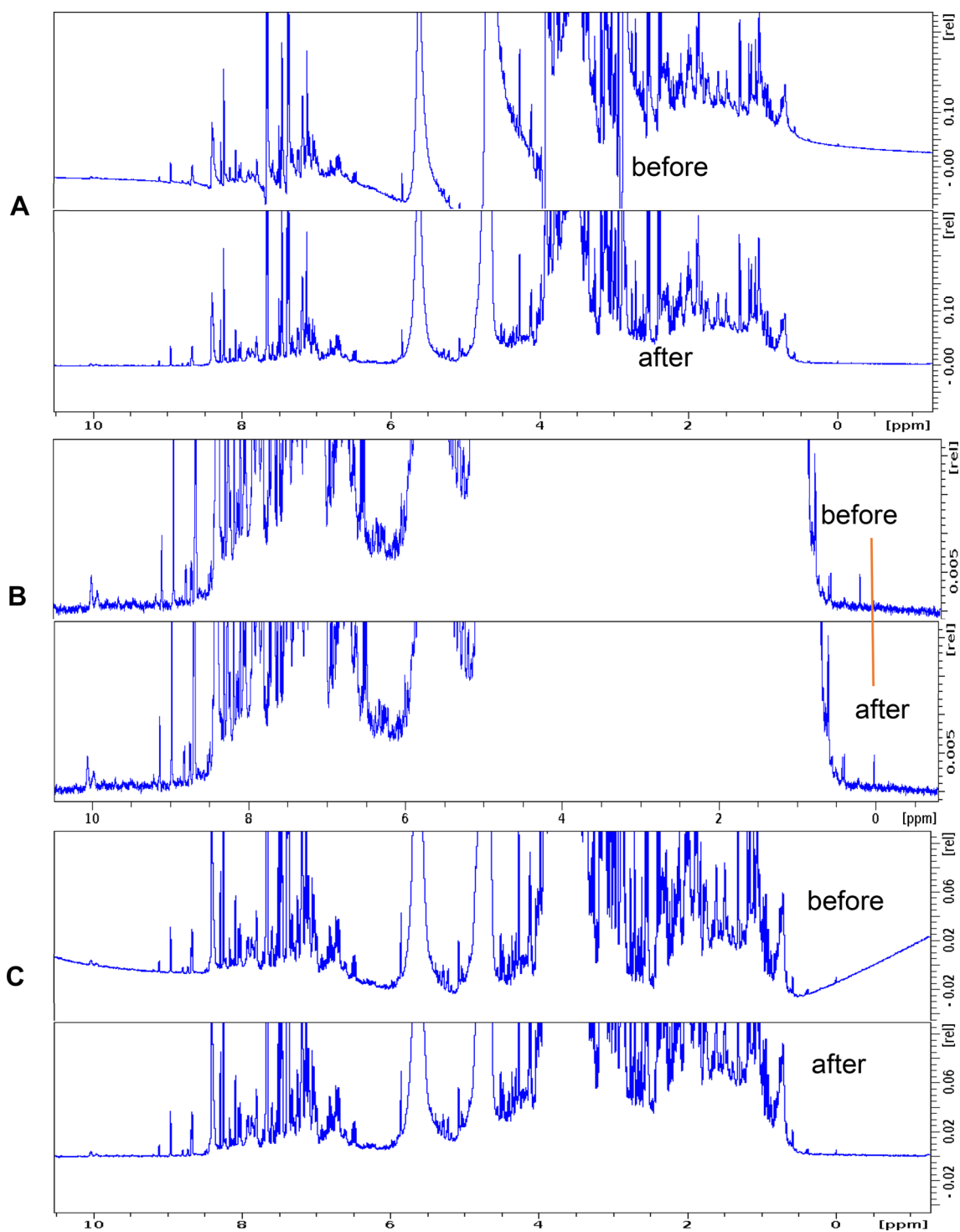


Fig. 2 A simple visualization of the effects of **a** phasing, **b** referencing and **c** baseline correction and on an NMR spectrum

quantification (such as sodium acetate or sodium formate) is recommended. The use of the solvent water peak (i.e. H_2O , and HDO in rapid exchange with non-observed D_2O) for chemical shift referencing is very strongly discouraged since the signal position is sensitive to a wide

variety of sample parameters, including temperature, pH, exchangeable moieties, salts and demagnetization field effects (Edzes 1990; Levitt 1996).

2.2 Phasing

Phasing is a NMR spectral adjustment process that is intended to maximize the absorptive character and the symmetry of all NMR peaks over all regions of an NMR spectrum. Phasing is one of the most important steps in spectral processing as even small phasing errors can lead to significant problems that will ripple down through all remaining spectral processing and post-processing steps Fig. 2. In particular, phasing errors can affect spectral alignment, spectral binning and the measured peak areas (Wishart 2008). Even though automatic phasing is available in most modern NMR spectrometers, manual phasing is often required in metabolomics studies since many auto-phasing routines will distort low-intensity peaks. Phasing is particularly important for handling the residual (but often still prominent) water signal. A phase distortion in the solvent signal can substantially perturb the surrounding regions (~4.7 ppm). Auto-phasing programs may sometimes distort the entire NMR spectrum while attempting to correct for the residual solvent signal. Exclusion of the solvent region from auto-phasing procedures may help reduce this problem, however, manual phasing generally gives better results. Despite these caveats, auto-phasing is still widely used in the metabolomics community. This is because it is fast (allowing greater throughput) and it avoids operator bias.

We recommend that auto-phasing should be used as an initial phasing step. Subsequently, all NMR spectra should be manually inspected for phase distortions and, if necessary, those spectra exhibiting phase distortions should be phased manually. During manual phasing, the vertical scale should be increased as much as possible to allow for proper adjustment of the smaller signals. Even when manual phasing is performed by an experienced operator there are still some cases where it fails to improve spectral quality. Errors in executing or optimizing pulse sequence parameters can be manifested in some “phase-recalcitrant” spectra. The only way to correct for these problems is to re-acquire the spectrum using a standardized pulse sequence and using correct instrument parameters. Careful testing of a new pulse sequence’s performance on known, standardized samples (e.g. DSS with 90% H₂O/10% D₂O with several known small molecules in various spectral regions) is often necessary to ensure that any undetected or phase-distorting pulse-sequence errors will not propagate into the NMR spectra collected for “real” biofluids. In many cases, timing errors in the pulse sequence and/or instrument delays not properly taken into account are the main culprits leading to phase-recalcitrant spectra. These can be difficult to track down, but it is essential that they be detected and dealt with prior to acquiring a large number of spectra.

2.3 Baseline correction

Baseline correction is another spectral processing technique that is critical for removing spectral artefacts that can arise from electronic distortions, inadequate digital filtering or incomplete digital sampling. When properly done, baseline correction yields a more pleasant looking spectrum where signal-free regions are completely flat, horizontal lines with zero intensity Fig. 2. While baseline correction is trivial for simple spectra with just a few peaks, it is somewhat more difficult for NMR spectra containing thousands of peaks with large differences in intensities (as is seen in urine). Correct baselines are critical for proper spectral alignment and proper peak integration (i.e. relative and absolute quantification). Small errors in the baseline structure can easily lead to errors (by orders of magnitude) in the quantification of low abundance metabolites. We recommend that all NMR spectra should be manually inspected for baseline distortions and, if necessary, those spectra exhibiting baseline distortions should be corrected using high quality baseline correction software.

Baseline correction in NMR is normally done via semi-automatic approaches that involve manual identification of reliable baseline regions followed by a computer-generated spline fit. Just as with phasing, baseline correction requires that the vertical scale should be increased as much as possible to allow for proper detection of those baseline regions needing correction. Software from all the major NMR vendors along with many third party software packages, such as NMRPipe (Delaglio et al. 1995), Chenomx NMR Suite (Mercier et al. 2011), or MestreLab Inc.’s MNova (to name just a few), can perform high quality baseline correction. All of these packages work in a semi-automated fashion, meaning that the baseline regions are first identified manually and then the programs complete the remaining baseline correction process. This correction process may use either time domain methods or frequency domain methods (Xi and Rocke 2008; Marion and Bax 1988; Halamek et al. 1994; Bao et al. 2012; Golotvin and Williams 2000; Wang et al. 2013; Bartels et al. 1995). We recommend the frequency domain correction methods as they are more widely used. Frequency domain methods attempt to construct a new baseline curve within the processed spectra directly using techniques such as asymmetric least squares (Peng et al. 2010; Eilers 2003), regular polynomial fitting or spline curve fitting and iterative polynomial fitting with automatic thresholding (Feng et al. 2006). More recently, a parametric approach that employs weighted scatter plot smoothing (LOWESS) has been used to estimate noise levels and generate more accurate baselines for metabolomic studies (Xi and Rocke 2008).

Fully automated baseline correction has been implemented in certain packages such as Bayesil (Ravanbakhsh

et al. 2015) and MestreLab's MNova suite, but these methods are currently limited to simpler biofluid spectra of serum, plasma, fecal water or cerebrospinal fluid. If and when fully automated methods appear for urine analysis, we would recommend them over manual methods as these automated methods would remove any user bias in baseline correction.

3 Data post-processing

Data post-processing refers to the steps involved in assessing processed NMR spectra prior to the identification and comparison of important peaks and peak intensities. As mentioned in the introduction, NMR spectra of urine (or other very complex biofluids with > 75 detectable metabolites) require some degree of spectral simplification. This simplification can be achieved through several data post-processing steps: (1) sub-spectral selection; (2) spectral alignment; (3) spectral binning to extract peak intensities; (4) scaling and normalization, and finally (5) important peak identification (via multivariate statistics). Together, these approaches allow users to identify and quantify the most informative peaks in a given biofluid or urine NMR spectrum.

3.1 Sub-spectral selection and filtering

Sub-spectral selection is a filtering technique involving the selection of only the interesting regions and discarding the uninformative areas of a given NMR spectrum. In general, not all parts of a recorded NMR spectrum are important for identifying and quantifying metabolites. For instance, in urine, the region between 0.00 and 0.60 ppm can be safely removed before alignment and/or binning since no metabolite signals (except possibly those from vacuum grease and other contaminants) exist in this portion of the spectrum. The water signal region from 4.50 to 4.90 ppm is also commonly excluded, as the residual solvent signal after suppression is not of interest and often interferes with the analysis of other metabolites signals. In urine samples, urea is one of the most highly concentrated metabolites and its peak is relatively close to the water resonance (near 6.00 ppm). Urea's exchangeable protons are significantly affected by most water suppression techniques and so urea's signal intensity changes significantly with the degree or quality of water suppression. Therefore, the urea peak (and the surrounding region, if affected) is normally excluded from further analysis. To summarize, we recommend the removal of the upfield region (0.00–0.60 ppm), the residual water region (~4.50 to 4.9 ppm) and the urea region (5.5–6.1 ppm) when analysing urine NMR spectra.

3.2 Spectral alignment

Spectral alignment is a process that iteratively shifts peak positions in multiple spectra so that the peaks corresponding to the same compounds can be directly overlaid or aligned. Spectral alignment is needed to ensure that the same peaks, from the same compounds, can be compared and quantified across multiple NMR spectra. Signals or peaks that are inconsistently shifted across different NMR spectra, will not be properly matched and subsequent binning steps, scaling steps and multivariate analysis of the binned/scaled intensities will be compromised. While spectral alignment is widely used in NMR spectral analysis, it is also important to remember that alignment can hide important information encoded in chemical shift data, including sample pH, metal ion concentrations, ionic strengths and temperature.

Spectral alignment is trivial for NMR spectra with a small number (< 20) of peaks. However, it is not trivial for NMR spectra with thousands of peaks as is frequently seen for NMR spectra of biofluids such as urine. Even when properly referenced, the chemical shifts of many compounds in urine are often subject to a phenomenon known as chemical shift drift (Giskeodegard et al. 2010; Wu et al. 2006a), which is shown in Fig. 3. Chemical shift drift is an environmental effect that can be due to several factors such as sample pH, ionic strength, changes of temperature, instrumental factors, level of compound dilution and relative concentration of specific ions (Defernez and Colquhoun 2003; Cloarec et al. 2005b). The net result of chemical shift drift is that it is often quite difficult to determine which peaks match to which compounds when comparing one urine spectrum to another. One experimental approach to address chemical shift drift is to precisely control the pH and salt concentration of the sample by adding a strong buffer solution to the sample (pH 7.0, 400 mM phosphate, 20–30% by volume). However, this is often not practical for large numbers of samples and it may not always correct other ionic contributions to chemical shift drift. As a result, several computational methods have been developed to correct the movement of NMR peaks. These are called peak alignment or spectral alignment methods and they include such processes as correlation optimized warping (COW) (Nielsen et al. 1998), fuzzy warping (Wu et al. 2006b), peak alignment by beam search (Forshed et al. 2003; Lee and Woodruff 2004), and interval correlation shifting (icoshift) (Savorani et al. 2010). These methods are known as pairwise alignment techniques because they align each NMR spectrum to a chosen reference NMR spectrum, one by one. The reference spectrum can either be real or virtual and should always be representative for the whole dataset. More details about these spectral alignment algorithms are given below.

COW is an older alignment approach developed in the late 1990s that uses a technique called segment warping

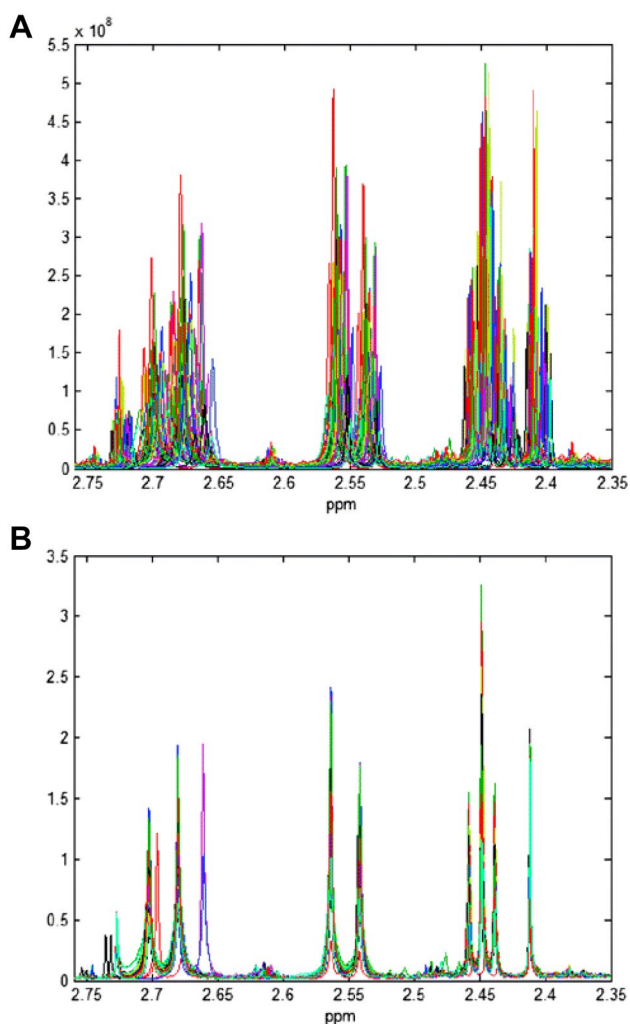


Fig. 3 NMR spectra of urine samples, **a** original spectra in the selected region, and **b** normalized spectra warped to spectrum number 52 in the same region, from (Wu et al. 2006a)

(Tomasi et al. 2004). More specifically, COW is a piecewise or segmented data preprocessing method (where the spectrum is divided into equal sized segments) aimed at aligning a sample spectrum towards a reference spectrum by allowing limited changes in segments lengths on the sample spectrum. This method was originally designed to be used for the alignment of chromatographic data, but it has proven to be useful for the alignment of NMR spectra as well (Tomasi et al. 2004; Smolinska et al. 2012b).

The Beam search method for peak alignment of NMR signals was developed in the early 2000's based on genetic algorithms for optimization (Lee and Woodruff 2004; Forshed et al. 2003). In this method each spectrum is divided into a number of segments then each segment is aligned to a corresponding region in a reference spectrum using a genetic algorithm (Forshed et al. 2002). A smaller part of the spectrum (covering a region spanning ~0.15 ppm)

is aligned to a corresponding reference by shifting (right or left) and then using linear interpolation to adjust the spectra piecewise (Forshed et al. 2003).

Another technique for NMR peak alignment is called the fuzzy warping method which was originally developed and used for the alignment of urine NMR spectra (Wu et al. 2006a). Fuzzy warping seeks to establish a correspondence between the most intense peaks in the spectra to be aligned, where iterative procedures alternate between fuzzy matching and signal transformation. The parameters are weighted according to the corresponding of target spectrum. The performance of the peak alignment can be carried out to assess the alignment procedure in terms of any erroneous alignment or change of peak shape (Wu et al. 2006a).

The interval correlated optimized shifting (icoshift) method is the newest approach to NMR spectral alignment (Savorani et al. 2010). It is based on dividing a given NMR spectrum into different segments or intervals, then aligning the spectral intervals to the corresponding segment of a reference spectrum. Icoshift optimizes the piecewise cross-correlation using a fast Fourier transform (FFT) and a greedy algorithm that allows for user-defined recursion. In particular, each spectrum or interval is allowed to shift right or left until the maximum correlation to the target spectrum is achieved. The use of the FFT approach allows for simultaneous processing and alignment of all spectra. Icoshift has been found to be substantially faster than other algorithms (such as COW, fuzzy warping and beam search) thereby making full-resolution alignment of large 1D ¹H-NMR datasets possible in just a few seconds—even on a desk-top computer. Unlike most other tools, icoshift also allows users to customize peak shape, peak multiplicity, peak position and peak height to better match the target spectrum. Icoshift is available as both an open source MatLab package and a Python package. While icoshift only achieves local alignment optimization and it cannot deal with strongly overlapped regions, the fact that it is open access and substantially faster/better than previously published methods, we recommend that icoshift should be used in the alignment of biofluid (esp. urine) NMR spectra.

Table 1 summarizes the four spectral alignment algorithms discussed above. A much more detailed discussion and assessment of NMR spectral alignment algorithms is provided in a recent review (Vu and Laukens 2013). While icoshift goes a long way towards simplifying and improving the quality of NMR spectral alignment, a fully automated, perfectly functioning NMR spectral alignment tool is still not available. In particular, the problem of peak order changes (Csenki et al. 2007) has yet to be addressed, as all existing alignment methods assume the same peak order between spectra.

Table 1 List of the alignment methods and presents their features

Short name	Full name	Technique	Target function	Peak picking based	NP	SW	Pair-Wise(#)	Correction method	Software	Remarks
COW	Correlation Optimized Warping (Nielsen et al. 1998)	Dynamic programming	Pearson correlation coefficient	No	2	Yes	Yes	Insert and deletion	(1)	Slow
PABS	Peak alignment by Beam search (Lee and Woodruff 2004)	Beam search algorithm	Pearson correlation coefficient	No	3	Yes	Yes	Shift & Insert and deletion	(+)	
FW	Fuzzy Warping (Wu et al. 2006a)	Fuzzy logic for matching most intense peaks	Maximize fuzzy membership Gaussian function	Yes	1	No	Yes	Insert and deletion	(4)	
icoshift	Interval correlation shifting (Savorani et al. 2010)	Segmentation model by equal size segments or manually selecting segments	FFT cross-correlation	No	2	Yes	Yes	Shift & Insert and deletion	(6)	The peak-picking step adds complexity

3.3 Binning and peak picking

The next “post-processing” step is usually some form of binning. Binning can be a very simple method, not even requiring alignment, to extract peak intensities from multiple NMR spectra prior to performing multivariate statistical analysis. Binning involves dividing NMR spectra into small regions (typically spanning 0.04–0.05 ppm), which are sufficiently wide to include one or more NMR peaks. The intensity of each bin is determined by calculating the area under the curve (AUC). As a result, a typical urine NMR spectrum will often generate 500–1000 bins with non-zero intensities. Multivariate statistical analysis is then carried out on the extracted bin intensities and the most significant peaks (or bins) are then assigned to specific metabolites. Binning can be done using prior knowledge (i.e. knowing where metabolite peaks appear) or naively using an automatic algorithm.

Table 2 describes a number of common binning techniques including equidistant (equal size) binning (Izquierdo-Garcia et al. 2011), Gaussian binning (Anderson et al. 2008), adaptive-intelligent binning (De Meyer et al. 2008), dynamic adaptive binning (Anderson et al. 2011), adaptive binning using wavelet transforms (Davis et al. 2007) and an optimized bucketing algorithm (Sousa et al. 2013). Equidistant binning takes a spectrum and then divides it into equal spectral widths (i.e. 0.02, 0.04 or 0.05 ppm) and is the most commonly used binning method (Craig et al. 2006; De Meyer et al. 2010; Izquierdo-Garcia et al. 2011). However, a disadvantage of this method is the lack of flexibility with regard to boundaries in cases where peaks are split between two adjacent bins. Other methods such as adaptive-intelligent binning (De Meyer et al. 2008), dynamic adaptive binning (Anderson et al. 2011) and adaptive binning using wavelet transforms (Davis et al. 2007) can be utilized to overcome this problem by adjusting the bin position so that one bin can only cover complete peaks. We cannot recommend a single binning method because all of them have pros and cons, and their efficiency is somewhat dataset-dependent. As a general rule, equidistant binning is the most commonly used method (Smolinska et al. 2012a), and often works quite well despite its simplicity.

Several non-binning methods such as spectral deconvolution (Weljie et al. 2006), curve-fitting (Bollard et al. 2005), direct peak fitting (Schuyler et al. 2015), and peak alignment have been developed to overcome the drawbacks to binning. However, these methods are generally best for simpler biofluids (serum, plasma, CSF, saliva) and are not yet suited to handling the spectral complexity of urine.

3.4 Normalization

After NMR peaks have been aligned, identified or binned, and their respective intensities determined, the next step in

Table 2 Summary of binning methods with a brief description of the main features of each method

Method	Description	Remarks	References
Equal size binning	Spectrum is divided in simple rectangular regions of the same size. Each bin span the same number of spectral points	Straightforward and fast to apply. Works quite well despite simplicity. The bins size in ppm needs to be specified (0.04, 0.02, 0.01 ppm the most common choices). Peaks can be splitted across multiple bins	Izquierdo-Garcia et al. (2011)
Gaussian binning	A Gaussian kernel weights the signal contribution relative to distance from bin center, and the overlap between bins is controlled by the kernel standard deviation	Overlapping bins are used. Very robust to peak shifts. Two parameters (not easy to tune): standard deviation and step size that make a trade-off between loose of information and robustness	Anderson et al. (2008)
Adaptive-intelligent binning	Iterative algorithm that uses variable bin sizes adaptively inferred from spectra	No arbitrary parameters, reference spectra, a priori knowledge, or data modifications are required. Low-intensity peaks could be troublesome. Noise regions need to be specified	De Meyer et al. (2008)
Dynamic adaptive binning	Bin boundaries are determined by optimizing an objective function using a dynamic programming strategy. The objective function measures the quality of a bin configuration based on the number of peaks per bin	Ability to create bins containing a single peak. Two main parameters and several other parameters for peaks identifications	Anderson et al. (2011)
Adaptive binning using wavelet transform	Wavelet transforms are used to detect peaks in a reference spectrum. Integration is then performed over these peaks in each of the sample spectra. What constitutes a peak is determined by the amount of smoothing implicit in the wavelet transform	Noise regions are excluded. The amount of smoothing depends on the number of levels of the wavelet transform and can be adjusted according to the data resolution and the shifts expected between samples	Davis et al. (2007)
Optimized bucketing algorithm	A bucketing method that optimizes bucket sizes by setting their boundaries at the local minima determined through the average NMR spectrum	A mathematically simple approach. Two parameters need to be chosen, requiring visual inspection of the result	Sousa et al. (2013)

the post-processing pipeline is to correct for inherent concentration differences. Plasma and serum are examples of biofluids that are under strict physiological control, so the spectra collected from these biofluids (at least for the same organism) can often be compared without further adjustment, normalization or scaling. On the other hand, most other biofluids are not under such strict physiological controls and so corrections for dilution effects must be made, for example urine is certainly subject to substantial metabolite concentration variation. Urine volume varies greatly with fluid intake and it is also affected by many other physiological and pathophysiological factors. More specifically, the concentrations of endogenous metabolites in urine (even from the same individual) can vary by several orders of magnitude (Emwas 2015). Therefore, proper adjustment to accommodate these large intensity/concentration variations is critical. The best approach for doing this is called normalization, a well-known data processing technique that aims to make all samples comparable to each other. Note that normalization can mean different things under different situations. In statistics, normalization means transforming a collection of data so that it is normally distributed (i.e. follows a Gaussian distribution). In clinical science, normalization means multiplying the data by some correction factor to make the values more comparable. In this regard, normalization for clinical scientists is similar to the statistical definition of scaling.

Many approaches for sample normalization of urine have been proposed and reviewed in the literature (see Table 3). As a general rule, sample-to-sample normalization can be divided in two broad categories: physiological (normalization to the urine output relative to creatinine or osmolality) or numerical (i.e. all the others). Fig. 4 shows how metabolite concentration profiles change when different normalization strategies are applied to the data. Physiological normalization generally requires a separate measurement using: (1) an osmometer (or osmality meter) to measure the electrolyte to water balance, (2) a refractometer to measure refractive index (a proxy for specific gravity) or (3) a creatinine test (via direct measurement using an enzyme assay or by NMR analysis/integration of the creatinine peaks). Physiological normalization (especially to creatinine) is how most urine concentrations are reported in the clinical and biochemical literature. Its widespread use in the medical community made it a preferred normalization option in the past. However, normalization to creatinine assumes that creatinine clearance is constant and this may not be true in presence of metabolic dysregulation. Therefore, normalization to creatinine should be used only when significant metabolic dysregulation is not suspected (which is not always the case). Measures of urinary specific gravity and osmolality are not as highly dependent on the state of an individual's metabolic regulation. As a result they are gaining increasing traction

in the urinalysis community (Miller et al. 2004; Edmands et al. 2014; Sauve et al. 2015; Waikar et al. 2010; Tang et al. 2015). Therefore, for physiological normalization of NMR-based urinary metabolomic data we recommend the use of specific gravity over creatinine. However, physiological normalization assumes one is working with real concentration data (μM or mM) and in many cases with NMR-based urine metabolomics, only relative concentration data (i.e. no concentration units) are available.

When physiological normalization is not possible, numerical normalization is a viable alternative and, in some cases, can yield even better normalization results than physiological normalization. There is now a large body of literature covering numerical normalization techniques for urine analysis (see Table 3 for a list of methods, abbreviations, short descriptions and references). Different approaches work better for different situations. Luszczek et al. (2013b) found constant sum (CS), constant sum excluding lactate, glucose, and urea concentrations CS-LGU and total spectral area TSA normalized data appear to correlate well with each other. They also do a good job of representing NMR spectral intensities. probabilistic quotient normalization (PQN) normalized data was found to be moderately correlated with UO and osmolality (OSM) data and not with CS, CS-LGU and total spectral intensity (TSI) normalized data.

Kohl et al. (2012) recently reviewed and compared many of the more advanced numerical normalization methods. In particular, they tested the impact of these normalization methods on data structures and sample classification using NMR data from healthy and autosomal polycystic kidney disease (ADPKD) patients. They found only four methods (Loess, Quantile, Linear and Spline normalization) that were able to perform better than methods without normalization for the detection of differentially expressed metabolites. For the accurate determination of metabolite concentration changes, the same four methods provided the most uniform results for all tested metabolites investigated.

In a sample classification context, Quantile and Spline normalization were found to be the best performing methods. Overall, they found that Quantile normalization outperformed all of the most common normalization methods, but achieved mediocre classification performance for small data sets. The opposite was found for Spline normalization. In contrast, Filzmoser and Walczak (2014) found PQN to outperform other methods and recommended it over other numerical normalization techniques. However, Saccenti (2017) found that PQN did not perform particularly well in discriminant/classification setting (see the results of partial least squares discriminant analysis shown in Table 4).

It is interesting to note that total content normalization, urinary output normalization, internal standard normalization, and probabilistic quotient normalization were originally developed for processing metabolomic data. All of the

Table 3 Characteristics of common normalization procedures. Descriptions are based on references (Hochrein et al. 2015; Kohl et al. 2012; Saccenti 2017)

Method	Abbr.	Description	Remarks	References
Urine output (Elizabeth R. Lusczek et al., 2013b)	UO	Urine output is the volume of urine excreted per hour per kilogram of mass. Normalization is performed by multiplying raw metabolite concentrations by urine output	Total urine output may be difficult to obtain	Lusczek et al. (2013a)
Osmolality	OSM	Osmolality is the concentration of solute particles in the urine. Normalization is carried on by dividing concentration by the osmolality of each sample, returning	It reflects physiological metabolite-concentrating mechanisms in the kidneys Require intensive lab work	Lusczek et al. (2013a)
Normalization to an internal standard	IS	The concentration of each metabolite is divided by the concentration of an internal standard, usually creatinine	Constant excretion of creatinine into urine (creatinine clearance) is assumed to be constant. Exact creatinine concentration may be difficult to estimate from NMR spectra	Jatlow et al. (2003)
Constant sum (1)	SC	Metabolites concentration are divided by the sum of the concentration for all the measured metabolites	Assumes that the concentration of each metabolite is increased by the same amount due to the increased concentration of the urine sample	Lusczek et al. (2013a)
Constant sum (2)	SC-LUG	Same as SC but lactate, glucose, and urea concentrations are excluded	Implies that lactate, glucose, and urea concentrations are highly variable during the experiment	Lusczek et al. (2013a)
Total spectrum area	TSA	The (binned) spectra are summed over the entire spectral area excluding regions containing water and urea resonances	See CS and CS-LUG	Lusczek et al. (2013a)
Probabilistic quotient normalization	PQN	It starts with an integral normalization of each spectrum, followed by the calculation of a reference spectrum (median or baseline). For each variable of interest the quotient of a given test spectrum and the reference spectrum is calculated and the median of all quotients is estimated. All variables of the test spectrum are then divided by the median quotient	It assumes that biologically interesting concentration changes influence only parts of the NMR spectrum, while dilution effects will affect all metabolite signals. Choice of reference spectrum is arbitrary	Dieterle et al. (2006)
Cyclic loess normalization	Loess	The log-transformed ratio of any two spectra is compared to their average feature by feature; then, a normalization curve is fitted using non-linear local regression (loess) and subtracted from the original values	Assumes the presence of non-linear biases, such as intensity-dependent biases, is assumed	Dudoit et al. (2002); Cleveland and Devlin, 1988)
Contrast normalization	Contrast	By means of an orthonormal transformation the matrix into a transformed onto a contrast space. Normalizing curves are fitted similarly to those in Cyclic Loess Normalization, using a robust distance measure based on the Euclidean norm. The contrasts are evened out by a smooth transformation and data are mapped back to the original input space	Cyclic loess normalization	Åstrand (2003)

Table 3 (continued)

Method	Abbr.	Description	Remarks	References
Quantile normalization	Quantile	All spectra are brought to an identical distribution of intensities across features (bins or metabolites). quantile	After QN the vectors of feature intensities consist of the same set of values, however, these values are distributed differently among features	Bolstad et al. (2003)
Linear baseline normalization	Linear	A scaling factor is used to map linearly from each spectrum to the baseline. The scaling factor is computed for each spectrum as the ratio of the mean intensity of the baseline to the mean intensity of the spectrum	It assumes a constant linear relationship between each feature of a given spectrum and the baseline. Baseline is arbitrary: it can be constructed by calculating the median of each feature over all spectra	Bolstad et al. (2003)
Non-linear baseline normalization	Li-Wong	A normalization curve is fitted to map a spectrum to the baseline spectrum (having the median overall intensity) on a scatter plot. The normalization curve is fitted only on non-differentially expressed features which are used for finding the normalizing piecewise linear running median line	It is assumed that features corresponding to unregulated metabolites have similar intensity ranks in two spectra. Possible non-linear relationships between the baseline and the individual spectra are also assumed	Li and Wong (2001)
Cubic-spline normalization	Spline	A baseline spectrum is built by computing the geometric mean of the intensities of each feature over all spectra A set of evenly distributed quantiles is taken from both the target spectrum and the sample spectrum and used to fit a smooth cubic spline. The process is iterated several times shifting the set of quantiles by a small offset each time. Finally a spline function generator uses the generated set of interpolated splines to fit the parameters of a natural cubic spline	The existence of non-linear relationships between baseline and individual spectra are assumed The geometric mean can be substituted by the arithmetic mean for reasons of robustness to negative values	Workman et al. (2002)
Shapiro-Wilk	SW	Features showing high variability in concentration are iteratively removed until mostly nonregulated features remain. These are used as reference features for subsequent data normalization	It assumes that the variance of regulated features across all specimens is larger than that of non-regulated features	Hochrein et al. (2015)
Linear mixed	LM	It fits a mixed model to metabolite concentration with simultaneous estimation of the correlation matrix	Assume data as coming from a larger pool cohorts; same for the batches and samples	Jauhainen et al. (2014)
EigenMS		Estimating treatment effects with an ANOVA model; singular value decomposition of the residuals matrix is then used to determine bias trends in the data. The number of bias trends is then estimated via a permutation test and the effects of the bias trends are eliminated	Not tested on NMR data	Karpivitch et al. (2014)
Variance stabilization normalization	VSN	VSN approaches are set of non-linear methods that used to keep the variance constant over the entire data range, leading to roughly equal variable variance. Found to work well with NMR data	Found to work well with NMR data	Huber (2002)

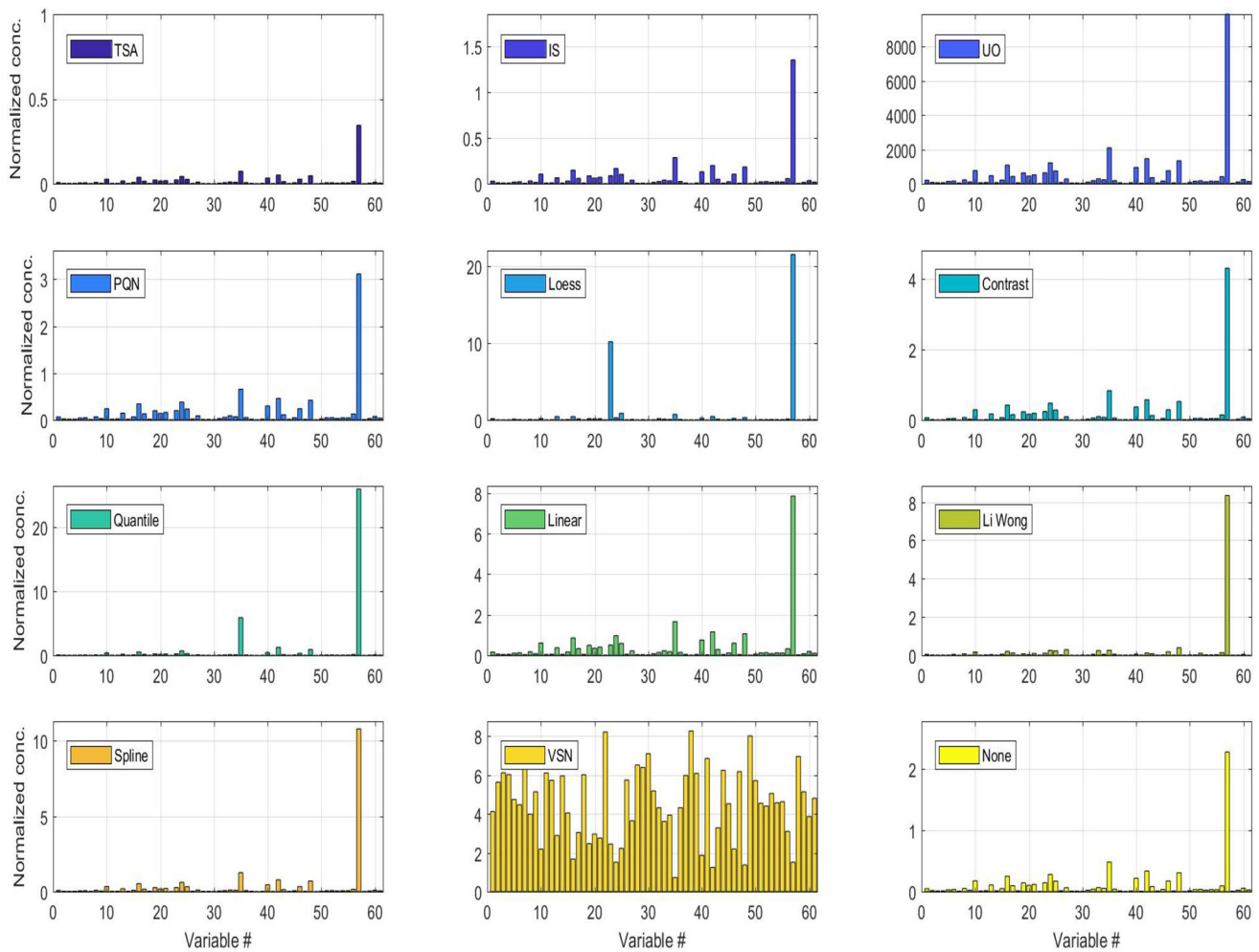


Fig. 4 Metabolite concentrations in a urine sample after different normalization procedures have been applied. For a full description of the methods see Table 3. The data are originally from (Lusczeck et al. 2013b) and were retrieved at <http://www.ebi.ac.uk/metabolights/MTBLS123>

Table 4 Quality of PLS-DA model for the discrimination of two groups when different normalization approaches are applied on the data. The table is reproduced with permission from (Saccenti 2017)

	Method	NMC	Q^2	DQ^2	AUROC
1	TSA	64	0.02	0.02	0.60
2	IS	77	-0.33	0.005	0.37
3	UO	57	-0.04	0.007	0.61
4	PQN	62	-0.05	-0.03	0.64
5	Loess	1	0.89	0.92	1
6	Contrast	23	0.41	0.56	0.99
7	Quantile	7	0.63	0.81	0.99
8	Linear	86	-0.69	-0.02	0.28
9	Li Wong	64	-0.25	-0.03	0.56
10	Spline	7	0.36	0.7	1
11	VSN	0	0.95	0.97	1
12	None	33	0.21	0.25	0.77

NMC Number of misclassifications, DQ^2 discriminant Q^2 , AUROC area under the receiver operating curve. The measures are discussed in (Saccenti 2017)

other methods were developed to normalize microarray data, which have inherently different properties in terms of variance and covariance patterns and error structure. Indeed, the performance of the latter normalization methods on metabolomics data can be quite inconsistent, as observed by a number of different authors (Hochrein et al. 2015; Saccenti 2017).

Many of the numerical methods used for normalization implicitly assume that the average sum of measured metabolite concentrations is constant across samples or group of samples. In other words, it is assumed that the total quantity of dissolved metabolites is invariable. Unfortunately, this is often an unrealistic assumption. In particular, Hochrein et al. (2015) showed that commonly used normalization and scaling methods fail to retrieve true metabolite concentrations in the presence of increasing amounts of glucose added to simulate unbalanced metabolic regulation. They also proposed an alternative method to compensate for these effects in the presence of marked unbalanced metabolic regulation.

All normalization methods alter the structure of the data and the results of subsequent analysis will be affected by the choice of the normalization method applied, especially when the data are used to infer correlations and biological networks as described in (Saccenti 2017). Jauhainen et al. (2014) proposed a method based on linear mixed modelling, and found that it performed well when assessing robustness and its ability to discover true correlations. Figure 5 shows the results of a principal component analysis, which is one of the most commonly used multivariate tools in metabolomics (Table 3), after it has been applied to the data. While this is just one example taken for one particular data set, it clearly illustrates how normalization not only affected the results of this exploratory analysis but also the performance of the methods used to discriminate between groups of samples, which is a typical problem in metabolomics studies.

It is evident from the reported literature that there is no consensus on which numerical method should be applied to normalize data and that a consensus is difficult to establish. Therefore, we are unable to make a formal recommendation

on which numerical normalization method should be used for NMR-based urinary metabolomics. Based on the data at hand, it seems advisable to use PQN when the goal is biomarker selection but when the goal is discrimination/classification Quantile normalization for large (> 50 samples) data sets would seem to perform best, while Spline normalization seems to work better for smaller data sets.

3.5 Scaling and transformation

Scaling and transformation refer to statistical techniques that help to make data more normally distributed or to reduce the spread in values by employing a mathematical operation on the spectral signal intensities (or concentrations) for all samples. As mentioned earlier, urinary metabolite concentrations can range over several orders of magnitude. The detectable variations in metabolites with higher concentrations will of course be easier to detect than the ones with low concentrations. This can lead to a bias or an undue influence from highly concentrated metabolites on the results of

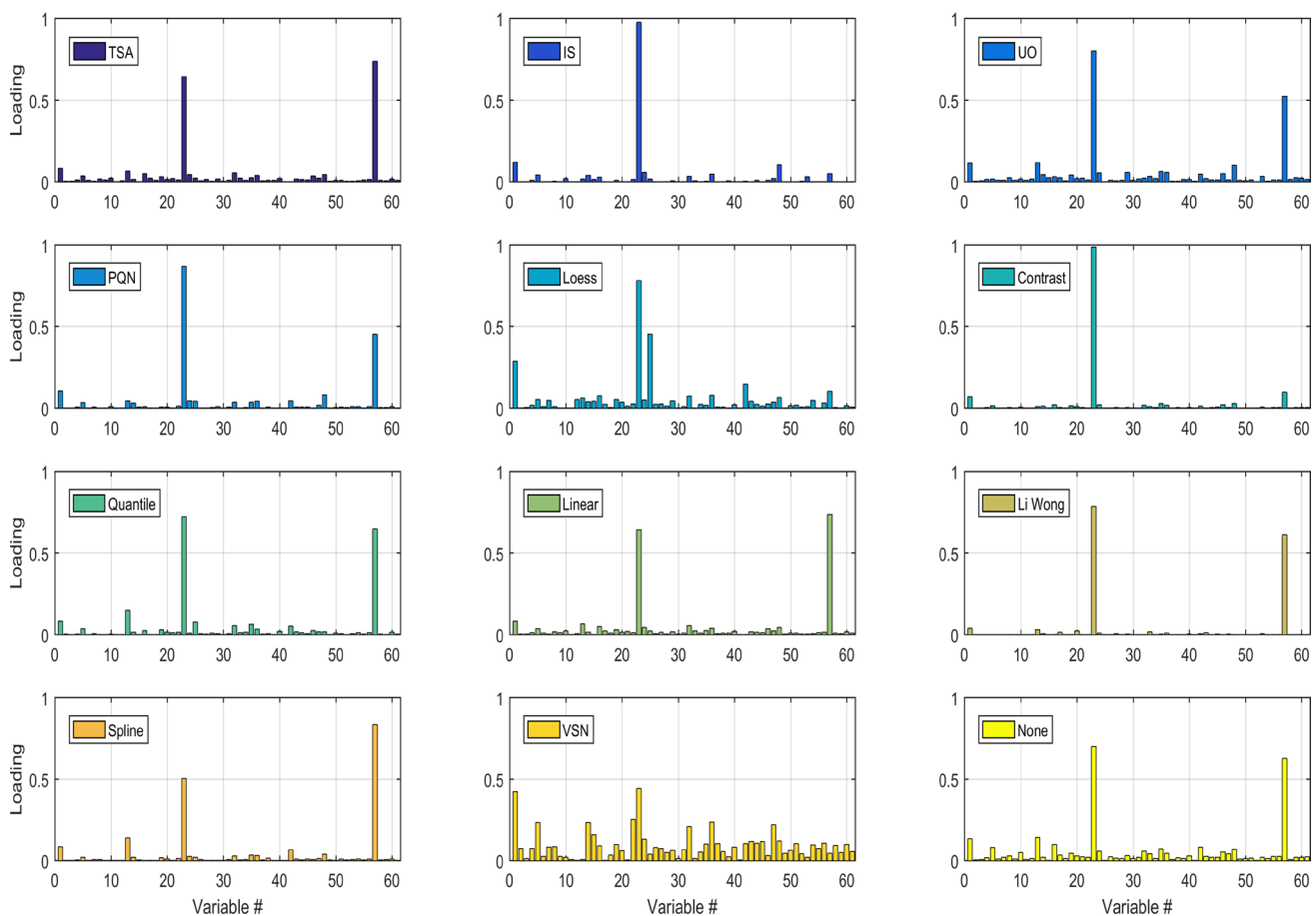


Fig. 5 Loadings for the first principal components for a PCA model fitted on the data normalized with the different procedures. Data are Pareto scaled before PCA.) Data are from (Luszczek et al. 2013b)

and have been retrieved at <http://www.ebi.ac.uk/metabolights/MTBLS123> The figure is from reference (Hochrein et al. 2015)

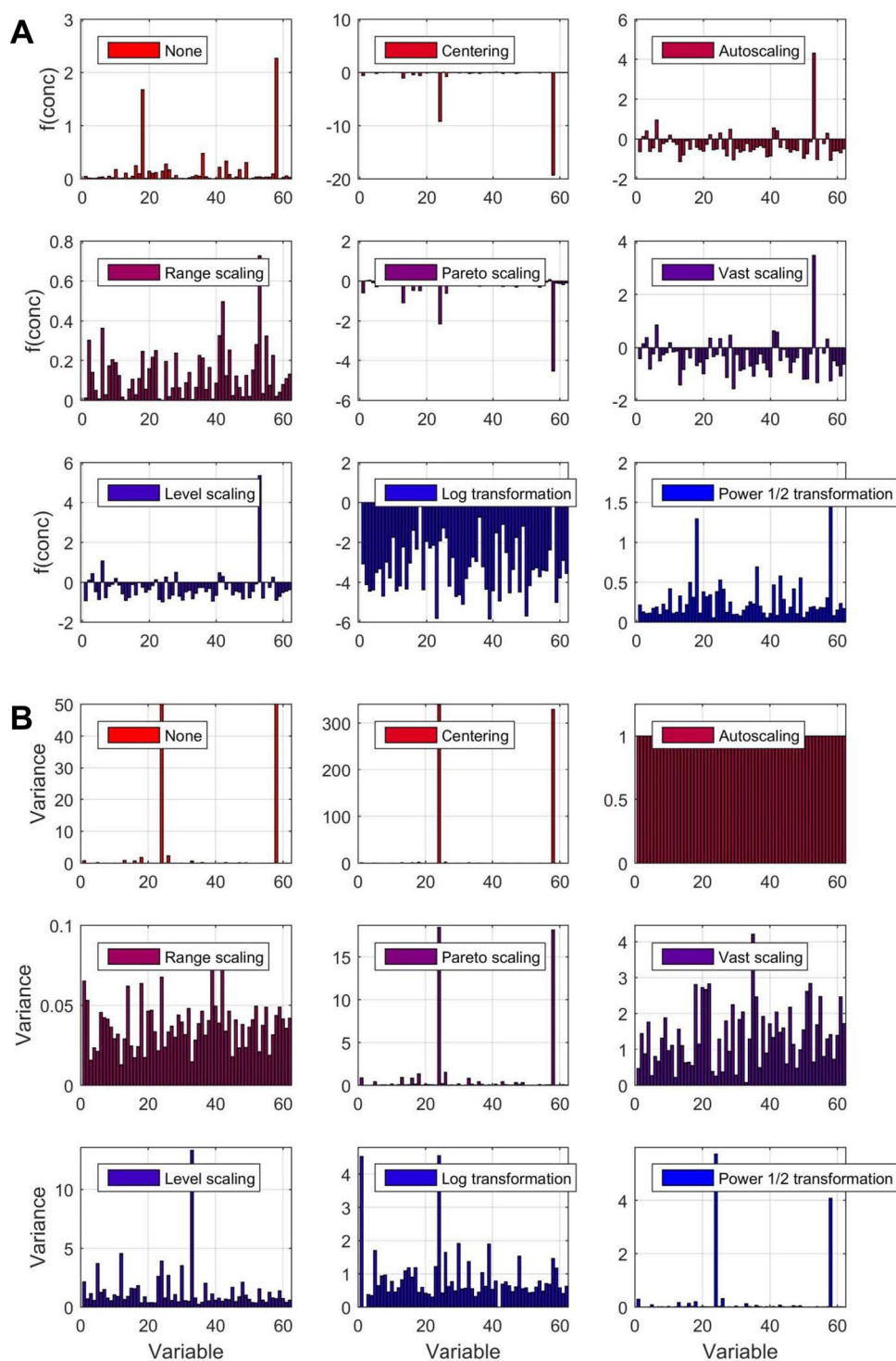
Table 5 Summary of centering, scaling, and transformation methods based on reference (van den Berg et al. 2006), where \bar{x}_i indicate the average and S_i the standard deviation

Method	Description	Drawbacks
Centering $x_{ij} = x_{ij} - \bar{x}_j$	In this method the mean is subtracted from each column. Mean of transformed variance will always be "zero". It is usually applied in most data analysis methods like PCA and PLS regression	In heteroscedastic data, it is not always sufficient
Autoscaling $x_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$	In this method each column is centred by subtracting the mean from each peak intensity then divided by the standard deviation of that column. Makes all variable variances equal; important in PCA and PLS application to avoid that high variance variable dominate the model	All variables become equally important, even noise. This may be a problem in NMR applications where large portions of the spectra contain only noise
Pareto scaling $x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_j}}$	Pareto scaling is similar to autoscaling but each column divided by square root of the standard deviation after mean centring. Good method to reduce the influence of intense peaks while emphasizing the weak peaks. Makes variable variance roughly equal. Well suited for the analysis of NMR data since it downgrade the importance of the noise while preserving the variance structure of the data	Like autoscaling, sensitive to large fold changes but at less extent
Range scaling $x_{ij} = \frac{x_{ij} - \bar{x}_j}{(x_i \max - x_i \min)}$	In this method each column center to the mean, then divided minimum and maximum range of that particular metabolite. Makes variable variance roughly equal; better for explorative analysis	Enhances variables with smaller variability; reduce consistently the variance. Very insensitive when the minimum and maximum range is very large
Level scaling $x_{ij} = \frac{x_{ij} - \bar{x}_j}{x_i}$	In this method each column is centre to the mean then divide by the same mean. Makes variable variance roughly equal while preserving the variance structure. Suitable for biomarker identification	May be problematic for NMR data since it may inflate low value variables
Logarithmic transformation $x_{ij} = \log(x_{ij})$	Correct for heteroscedasticity and non-normality; pseudo scaling effects. Suited for concentration data and multivariate analysis	Enhances small values in the case of noisy NMR data. Not applicable to 0 values; ineffective in bi-modal data distribution
Power transformed $x_{ij} = \sqrt[x_{ij}]{x_{ij}}$	Correct for heteroscedasticity and non-normality; pseudo scaling effects; applicable to 0 values. It is a good alternative to situation when logarithmic scaling is not possible	Choice of power is arbitrary; cannot be applied for negative values
Generalized log transformation $x_{ij} = \ln \left(x_{ij} + \sqrt{\frac{x_{ij}^2 + \lambda}{x_{ij}}} \right)$	Suitable for classification/discrimination applications	May suffer from numerical instability and not suitable in the case of very small values

a urinary metabolomic study (Ebbels et al. 2011). This influence can, in turn, make a small number of metabolites dominate the outcomes from multivariate statistical analyses. To avoid this kind of bias it is often necessary to scale metabolite intensities before undertaking any further analysis (van den Berg et al. 2006). Table 5 shows a list of scaling and transformation methods, several of which were investigated

and compared by (van den Berg et al. 2006). Centering is commonly used to adjust the differences between low-concentration and high-concentration metabolites by scaling all values so that they vary around zero (zero becomes the mean metabolite level). Mean-centering, on its own, is not sufficiently powerful to correct for scaling issues if the data is composed of sub-groups with different variability. As a

Fig. 6 Effect of different centering, scaling and transformation approaches on concentration values (a) and variance (b). For a description of the methods see Table 4. Data are from (Luszczek et al. 2013b) and have been retrieved at <http://www.ebi.ac.uk/metabolights/MTBLS123>



result, mean centering is usually combined with other scaling methods.

These “other” scaling methods include level scaling, range scaling, VAST scaling, Pareto scaling, and autoscaling (Ebbels et al. 2011; Craig et al. 2006). In Fig. 6, we show the effects of several scaling and transformation methods on urine metabolite concentration data. Each scaling method had its own strengths and weaknesses. For example, autoscaling can often increase noise artefacts from spectral regions devoid of usable signals. To address this problem, Pareto scaling uses the square root of the standard deviation instead of the standard deviation as the scaling factor. This increases the sensitivity and reduces noise, while still allowing the data to remain closer to the original measurements (Ebbels et al. 2011). Variable stability scaling (VAST) is another method that weighs each variable according to its measured stability and then down-weights the variables that are less stable. This approach is believed to improve the distinction between different classes in subsequent multivariate analysis (Keun et al. 2003). The advantages of this method were first demonstrated by analysing NMR spectra of urine in an animal model of bilateral nephrectomy (Keun et al. 2003).

Numerical transformations (e.g. power or logarithmic transformation) are another example of scaling or statistical normalization. Transformations are mostly used to correct for *heteroscedasticity* or to correct for data skewness and non-normality before statistical testing. When power and log transformations (or more sophisticated transformation like the Box–Cox’s transformation) are used, large values are more heavily penalized than small values. This provides a pseudo-scaling effect that can be particularly relevant to NMR data as it enhances the importance of small peaks relative to larger ones (Sakia 1992; Kvalheim et al. 1994). Although working on a different context, Feng et al. cautioned against the use of logarithmic transformation noting that the results of standard statistical tests performed on log-transformed data are often not relevant for the original, non-transformed data (Changyong 2014).

The optimal transformation method should be capable of reducing or removing heteroscedastic noise (i.e. variables of sub-group are different than other sub-groups) into homoscedastic information (i.e. variables are similar in sub-groups). These methods are more relevant when reducing non-linear, non-additive, non-normalized or heteroscedastic noise in NMR data and will enhance the information contained in small peaks (Sakia 1992; Kvalheim et al. 1994). For instance, the Box–Cox transformation is a parametric power transformation method used for nonlinear conversion of data where large values are reduced relatively more than the small values (Ebbels et al. 2011; Sakia 1992).

Van den Berg et al. reviewed most of the methods presented in Table 5 using MS data and found that

auto-scaling and range scaling performed better with regard to biological interpretation when data were analysed using PCA. In particular, these two methods were able to remove the dependence of metabolite rank importance in the PCA model from the average concentrations and the magnitude of fold changes. They also found that centering, log transformations, and power transformations, along with level and Pareto scaling showed a strong dependence on concentration and fold changes leading to poorly interpretable PCA results. However, Kohl et al. (2012) found VSN to outperform the latter two methods in a more exploratory setting.

In many situations, high concentration and high variance metabolites may not be the most relevant to the biological problem being studied. However, since most (multivariate) statistical approaches use the information embedded in the variance/covariance matrix, it is crucial that the variance structure of the data is preserved because it contains valuable (biological) information. However, the choice of the scaling methods needs to be tailored on both the application and the data type. NMR and MS data have inherently different properties in term of range and error structure and this may explain the different performance of the same method when applied on different data from different platforms. Depending on the final application, for NMR binned data, Pareto scaling may be the most sensible choice when the aim is data exploration through PCA. In a more discriminant setting, Parsons et al. (2007) found generalised logarithm transformations to significantly improve the discrimination between sample classes yielding higher classification accuracies compared to unscaled, auto-scaled, or Pareto scaled data (Parsons et al. 2007).

Gromski et al. (2015) investigated the effect of autoscaling, range scaling, level scaling, Pareto scaling and VAST scaling on four classification models [principal components-discriminant function analysis (PC-DFA), support vector machines (SVM), random forests (RF) and *k*-nearest neighbours (kNN)] and found that VAST scaling was more stable and robust across all the classifiers considered and advocated for its use.

Our recommendation is that scaling and transformation should be done on all NMR-derived biofluid data prior to conducting multivariate statistical analyses. Visualization and assessment of the scaling and/or transformation effects on the data is necessary to ensure that these scaling or transformation efforts make the data more centred and more Gaussian in its overall distribution (i.e. reducing heteroscedasticity). Researchers must refrain from blindly (i.e. without visualizing the consequences) applying different transformation and scaling methods until the results of the analysis match some predefined hypothesis, as this is scientifically and statistically improper.

3.6 Multivariate statistics, compound identification and biological interpretation

Once all the NMR data has been properly prepared through the careful use of phasing, weighting functions (apodization), zero filling, baseline correction, normalization, and scaling (among other methods described previously and in the referenced materials), then the specialized work of statistical analysis, compound identification and biological interpretation may begin. There are many excellent reviews on how to conduct multivariate statistics with MS or NMR-based metabolomics data (Ren et al. 2015; Emwas et al. 2013; Izquierdo-Garcia et al. 2011) as well as on methods to perform compound identification and biological interpretation from NMR data (Karaman et al. 2016; Donaa et al. 2016; Schleif et al. 2011). It is well beyond the scope of this paper to provide an overview or an assessment of these subjects. However, a few comments or suggestions are perhaps worthwhile.

In the field of NMR-based metabolomics there are a number of well-regarded, freely available software tools and resources that are widely used and which we highly recommend. These include: MetaboAnalyst (Xia et al. 2009, 2015; Xia and Wishart 2010) for multivariate analysis, metabolite annotation and biological interpretation, MVAPACK for multivariate analysis (Worley and Powers 2014), Workflow4Metabolomics for multivariate analysis (Giacomoni et al. 2015), Metassimulo for multivariate analysis (Muncey et al. 2010), the Human Metabolome Database (HMDB) for metabolite annotation and biological interpretation (Wishart et al. 2013), and the BioMagResBank (BMRB) for metabolite identification (Markley et al. 2008). There are also a number of commercial tools such as Chenomx's NMR Suite, Bruker's AMIX software, MestreLab's MNovo and Umetrics SIMCA that offer tools for multivariate analysis and/or metabolite identification. While many researchers prefer to do their own statistical analysis and data interpretation, our recommendation is, for those who are new to metabolomics, that they should collaborate with an individual who has already had significant prior experience in metabolomic data analysis and data interpretation. Alternately, statistical neophytes should dedicate considerable time and effort to become a proficient in this area as possible, prior to embarking on this sort of analysis.

4 Conclusion

The intent of this review was to provide readers with some guidance and recommendations regarding how to process and post-process NMR spectral data collected on biofluids, with a particular focus on urine. The wide disparity in published practices and outcomes from different NMR

metabolomics laboratories led us to investigate existing practices and to systematically assess which methods worked best under which situations. In doing so, we have tried to highlight the advantages and disadvantages of different NMR spectral collection and spectral data processing steps that are common to NMR-based metabolomic studies of biofluids such as urine. More specifically we reviewed the existing literature, assessed the methods in our laboratories and made the following best-practice recommendations:

1. We recommend the use of DSS (especially deuterated DSS) as the chemical shift reference standard for all urinary NMR spectroscopy.
2. We recommend that auto-phasing should be used as an initial phasing step. Subsequently, all biofluid NMR spectra should be manually inspected for phase distortions and, if necessary, those spectra exhibiting phase distortions should be phased manually.
3. We recommend that all biofluid NMR spectra should be manually inspected for baseline distortions and, if necessary, those spectra exhibiting baseline distortions should be corrected using specific high quality baseline correction software (mentioned in this document).
4. For urine NMR spectra we recommend the removal of the upfield region (0.00–0.60 ppm), the residual water region (~4.50–4.9 ppm) and the urea region (5.5–6.1 ppm), especially prior to alignment and binning.
5. We recommend that icoshift should be used in the alignment of biofluid (esp. urine) NMR spectra.
6. No specific recommendation on the best spectral binning method is possible, although equidistant binning appears to be the simplest and fastest approach.
7. When possible, we recommend physiological normalization for NMR-based urinary metabolomic studies, with specific gravity being preferred over creatinine normalization. In situations where physiological normalization is not possible, we recommend Quantile normalization for large (> 50 samples) data sets while Spline normalization is recommended for smaller data sets.
8. We recommend that scaling and transformation should be done on all NMR-derived biofluid data prior to conducting multivariate statistical analyses and subsequent compound identification or biological interpretation. Furthermore, this scaling and transformation must be visualized and assessed by users to determine if the heteroscedasticity has been properly reduced.

Following these recommendations should allow users not only to get consistent, reproducible NMR data but also to optimize the outcome for their multivariate statistical analysis as well as their subsequent final data interpretation.

This review is not intended to be prescriptive. Describing a single protocol that works for all situations is simply not

practical. Indeed, the optimal choice of data processing (and post-processing) options depends on the experiment being conducted, the quality of the data at hand, along with an appreciation of the problem being addressed. For example, if the focus of a study is on exploring differences between groups or subgroups, one should always try to employ a normalization and scaling strategy that will not level out possible differences. If the focus is on data exploration, it is advisable to scale the data in such a way as to avoid using high variance values that will dominate the final model. In all cases, careful experimental preparation prior to any NMR data acquisition, followed by careful, consistent spectral processing and post-processing is necessary before a truly productive NMR data analysis can begin. Otherwise precious time and resources will be wasted on trying to interpret inconsistent data and inaccurate results.

Acknowledgements The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. URF/1/1976-04.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval We declared that, all authors comply with Springer's ethical policies.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson, P. E., Mahle, D. A., Doom, T. E., Reo, N. V., DelRaso, N. J., & Raymer, M. L. (2011). Dynamic adaptive binning: An improved quantification technique for NMR spectroscopic data. *Metabolomics*, 7(2), 179–190. <https://doi.org/10.1007/s11306-010-0242-7>.
- Anderson, P. E., Reo, N. V., DelRaso, N. J., Doom, T. E., & Raymer, M. L. (2008). Gaussian binning: A new kernel-based method for processing NMR spectroscopic data for metabolomics. *Metabolomics*, 4(3), 261–272. <https://doi.org/10.1007/s11306-008-0117-3>.
- Åstrand, M. (2003). Contrast normalization of oligonucleotide arrays. *Journal of Computational Biology*, 10(1), 95–102.
- Bales, J. R., Bell, J. D., Nicholson, J. K., & Sadler, P. J. (1986). 1H NMR studies of urine during fasting: Excretion of ketone bodies and acetylcarnitine. *Magnetic Resonance in Medicine*, 3(6), 849–856. <https://doi.org/10.1002/mrm.1910030605>.
- Bao, Q. J., Feng, J. W., Chen, F., Mao, W. P., Liu, Z., Liu, K. W., et al. (2012). A new automatic baseline correction method based on iterative method. *Journal of Magnetic Resonance*, 218, 35–43. <https://doi.org/10.1016/j.jmr.2012.03.010>.
- Bartels, C., Guntert, P., & Wuthrich, K. (1995). IFLAT - A new automatic baseline-correction method for multidimensional NMR spectra with strong solvent signals. *Journal of Magnetic Resonance Series A*, 117(2), 330–333. <https://doi.org/10.1006/jmra.1995.0780>.
- Barton, R. H., Nicholson, J. K., Elliott, P., & Holmes, E. (2008). High-throughput H-1 NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: Validation study. *International Journal of Epidemiology*, 37, 31–40. <https://doi.org/10.1093/ije/dym284>.
- Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C., et al. (2007a). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols*, 2(11), 2692–2703.
- Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J. G., Holmes, E., Lindon, J. C., et al. (2007b). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*, 2(11), 2692–2703. <https://doi.org/10.1038/nprot.2007.376>.
- Bertram, H. C., Knudsen, K. E. B., Serena, A., Malmendal, A., Nielsen, N. C., Frette, X. C., et al. (2006). NMR-based metabonomic studies reveal changes in the biochemical profile of plasma and urine from pigs fed high-fibre rye bread. *British Journal of Nutrition*, 95(5), 955–962. <https://doi.org/10.1079/bjn20061761>.
- Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K., & Holmes, E. (2005). NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR in Biomedicine*, 18(3), 143–162. <https://doi.org/10.1002/nbm.935>.
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), 185–193.
- Changyong, F. E. N. G., Naiji, H. W., Tian, L. U., Hua, C. H. E. N., Ying, H. E. LU, and Xin, M. T. U. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*, 26(2), 5. <https://doi.org/10.3969/j.issn.1002-0829.2014.02.009>.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83(403), 596–610.
- Cloarec, O., Dumas, M. E., Craig, A., Barton, R. H., Trygg, J., Hudson, J., et al. (2005a). Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Analytical Chemistry*, 77(5), 1282–1289. <https://doi.org/10.1021/ac048630x>.
- Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., et al. (2005b). Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in H-1 NMR spectroscopic metabonomic studies. *Analytical Chemistry*, 77(2), 517–526. <https://doi.org/10.1021/ac048803i>.
- Cousin, S. F., Kaderavek, P., Haddou, B., Charlier, C., Marquardsen, T., Tyburn, J. M., et al. (2016). Recovering Invisible Signals by Two-Field NMR Spectroscopy. *Angewandte Chemie-International Edition*, 55(34), 9886–9889. <https://doi.org/10.1002/anie.201602978>.
- Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*, 78(7), 2262–2267. <https://doi.org/10.1021/ac0519312>.
- Csenki, L., Alm, E., Torgrip, R. J. O., Aberg, K. M., Nord, L. I., Schuppe-Koistinen, I., et al. (2007). Proof of principle of a generalized fuzzy Hough transform approach to peak alignment of

- one-dimensional H-1 NMR data. *Analytical and Bioanalytical Chemistry*, 389(3), 875–885. <https://doi.org/10.1007/s00216-007-1475-9>.
- Czaplicki, J., & Ponthus, C. (1998). A specialized database manager for interpretation of NMR spectra of synthetic glucides: JPD. *Journal De Chimie Physique Et De Physico-Chimie Biologique*, 95(2), 216–220.
- Davis, R. A., Charlton, A. J., Godward, J., Jones, S. A., Harrison, M., & Wilson, J. C. (2007). Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems*, 85(1), 144–154. <https://doi.org/10.1016/j.chemolab.2006.08.014>.
- De Meyer, T., Sinnaeve, D., Van Gasse, B., Rietzschel, E.-R., De Buyzere, M. L., Langlois, M. R., et al. (2010). Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum H-1-NMR spectra. *Analytical and Bioanalytical Chemistry*, 398(4), 1781–1790. <https://doi.org/10.1007/s00216-010-4085-x>.
- De Meyer, T., Sinnaeve, D., Van Gasse, B., Tsiportkova, E., Rietzschel, E. R., De Buyzere, M. L., et al. (2008). NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analytical Chemistry*, 80(10), 3783–3790. <https://doi.org/10.1021/ac7025964>.
- Defernez, M., & Colquhoun, I. J. (2003). Factors affecting the robustness of metabolite fingerprinting using H-1 NMR spectra. *Phytochemistry*, 62(6), 1009–1017. [https://doi.org/10.1016/s0031-9422\(02\)00704-5](https://doi.org/10.1016/s0031-9422(02)00704-5).
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), 277–293. <https://doi.org/10.1007/bf00197809>.
- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Analytical Chemistry*, 78(13), 4281–4290.
- Dona, A. C., Kyriakides, M., Scott, F., Shephard, E. A., Varshavi, D., Veselkov, K., et al. (2016). A guide to the identification of metabolites in NMR-based metabolomics/metabolomics experiments. *Computational and Structural Biotechnology Journal*, 14, 19.
- Dudoit, S., Yang, Y. H., Callow, M. J., & Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, 12, 111–139.
- Ebbels, T. M. D., Lindon J. C., & Coen, M. (2011). Processing and modeling of nuclear magnetic resonance (NMR) metabolic profiles. In T. O. Metz (Ed.), *Metabolic profiling: Methods and protocols*. vol. 708, (pp. 365–388). New York, NY: Humana Press. https://doi.org/10.1007/978-1-61737-985-7_21.
- Edmands, W. M. B., Ferrari, P., & Scalbert, A. (2014). Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Analytical Chemistry*, 86(21), 10925–10931. <https://doi.org/10.1021/ac503190m>.
- Edzes, H. T. (1990). The nuclear magnetization as the origin of transient changes in the magnetic field in pulsed NMR experiments. *Journal of Magnetic Resonance*, 86(2), 293–303. [https://doi.org/10.1016/0022-2364\(90\)90261-7](https://doi.org/10.1016/0022-2364(90)90261-7).
- Eilers, P. H. C. (2003). A perfect smoother. *Analytical Chemistry*, 75(14), 3631–3636. <https://doi.org/10.1021/ac034173t>.
- Emwas, A. H., Saunders, M., Ludwig, C., & Günther, U. (2008). Determinants for optimal enhancement in ex situ DNP experiments. *Applied Magnetic Resonance*, 34(3), 483–494.
- Emwas, A. H. M., Al-Talla, Z. A., Yang, Y. & Kharbatia, N. M. (2015). Gas Chromatography–Mass Spectrometry of Biofluids and Extracts. In J. T. Bjerrum (Ed.) *Metabonomics methods and protocols* (pp. 91–112). New York, NY: Springer. <https://doi.org/10.1007/978-1-4939-2377-9>.
- Emwas, A. H., Al-Talla, Z. A., & Kharbatia, N. M. (2015). Sample collection and preparation of biofluids and extracts for gas chromatography–mass spectrometry. In J. T. Bjerrum (Ed.) *Metabonomics: Methods and protocols*. vol. 1277, (pp. 75–90). New York, NY: Springer. <https://doi.org/10.1007/978-1-4939-2377-9>.
- Emwas, A.-H., Roy, R., McKay, R. T., Ryan, D., Brennan, L., Tenori, L., et al. (2016). Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. *Journal of Proteome Research*, 15(2), 360–373. <https://doi.org/10.1021/acs.jproteome.5b00885>.
- Emwas, A.-H. M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. In J. T. Bjerrum (Ed.), *Metabonomics methods and protocols* (pp. 161–193). New York, NY: Springer. <https://doi.org/10.1007/978-1-4939-2377-9>.
- Emwas, A.-H. M., Salek, R. M., Griffin, J. L., & Merzaban, J. (2013). NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. *Metabolomics*, 9(5), 1048–1072. <https://doi.org/10.1007/s11306-013-0524-y>.
- Feng, G., Guihua, R., & Jinyuan, M. (2006). Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemometrics and Intelligent Laboratory Systems*, 82(1–2), 59–65. <https://doi.org/10.1016/j.chemolab.2005.08.009>.
- Filzmoser, P., & Walczak, B. (2014). What can go wrong at the data normalization step for identification of biomarkers? *Journal of Chromatography A*, 1362, 194–205. <https://doi.org/10.1016/j.chroma.2014.08.050>.
- Forshed, J., Andersson, F. O., & Jacobsson, S. P. (2002). NMR and Bayesian regularized neural network regression for impurity determination of 4-aminophenol. *Journal of Pharmaceutical and Biomedical Analysis*, 29(3), 495–505. [https://doi.org/10.1016/s0731-7085\(02\)00086-9](https://doi.org/10.1016/s0731-7085(02)00086-9).
- Forshed, J., Schuppe-Koistinen, I., & Jacobsson, S. P. (2003). Peak alignment of NMR signals by means of a genetic algorithm. *Analytica Chimica Acta*, 487(2), 189–199. [https://doi.org/10.1016/s0003-2670\(03\)00570-1](https://doi.org/10.1016/s0003-2670(03)00570-1).
- Giacomini, F., Le Corguille, G., Monsoor, M., Landi, M., Pericard, P., Petera, M., et al. (2015). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, 31(9), 1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>.
- Gibney, M. J., Walsh, M., Brennan, L., Roche, H. M., German, B., & van Ommen, B. (2005). Metabolomics in human nutrition: Opportunities and challenges. *American Journal of Clinical Nutrition*, 82(3), 497–503.
- Giskeodegard, G. F., Bloemberg, T. G., Postma, G., Sitter, B., Tessem, M.-B., Gribbestad, I. S., et al. (2010). Alignment of high resolution magic angle spinning magnetic resonance spectra using warping methods. *Analytica Chimica Acta*, 683(1), 1–11. <https://doi.org/10.1016/j.aca.2010.09.026>.
- Golotvin, S., & Williams, A. (2000). Improved baseline recognition and modeling of FT NMR spectra. *Journal of Magnetic Resonance*, 146(1), 122–125. <https://doi.org/10.1006/jmre.2000.2121>.
- Gonzalez-Gil, G., Thomas, L., Emwas, A. H., Lens, P. N. L., & Saikaly, P. E. (2015). NMR and MALDI-TOF MS based characterization of exopolysaccharides in anaerobic microbial aggregates from full-scale reactors. *Scientific Reports*. <https://doi.org/10.1038/srep14316>.
- Grimes, J. H., & O’Connell, T. M. (2011). The application of micro-coil NMR probe technology to metabolomics of urine and serum. *Journal of Biomolecular NMR*, 49(3–4), 297–305. <https://doi.org/10.1007/s10858-011-9488-2>.
- Grison, S., Fave, G., Maillot, M., Manens, L., Delissen, O., Blanchard, E., et al. (2016). Metabolomics reveals dose effects of

- low-dose chronic exposure to uranium in rats: Identification of candidate biomarkers in urine samples. *Metabolomics*. <https://doi.org/10.1007/s11306-016-1092-8>.
- Gromski, P. S., Xu, Y., Hollywood, K. A., Turner, M. L., & Goodacre, R. (2015). The influence of scaling metabolomics data on model classification accuracy. *Metabolomics*, *11*(3), 684–695. <https://doi.org/10.1007/s11306-014-0738-7>.
- Hajjar, D., Kremb, S., Sioud, S., Emwas, A. H., Voolstra, C. R., & Ravasi, T. (2017). Anti-cancer agents in Saudi Arabian herbals revealed by automated high-content imaging. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0177316>.
- Halamek, J., Vondra, V., & Kasal, M. (1994). The elimination of baseline distortions induced by audio filters. *Journal of Magnetic Resonance Series A*, *110*(2), 194–197. <https://doi.org/10.1006/jmra.1994.1204>.
- Hao, J., Liebecke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. D. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature Protocols*, *9*(6), 1416–1427. <https://doi.org/10.1038/nprot.2014.090>.
- Harris, R. K., Becker, E. D., De Menezes, S. M. C., Granger, P., Hoffman, R. E., & Zilm, K. W. (2008a). Further conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008) (Reprinted from Pure Appl. Chem., vol 80, pg 59, 2008). *Magnetic Resonance in Chemistry*, *46*(6), 582–598. <https://doi.org/10.1002/mrc.2225>.
- Harris, R. K., Becker, E. D., De Menezes, S. M. C., Granger, P., Hoffman, R. E., & Zilm, K. W. (2008b). Further conventions for NMR shielding and chemical shifts IUPAC recommendations 2008 (Reprinted from Pure Appl Chem, vol 80, pg 59, 2008). *Solid State Nuclear Magnetic Resonance*, *33*(3), 41–56. <https://doi.org/10.1016/j.ssnmr.2008.02.004>.
- Hochrein, J., Zacharias, H. U., Taruttis, F., Samol, C., Engelmann, J. C., Spang, R., et al. (2015). Data normalization of H-1 NMR metabolite fingerprinting data sets in the presence of unbalanced metabolite regulation. *Journal of Proteome Research*, *14*(8), 3217–3228. <https://doi.org/10.1021/acs.jproteome.5b00192>.
- Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, *18*, 9.
- Izquierdo-Garcia, J. L., Villa, P., Kyriazis, A., del Puerto-Nevaldo, L., Perez-Rial, S., Rodriguez, I., et al. (2011). Descriptive review of current NMR-based metabolomic data analysis packages. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *59*(3), 263–270. <https://doi.org/10.1016/j.pnmrs.2011.02.001>.
- Jatlow, P., McKee, S., & O'Malley, S. S. (2003). Correction of urine cotinine concentrations for creatinine excretion: Is it useful? *Clinical Chemistry*, *49*(11), 1932–1934.
- Jauhiainen, A., Madhu, B., Narita, M., Narita, M., Griffiths, J., & Tavare, S. (2014). Normalization of metabolomics data with applications to correlation maps. *Bioinformatics*, *30*(15), 2155–2161. <https://doi.org/10.1093/bioinformatics/btu175>.
- Karaman, I., Ferreira, D. L. S., Boulange, C. L., Kaluarachchi, M. R., Herrington, D., Dona, A. C., et al. (2016). Workflow for integrated processing of multicohort untargeted H-1 NMR metabolomics data in large-scale metabolic epidemiology. *Journal of Proteome Research*, *15*(12), 4188–4194. <https://doi.org/10.1021/acs.jproteome.6b00125>.
- Karpievitch, Y. V., Nikolic, S. B., Wilson, R., Sharman, J. E., & Edwards, L. M. (2014). Metabolomics data normalization with EigenMS. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0116221>.
- Keun, H. C., Beckonert, O., Griffin, J. L., Richter, C., Moskau, D., Lindon, J. C., et al. (2002). Cryogenic probe ¹³C NMR spectroscopy of urine for metabolomic studies. *Analytical Chemistry*, *74*(17), 4588–4593.
- Keun, H. C., Ebbels, T. M. D., Antti, H., Bollard, M. E., Beckonert, O., Holmes, E., et al. (2003). Improved analysis of multivariate data by variable stability scaling: Application to NMR-based metabolic profiling. *Analytica Chimica Acta*, *490*(1–2), 265–276. [https://doi.org/10.1016/s0003-2670\(03\)00094-1](https://doi.org/10.1016/s0003-2670(03)00094-1).
- Kohl, S. M., Klein, M. S., Hochrein, J., Oefner, P. J., Spang, R., & Gronwald, W. (2012). State-of-the-art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics*, *8*(1), S146–S160. <https://doi.org/10.1007/s11306-011-0350-z>.
- Kvalheim, O. M., Brakstad, F., & Liang, Y. Z. (1994). Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Analytical Chemistry*, *66*(1), 43–51. <https://doi.org/10.1021/ac00073a010>.
- Lee, G. C., & Woodruff, D. L. (2004). Beam search for peak alignment of NMR signals. *Analytica Chimica Acta*, *513*(2), 413–416. <https://doi.org/10.1016/j.aca.2004.02.068>.
- Levitt, M. H. (1996). Demagnetization field effects in two-dimensional solution NMR. *Concepts in Magnetic Resonance*, *8*(2), 77–103.
- Li, C., & Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biology*, *2*(8), 1–11.
- Li, S., Winters, H., Jeong, S., Emwas, A. H., Vigneswaran, S., & Amy, G. L. (2016). Marine bacterial transparent exopolymer particles (TEP) and TEP precursors: Characterization and RO fouling potential. *Desalination*, *379*, 68–74. <https://doi.org/10.1016/j.desal.2015.10.005>.
- Lindon, J. C., Holmes, E., & Nicholson, J. K. (2007). Metabonomics in pharmaceutical R & D. *FEBS Journal*, *274*(5), 1140–1151. <https://doi.org/10.1111/j.1742-4658.2007.05673.x>.
- Ludwig, C., Marin-Montesinos, I., Saunders, M. G., Emwas, A. H., Pikramenou, Z., Hammond, S. P., et al. (2010). Application of ex situ dynamic nuclear polarization in studying small molecules. *Physical Chemistry Chemical Physics*, *12*(22), 5868–5871. <https://doi.org/10.1039/c002700f>.
- Luszczek, E., Lexcen, D., Witowski, N., Mulier, K., & Beilman, G. (2013a). Urinary metabolic network analysis in trauma, hemorrhagic shock, and resuscitation. *Journal of Proteome Research*, *12*(1), 223–235. <https://doi.org/10.1007/s11306-012-0441-5>.
- Luszczek, E. R., Lexcen, D. R., Witowski, N. E., Mulier, K. E., & Beilman, G. (2013b). Urinary metabolic network analysis in trauma, hemorrhagic shock, and resuscitation. *Metabolomics*, *9*(1), 223–235. <https://doi.org/10.1007/s11306-012-0441-5>.
- Marion, D., & Bax, A. (1988). Baseline distortion in real-Fourier-transform NMR spectra. *Journal of Magnetic Resonance*, *79*(2), 352–356. [https://doi.org/10.1016/0022-2364\(88\)90230-2](https://doi.org/10.1016/0022-2364(88)90230-2).
- Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., et al. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids - IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *European Journal of Biochemistry*, *256*(1), 1–15. <https://doi.org/10.1046/j.1432-1327.1998.2560001.x>.
- Markley, J. L., Ulrich, E. L., Berman, H. M., Henrick, K., Nakamura, H., & Akutsu, H. (2008). BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): New policies affecting biomolecular NMR depositions. *Journal of Biomolecular NMR*, *40*(3), 153–155. <https://doi.org/10.1007/s10858-008-9221-y>.
- Martinez-Arranz, I., Mayo, R., Perez-Cormenzana, M., Mincholé, I., Salazar, L., Alonso, C., et al. (2015). Enhancing metabolomics research through data mining. *Journal of Proteomics*, *127*, 275–288. <https://doi.org/10.1016/j.jprot.2015.01.019>.

- Mercier, P., Lewis, M. J., Chang, D., Baker, D., & Wishart, D. S. (2011). Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *Journal of Biomolecular NMR*, 49(3–4), 307–323. <https://doi.org/10.1007/s10858-011-9480-x>.
- Miao, Z. X., Jin, M. X., Liu, X., Guo, W., Jin, X. J., Liu, H. Y., et al. (2015). The application of HPLC and microprobe NMR spectroscopy in the identification of metabolites in complex biological matrices. *Analytical and Bioanalytical Chemistry*, 407(12), 3405–3416. <https://doi.org/10.1007/s00216-015-8556-y>.
- Miller, R. C., Brindle, E., Holman, D. J., Shofer, J., Klein, N. A., Soules, M. R., et al. (2004). Comparison of specific gravity and creatinine for normalizing urinary reproductive hormone concentrations. *Clinical Chemistry*, 50(5), 924–932. <https://doi.org/10.1373/clinchem.2004.032292>.
- Monakhova, Y. B., Schutz, B., Schafer, H., Spraul, M., Kuballa, T., Hahn, H., et al. (2014). Validation studies for multicomponent quantitative NMR analysis: The example of apple fruit juice. *Accreditation and Quality Assurance*, 19(1), 17–29. <https://doi.org/10.1007/s00769-013-1026-3>.
- Muncey, H. J., Jones, R., De Iorio, M., & Ebbels, T. M. D. (2010). MetAssimulo: Simulation of realistic NMR metabolic profiles. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-11-496>.
- Nagato, E. G., Lankadurai, B. P., Soong, R., Simpson, A. J., & Simpson, M. J. (2015). Development of an NMR microprobe procedure for high-throughput environmental metabolomics of *Daphnia magna*. *Magnetic Resonance in Chemistry*, 53(9), 745–753. <https://doi.org/10.1002/mrc.4236>.
- Nielsen, N. P. V., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 805(1–2), 17–35. [https://doi.org/10.1016/S0021-9673\(98\)00021-1](https://doi.org/10.1016/S0021-9673(98)00021-1).
- Nowick, J. S., Khakshoor, O., Hashemzadeh, M., & Brower, J. O. (2003). DSA: A new internal standard for NMR studies in aqueous solution. *Organic Letters*, 5(19), 3511–3513. <https://doi.org/10.1021/ol035347w>.
- Parsons, H. M., Ludwig, C., Gunther, U. L., & Viant, M. R. (2007). Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-8-234>.
- Pearce, J. T. M., Athersuch, T. J., Ebbels, T. M. D., Lindon, J. C., Nicholson, J. K., & Keun, H. C. (2008). Robust algorithms for automated chemical shift calibration of 1D H-1 NMR spectra of blood serum. *Analytical Chemistry*, 80(18), 7158–7162. <https://doi.org/10.1021/ac8011494>.
- Peng, J. T., Peng, S. L., Jiang, A., Wei, J. P., Li, C. W., & Tan, J. (2010). Asymmetric least squares for multiple spectra baseline correction. *Analytica Chimica Acta*, 683(1), 63–68. <https://doi.org/10.1016/j.aca.2010.08.033>.
- Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., et al. (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0124219>.
- Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D., & Lu, L. J. (2015). Computational and statistical analysis of metabolomics data. *Metabolomics*, 11(6), 1492–1513. <https://doi.org/10.1007/s11306-015-0823-6>.
- Saccenti, E. (2017). Correlation patterns in experimental data are affected by normalization procedures: Consequences for data analysis and network inference. *Journal of Proteome Research*, 16(2), 15. <https://doi.org/10.1021/acs.jproteome.6b00704>.
- Sakia, R. M. (1992). THE BOX-COX TRANSFORMATION TECHNIQUE - A REVIEW. *Statistician*, 41(2), 169–178. <https://doi.org/10.2307/2348250>.
- Sauve, J. F., Levesque, M., Huard, M., Drolet, D., Lavoue, J., Tardif, R., et al. (2015). Creatinine and Specific gravity normalization in biological monitoring of occupational exposures. *Journal of Occupational and Environmental Hygiene*, 12(2), 123–129. <https://doi.org/10.1080/15459624.2014.955179>.
- Savorani, F., Tomasi, G., & Engelsen, S. B. (2010). icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, 202(2), 190–202.
- Schleif, F. M., Riemer, T., Borner, U., Schnapka-Hille, L., & Cross, M. (2011). Genetic algorithm for shift-uncertainty correction in 1-D NMR-based metabolite identifications and quantifications. *Bioinformatics*, 27(4), 524–533. <https://doi.org/10.1093/bioinformatics/btq661>.
- Schuyler, A. D., Maciejewski, M. W., Stern, A. S., & Hoch, J. C. (2015). Nonuniform sampling of hypercomplex multidimensional NMR experiments: Dimensionality, quadrature phase and randomization. *Journal of Magnetic Resonance*, 254, 121–130. <https://doi.org/10.1016/j.jmr.2015.02.015>.
- Serkova, N., Fuller, T. F., Klawitter, J., Freise, C. E., & Niemann, C. U. (2005). H-1-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants. *Kidney International*, 67(3), 1142–1151. <https://doi.org/10.1111/j.1523-1755.2005.00181.x>.
- Smith, L. M., Maher, A. D., Want, E. J., Elliott, P., Stampler, J., Hawkes, G. E., et al. (2009). Large-scale human metabolic phenotyping and molecular epidemiological studies via H-1 NMR spectroscopy of urine: Investigation of borate preservation. *Analytical Chemistry*, 81(12), 4847–4856. <https://doi.org/10.1021/ac9004875>.
- Smolinska, A., Blanchet, L., Buydens, L. M. C., & Wijmenga, S. S. (2012a). NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*, 750, 82–97. <https://doi.org/10.1016/j.aca.2012.05.049>.
- Smolinska, A., Blanchet, L., Coulier, L., Ampt, K. A. M., Luidert, T., Hintzen, R. Q., et al. (2012b). Interpretation and visualization of non-linear data fusion in kernel space: Study on metabolomic characterization of progression of multiple sclerosis. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0038163>.
- Sokolenko, S., McKay, R., Blondeel, E. J. M., Lewis, M. J., Chang, D., George, B., et al. (2013). Understanding the variability of compound quantification from targeted profiling metabolomics of 1D-H-1-NMR spectra in synthetic mixtures and urine with additional insights on choice of pulse sequences and robotic sampling. *Metabolomics*, 9(4), 887–903. <https://doi.org/10.1007/s11306-013-0503-3>.
- Sousa, S. A. A., Magalhaes, A., & Ferreira, M. M. C. (2013). Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems*, 122, 93–102. <https://doi.org/10.1016/j.chemolab.2013.01.006>.
- Spraul, M., Link, M., Schaefer, H., Fang, F., & Schuetz, B. (2015). Wine analysis to check quality and authenticity by fully-automated H-1-NMR. In A. JeanMarie (Ed.), 38th World Congress of Vine and Wine. <https://doi.org/10.1051/bioconf/20150502022>.
- Tang, K. W. A., Toh, Q. C., & Teo, B. W. (2015). Normalisation of urinary biomarkers to creatinine for clinical practice and research - when and why. *Singapore Medical Journal*, 56(1), 7–10.
- Tiziani, S., Emwas, A. H., Lodi, A., Ludwig, C., Bunce, C. M., Viant, M. R., et al. (2008). Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, 377(1), 16–23.
- Tkac, I., Oz, G., Adriany, G., Ugurbil, K., & Gruetter, R. (2009). In vivo H-1 NMR spectroscopy of the human brain at high magnetic fields: Metabolite quantification at 4T vs. 7T. *Magnetic*

- Resonance in Medicine*, 62(4), 868–879. <https://doi.org/10.1002/mrm.22086>.
- Tomasi, G., van den Berg, F., & Andersson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18(5), 231–241. <https://doi.org/10.1002/cem.859>.
- van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-7-142>.
- Vu, T. N., & Laukens, K. (2013). Getting your peaks in line: A review of alignment methods for NMR spectral data. *Metabolites*, 3(2), 259–276. <https://doi.org/10.3390/metabo3020259>.
- Waikar, S. S., Sabbiseti, V. S., & Bonventre, J. V. (2010). Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney International*, 78(5), 486–494. <https://doi.org/10.1038/ki.2010.165>.
- Wang, K. C., Wang, S. Y., Kuo, C. H., & Tseng, Y. F. J. (2013). Distribution-based classification method for baseline correction of metabolomic 1D proton nuclear magnetic resonance spectra. *Analytical Chemistry*, 85(2), 1231–1239. <https://doi.org/10.1021/ac303233c>.
- Wang, T., Shao, K., Chu, Q. Y., Ren, Y. F., Mu, Y. M., Qu, L. J., et al. (2009). Automics: An integrated platform for NMR-based metabolomics spectral processing and data analysis. *BMC Bioinformatics* <https://doi.org/10.1186/1471-2105-10-83>.
- Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of 1H NMR metabolomics data. *Analytical Chemistry*, 78(13), 4430–4442.
- Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., et al. (1995). 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR*, 6(2), 135–140.
- Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC-Trends in Analytical Chemistry*, 27(3), 228–237. <https://doi.org/10.1016/j.trac.2007.12.001>.
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y. F., et al. (2013). HMDB 3.0-The human metabolome database in 2013. *Nucleic Acids Research*, 41(D1), D801–D807. <https://doi.org/10.1093/nar/gks1065>.
- Workman, C., Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielser, H. B., et al. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology*, 3(9), 1–16.
- Worley, B., & Powers, R. (2014). MVAPACK: A complete data handling package for NMR metabolomics. *ACS Chemical Biology*, 9(5), 1138–1144. <https://doi.org/10.1021/cb4008937>.
- Wu, W., Daszykowski, M., Walczak, B., Sweatman, B. C., Connor, S. C., Haselden, J. N., et al. (2006a). Peak alignment of urine NMR spectra using fuzzy warping. *Journal of Chemical Information and Modeling*, 46(2), 863–875.
- Wu, W., Daszykowski, M., Walczak, B., Sweatman, B. C., Connor, S. C., Haseldeo, J. N., et al. (2006b). Peak alignment of urine NMR spectra using fuzzy warping. *Journal of Chemical Information and Modeling*, 46(2), 863–875. <https://doi.org/10.1021/ic050316w>.
- Xi, Y., & Rocke, D. M. (2008). Baseline correction for NMR spectroscopic metabolomics data analysis. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-9-324>.
- Xia, J. G., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, 37, W652–W660. <https://doi.org/10.1093/nar/gkp356>.
- Xia, J. G., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research*, 43(W1), W251–W257. <https://doi.org/10.1093/nar/gkv380>.
- Xia, J. G., & Wishart, D. S. (2010). MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research*, 38, W71–W77. <https://doi.org/10.1093/nar/gkq329>.
- Zhao, S., Luo, X., & Li, L. (2016). Chemical isotope labeling LC-MS for high coverage and quantitative profiling of the hydroxyl sub-metabolome in metabolomics. *Analytical Chemistry*, 88(21), 10617–10623. <https://doi.org/10.1021/acs.analchem.6b02967>.
- Zheng, C., Zhang, S. C., Ragg, S., Raftery, D., & Vitek, O. (2011). Identification and quantification of metabolites in H-1 NMR spectra by Bayesian model selection. *Bioinformatics*, 27(12), 1637–1644. <https://doi.org/10.1093/bioinformatics/btr118>.