Published in final edited form as:

J Magn Reson. 2017 December ; 285: 37–46. doi:10.1016/j.jmr.2017.10.005.

Robust and Transferable Quantification of NMR Spectral Quality using IROC Analysis

Matthew A Zambrello¹, Mark W Maciejewski¹, Adam D Schuyler¹, Gerard Weatherby¹, and Jeffrey C Hoch¹

¹UConn Health, Department of Molecular Biology and Biophysics, 263 Farmington Ave., Farmington, CT 06030-3305 USA

Abstract

Non-Fourier methods are increasingly utilized in NMR spectroscopy because of their ability to handle nonuniformly-sampled data. However, non-Fourier methods present unique challenges due to their nonlinearity, which can produce nonrandom noise and render conventional metrics for spectral quality such as signal-to-noise ratio unreliable. The lack of robust and transferable metrics (i.e. applicable to methods exhibiting different nonlinearities) has hampered comparison of non-Fourier methods and nonuniform sampling schemes, preventing the identification of best practices. We describe a novel method, *in situ* receiver operating characteristic analysis (IROC), for characterizing spectral quality based on the Receiver Operating Characteristic curve. IROC utilizes synthetic signals added to empirical data as “ground truth”, and provides several robust scalar-valued metrics for spectral quality. This approach avoids problems posed by nonlinear spectral estimates, and provides a versatile quantitative means of characterizing many aspects of spectral quality. We demonstrate applications to parameter optimization in Fourier and non-Fourier spectral estimation, critical comparison of different methods for spectrum analysis, and optimization of nonuniform sampling schemes. The approach will accelerate the discovery of optimal approaches to nonuniform sampling experiment design and non-Fourier spectrum analysis for multidimensional NMR.

Graphical abstract

Correspondence to hoch@uchc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Optimization of sampling schemes or data processing parameters, tests for convergence, and critical comparison of different methods for signal processing in NMR all require a measure of spectral quality to guide the analysis. Signal-to-noise ratio (SNR) or root-mean-square (RMS) difference from a reference spectrum are common metrics. When used in conjunction with linear methods of spectrum analysis such as the discrete Fourier Transform (DFT), they are both robust and transferable, in the sense that they can be used to compare spectra. With nonlinear methods of spectrum analysis, however, these metrics fail to accurately assess sensitivity, as it is possible to obtain cosmetic improvements in SNR that do not improve the ability to distinguish signal from noise¹. This presents a challenge to modern NMR spectroscopy, because non-Fourier methods are needed to process nonuniformly sampled (NUS) data, and nonlinearity is a hallmark of virtually all non-Fourier methods². The nature of the nonlinearities differs among methods, and often depend on the input data, further complicating the comparison of non-Fourier methods. Although anecdotal evidence exists for improvements in SNR through NUS³ and metrics for spectral quality that measure consistency with the experimental data have been suggested⁴, robust and transferable criteria for making critical comparisons of sensitivity and resolution remain elusive⁵. Here, we propose a method for quantifying spectral quality and characterizing best practices among non-Fourier approaches for reconstructing frequency spectra from NUS data, and for the design of efficient NUS schemes, that is both robust and transferable among methods exhibiting nonlinearities of varying degree. Because emerging biomolecular applications of multidimensional NMR frequently operate at the very limits of sensitivity, resolution, and experiment time^{6–10}, additional improvements in sensitivity and resolution and reductions in experiment time afforded by NUS and non-Fourier methods are needed to extend NMR spectroscopy to systems that are larger and more complex, fleetingly stable, sparingly soluble, or available in limited supply. Robust metrics for spectral quality will enable investigators to develop the needed improvements.

Prior approaches and the threshold problem

The challenge of finding a sensitivity metric that is transferable among nonlinear methods can be illuminated by considering two hypothetical methods of spectrum analysis used to determine an NMR spectrum from NUS data that contains few signals and mostly noise. Method A perfectly recovers the signals and completely suppresses the noise. Method B perfectly recovers the noise but suppresses the signals. Using a frequently-used approach, the RMS difference or l_2 -norm¹¹ to compare the recovered spectra to the DFT spectrum of a corresponding uniformly sampled data set, method B would be considered better than method A. For similar reasons, SNR is not a reliable proxy for sensitivity, and is rendered even more unreliable when methods for spectrum analysis result in nonrandom noise distributions (see e.g. Figs. 3C,D of Hyberts et al.¹²). The detection of synthetic signals, either in purely synthetic data^{13,14} or added to empirical data for “in-situ” error analysis^{15,16} avoids these problems, but presents a detection-threshold problem (discussed below). A recently-proposed metric for spectra quality⁴ utilizing cross-validation that in principle is applicable to arbitrary methods of spectrum analysis, quantifies the level of agreement

between a spectral reconstruction and the measured data, but does not address the sensitivity or resolution of spectral estimates.

In principle, a peak-picker¹⁷ able to distinguish signal from noise is a potential approach to quantifying sensitivity. However most peak-pickers require specification of a threshold^{18–20}. When comparing the results of two different non-Fourier spectral estimates, the threshold at which signal can be distinguished from noise may be quite different and therefore a single threshold is unable to provide an accurate characterization of relative sensitivity. The difficulty of selecting a peak discrimination threshold is highly analogous to the problem of selecting contour levels for visualizing two-dimensional spectra, illustrated in Figure 1. In Fig. 1A, the ¹H-¹⁵N heteronuclear single-quantum coherence (HSQC) spectrum of the protein Prolactin is shown at high contour levels appropriate for discerning intense peaks. Fig. 1B shows the same spectrum at lower contour levels suitable for observing weaker peaks, but obscuring more intense peaks. The problem is exacerbated when using contour plots to compare spectra computed using spectral estimates that have very different properties. Fig. 1C and 1D display HSQC spectra of Prolactin at the same contour levels, processed with Maximum Entropy (MaxEnt) reconstruction²¹ (1C) and with DFT (1D). Peaks observable above a particular contour level in a spectrum processed with a nonlinear method are not necessarily observed above the same contour level in a spectrum processed with a linear method and a threshold that discriminates signal from noise in one spectrum will not similarly discriminate signal from noise in the other spectrum.

In the 1950's, an approach for assessing sensitivity in RADAR signal processing was developed that circumvents the problem posed by the choice of a detection threshold. Called the Receiver Operating Characteristic (ROC)²², it compares the True Positive Rate (TPR) for detecting signals with the False Positive Rate (FPR) at which noise is mischaracterized as signal. The two rates are computed as a function of detector threshold, resulting in an ROC curve. The ROC curve reveals the optimal threshold setting for discriminating noise from signal, and provides a versatile and robust measure of sensitivity. Although ROC analysis and NMR both trace their origin to the development of RADAR^{23–25}, ROC analysis has not found widespread use in NMR. It has, however, been widely applied in areas such as medical diagnostics and machine learning²⁶. One obstacle to the application of ROC analysis in NMR is that it requires *ground truth*, a set of signals that are known *a priori*, in order to compute the TPR and FPR. The lack of a general approach to identifying ground truth has been an obstacle to the application of ROC analysis in NMR.

Here we describe a novel and general approach to generating ROC curves for spectral estimates based on discrimination of synthetic signals added to empirical NMR data. The synthetic signals are added to resonance-free regions of the spectrum, forming the ground truth of known signals used to compute the number of true and false positives. The use of synthetic signals as ground truth was previously described as *in situ* analysis¹⁵, used to correct for nonlinearities in quantitative applications of MaxEnt reconstruction²⁷. We refer to the combination of the two approaches as *in situ*/receiver operating characteristic analysis, or IROC analysis. Here, we demonstrate the applications of IROC analysis for selection of spectral processing parameters using both Fourier and non-Fourier processing and for comparison of nonuniform sampling schemes.

An alternative approach for comparing non-Fourier methods and NUS schedules is to compare the results to a spectrum obtained from a uniformly sampled data set collected from the same sample and processed using conventional DFT processing. A shortcoming of this approach is that dynamic range, noise level, and spectral sparsity have an impact on the frequency accuracy, noise distribution, and sensitivity of non-Fourier spectral estimates, a property that is not shared by the DFT. This is particularly troublesome for NUS data since sampling artifacts arise through convolution of the point-spread function (PSF)²⁸ with empirical signals, and the distribution of these artifacts is different for different sampling schedules. These properties indicate that different non-Fourier methods might work best for different experiment types and sampling schedules. Because of the data-dependence of non-Fourier methods, comprehensive critical comparison requires an exhaustive set of test cases. While test cases can and should be used to provide insights into the relative performance of non-Fourier and NUS methods, there is also a need for robust and transferable *a priori* metrics or proxies for sensitivity, in the way that signal-to-noise (SNR) has served as a proxy for sensitivity of Fourier methods. The IROC approach described here circumvents many of the challenges posed by non-Fourier methods and provides transferable metrics. Computer codes for IROC analysis are available on the NMRbox platform (<http://nmrbox.org>)²⁹.

METHODS

Detailed Workflow for IROC analysis

Our approach to IROC analysis entails determination of the *recovery rate* and *false discovery rate* as calculated for a set of test signals, or ground truth, that are added to the empirical dataset. Recovery rate is calculated as fraction of synthetic peaks that are recovered and *false discovery rate* is defined as the fraction of total picked peaks that are false positives. Traditional ROC curves comprise a graphical plot of the true positive rate (TPR) of a particular classifier as a function of the false positive rate (FPR). Since the number of true negatives is required for calculation of the FPR, it cannot be calculated directly from an NMR spectrum. We therefore invoke the evaluation metrics of Yang et al³⁰, in which the curve is generated using *recovery rate* as a substitute for true positive rate and false discovery rate as a substitute for false positive rate. It is known phenomenologically that NMR signals are exponentially decaying sinusoids. Using this knowledge, we can create ground truth from synthetic exponentially decaying sinusoids, constructed mathematically and added to the empirical data. Signals must be added to the data so that the reconstruction is performed in the presence of all the imperfections of real data: non-random noise, NUS artifacts, etc. Although separating the inject regions with these features from the empirical data is possible, adding signals to a subset of the data assumes that the spectrum is zero-valued everywhere else and as such does not account for leakage of signals from adjacent points. The presence or absence of the synthetic signals can be assessed since they are constructed with known characteristics.

In general, IROC analysis involves identification of a resonance-free spectral region, addition of synthetic time-domain signals with frequencies falling in this region to the empirical dataset, computation of the spectrum, peak detection using multiple thresholds,

and calculation of ROC statistics for the ground truth peaks. The process is elaborated below:

1. Identifying a test region of interest. To identify a signal-free region, a preliminary spectral estimate is computed from the experimental dataset. In order to preserve the norms of the noise and signals, this is accomplished via the discrete Fourier Transform (DFT), without zerofilling or apodization, or alternatively using non-Fourier methods that preserve norms (e.g MINT³¹ or Forward MaxEnt³²). Next, the noise level is estimated and the standard deviation of baseline noise is calculated. Finally, a region of interest for adding the ground truth peaks must be specified. This is taken to be the largest submatrix in the preliminary spectral estimate for which the intensity is equivalent to the noise, or an operator-specified multiple thereof. The median absolute deviation (MAD) of noise intensity values can also serve in this function. To identify the region, we implemented a dynamic programming approach. The operator must first set the threshold for what constitutes noise in the preliminary spectrum. The frequency-domain data matrix of the preliminary spectrum is then converted to a 2-dimensional bit map with values above the noise threshold set to 0 and values below set to 1. The maximum submatrix containing all ones is calculated using the following iterative memoization technique: A result matrix of the same size as the input is created with all cells initially empty. Note empty is a different state than zero. Starting from the top left of the input matrix, each cell is examined. If a cell is zero, zero is stored in the corresponding location in the result matrix. If a one is present in the input, the minimum of the non-empty cells in the result is found, and one is added to it, and the result stored in the output. This works because an all ones submatrix of size N can be broken down into three parts, Relative to the lower right cell ("R") of the submatrix, there must exist: (1) An all ones submatrix with its lower right cell directly above R, (2) An all ones submatrix with its lower right cell directly to the left of R and (3) A filled diagonal of length N - 1 with terminus diagonally to the upper left of R. In this way, the algorithm identifies the largest submatrix of the bitmap populated with 1's, and therefore the largest submatrix region that represents only noise in the data matrix. Although this approach reliably finds regions that are obviously resonance free by visual inspection (and further consistent with expected resonance frequencies per the excitation bandwidth of the experiment), we recommend repeating this algorithm using multiple thresholds until a suitable value is determined. Another approach is to use a purely synthetic spectrum for analysis optimal sample schedule construction. In this case, a Gaussian distribution of noise may be added to the data prior to applications of the sample schedule.
2. Generation of augmented time-domain data. Once the frequency limits of the blank spectral region are known, we generate a set of synthetic signals with known properties (location, linewidth, and intensity). Bounds on the linewidths and amplitudes used to generate the synthetic signals are determined from the results of peak-picking (using any suitable software; see below) the preliminary

spectrum. Master linewidths and amplitudes for the synthetic signals in each dimension are such that they provide a good approximation to the empirical peaks, but can also be chosen to span ranges that satisfy the processing requirements of the spectroscopist as described in the main text. Peak frequencies are distributed within the bounds of the resonance-free region identified in step 1 using either a pseudo-random or quasi- (sub-) random distribution. Phases are chosen to match the empirical peaks. Synthetic decaying sinusoids corresponding to peaks with the stated properties are generated using the “*inject*” tool from RNMRTK, and added to the empirical data to yield the augmented time-domain data containing empirical signals and noise and the synthetic ground-truth signals.

3. Spectral processing. The augmented data comprised of synthetic signals, empirical resonance signals, and experimental noise is then subjected any desired method of spectrum analysis, including apodization, linear prediction and Fourier Transform or non-Fourier processing regimes, such as Maximum Entropy Reconstruction, Maximum Likelihood, Bayesian, Compressed Sensing, etc.
4. Peak picking. The frequencies of the peaks in the test region of the final spectrum are determined using ordinary peak detection software. In this work, we use NMRPipe, which employs threshold-based detection and parabolic interpolation. In our workflow, the threshold for detection is set to a fraction of the noise estimate in the spectrum to ensure that all possible resonances are assigned a chemical shift. At this threshold, spectral information is reported in tabular format for all peaks (both noise peaks and any recovered ground-truth peaks).
5. IROC curve construction. The presence or absence of master peaks in final spectrum is assessed and used to generate IROC statistics. To accomplish this, the output peak table generated during peak detection is ordered from highest to lowest intensity. The Euclidean distance in Hz between the output peak location and all master peak locations is calculated. Each Hz distance is divided by the linewidth of the master peak linewidth to generate a measurement that reflects distance in multiples of the master linewidth. This analysis uses the standard that any experimental peak and master peak with locations that match to within the error of one half the master linewidth constitutes a true positive. Further, if greater than two peaks match within this standard, the closer of the two is accepted as the true peak and the further is not assigned to the current peak in question (it may be a false positive or it may be assigned to another nearby peak in later iterations). Consequently, the absolute number of true positives and false positive is quantified during each iteration of the peak table.
6. Determination of IROC metrics. IROC analysis concludes with the computation of one or more metrics derived from the ROC curve. The true and false positives quantified in the first phase of the analysis are used to calculate the *recovery rate* and *false discovery rate* as defined above. The point (0.0,1.0) on the ROC plot

corresponds to a perfect detector with a false discovery rate of 0 and recovery rate of 1.0. The quality of a spectral estimate can be quantified as one of several single scalar quantities derived from ROC curves. The minimum distance, among all points on the ROC curve, to the point (0.0,1.0), is called the distance to the perfect classifier (DPC)³³. An alternative is the point on an ROC curve that exhibits the maximum recovery rate while detecting no false positives, the maximum recovery rate at minimum false discovery rate (MRMF). Another commonly-used metric for ROC curves is the area under the curve (AUC)³⁴. Departing from a single scalar metric, comparison of the recovery rate distribution for a specific false discovery rate has been reported³⁰.

System Selection and Data Processing

The analysis presented in this work can be segmented into two subsets; an analysis of empirical data to investigate processing parameters and a simulation study using purely synthetic data to examine the effect of sampling scheme and coverage on peak recovery. All data processing was conducted using the Rowland NMR Toolkit and peak detection was performed using NMRPipe.

Parameter Analysis—Time-domain data from heteronuclear single quantum coherence spectroscopy (HSQC) of the 41.5 kDa catalytic domain of the protein USP7 was used for this analysis. 128 points were collected in the ¹⁵N dimension at a carrier frequency of 81.047 MHz and a spectral width of 2500.00 Hz. 1202 points were collected in the ¹HN dimension at a carrier frequency of 799.75 MHz and a spectral width of 12019.23 Hz. 0 Hz reference values were 119.0 ppm and 4.7 ppm respectively. All spectra were reconstructed or zero-filled to a final size of 2048 points in the direct dimension and 256 points in the indirect dimension. For MaxEnt reconstructions, the parameters *def* and *aim* were set to values of 4.265 and 8.53, respectively. These values were selected using the procedure of Mobli et al.³⁵ MaxEnt reconstructions were allowed 4000 iterations to converge. Reconstructions that did not converge were excluded from the analysis (IROC metrics for nonconverged spectra were nevertheless included in some figures for added clarity). Linewidth deconvolution was applied to each dimension with kernel parameters ranging from 0 Hz to 20 Hz in increments of 1 Hz. In total, 441 spectra were produced for each ground truth peak list for a total of 4410 spectra. For the DFT spectral estimates, linear prediction (LP) extrapolation was applied along the indirect dimension after transform of the acquisition dimension. The first 80 points of the dataset was used for prediction. Shifted sinebell apodization was applied along both dimensions with shift values ranging from 0° to 90° in increments of 10°. In total 100 spectra were generated for each ground truth peak lists, producing 1000 total spectra. IROC analysis was applied to each of 5410 spectra.

Sampling Schedule Analysis—Synthetic spectra were generated according to the following metadata. The first dimension is defined as ¹⁵N and spans 128 points data points with a spectral width of 1472.0 Hz, a carrier frequency of 50.64 MHz and a 0 Hz reference value of 119.0 ppm. The second dimension is ¹HN and spans 256 points with a spectral width of 4096.0 Hz, a carrier frequency of 499.73 MHz and a 0 Hz reference value of 11.0 ppm. These values are typical for an HSQC spectrum of a 20 kDa globular protein collected

on a 500 MHz spectrometer. After initializing a zero-valued time-domain dataset according to the preceding parameters, we used the RNMRTK *noise* function to add pseudo-random data values consisting of Gaussian noise to each blank dataset. *Noise* allows for an operator specified mean and standard deviation, which were set to 0.0 and 10.2, respectively. The resulting time-domain dataset consisted of random noise.

Master Dataset Construction

A set of 50 hypothetical two-dimensional peaks pairs was generated at the outset, termed the ground truth peak list, with each entry defined as a frequency, linewidth, and intensity (see main text). Selection of these parameters is discussed below.

Parameter Analysis—Intensity and linewidth selection: Peak intensities were allowed to take on fixed values within a predefined intensity range that mimicked the intensity profile of the empirical resonances. Linewidths were selected similarly and then broadened by 10Hz.

Chemical Shift Selection—The resonance-free region of interest is defined as a subset of the full sampling grid (see above). To generate a master peak pair, an integer point is randomly selected from within this region and converted to the corresponding chemical shift using the referencing information obtained from the empirical dataset. For each point selected and converted to a chemical shift, a vicinal chemical shift is generated as well. The vicinal chemical shifts were allowed to randomly vary in proximity to the original chemical shift by 0.02–0.12 ppm in the f_1 dimension and by 0.01 – 0.1 ppm in the f_2 dimension. This approach provides a thorough survey of proximities from highly overlapping to reasonably well-dispersed. Once the full ground truth peak list is generated, the RNMRTK *inject* function is then used to construct sinusoids for each entry that produce the specified characteristics subsequent to transform. Note that point selection from within the region of interest was restricted to points occurring sufficiently internal to the boundaries such that after transform, a full Lorentzian line shape for each lies within the boundaries, i.e. – ground truth peaks are not clipped by the boundaries of the region of interest.

Sample Schedule Analysis—To analyze sample schedule optimization, we generated a separate set of 20 peak lists using a similar approach as above. In this case, the time-domain data consisted of the Gaussian noise as described in the System Selection section and the entire spectral window was used as the *inject* region. Single peaks ranging in intensity from a factor of 2 to a factor 10 times the RMS noise estimate of the preliminary spectrum were added at pseudo-randomly distributed chemical shifts within the spectral window. Linewidths were set to 20 Hz in each dimension.

The *inject* function offers several features that warrant further discussion. First, although experimentally collected data commonly yields peaks with non-Lorentzian line shapes as well as distorted peaks due to inhomogeneity of the magnetic field, the *inject* function produces peaks with perfect Lorentzian line shapes. This feature provides a more thorough investigation of an algorithm's processing prowess because in theory, a given processing regime should more readily recover a perfect Lorentzian as opposed to distorted peak.

Second, *inject* provides a means of quantitatively assessing peak recovery from the naturally occurring noise distribution of the experimental data. There is no way of knowing the characteristics of an empirical resonance *a priori*, however recovery of ground truth peaks can be assessed since they have specified characteristics.

NUS Schedule Generation

In lieu of a defined theoretical underpinning to guide sample schedule construction, one can resort to simulation studies to examine the effect of various sampling schemes and coverages. In this work, sampling schedules were generated according to on-grid, full-component sampling schemes. One set of schedules was generated according to enveloped-matched filter sampling (EMS)³⁶. In this scheme, points are selected in such a way that earlier evolution time points are more likely to be selected relative to later evolution time points, with the probability of selection decreasing exponentially as later times are reached. In this study, sets of schedules were generated for which the magnitude of the exponential bias matches the decay or twice the decay rate of the time-domain dataset. Another set of schedules was generated according to stochastic random sampling (RS). This scheme consists of pseudo-random integer sampling from the full Nyquist grid space of the synthetic dataset with each point having an equal probability of selection. In simulation studies, RS serves as a control to evaluate the relative efficacy of various schemes. In principle, the PSF, i.e.- the DFT of a constant signal that is zero-valued at points that are not specified in the sample schedule, can be used to make comparison between sampling schemes.

Although points can be selected according to the various sampling schemes, the percentage of total points that is selected must be specified. The sampling coverage values used in this study are 10%, 14%, 18% and 22%. 20 different sample schedules were generated for each sampling scheme/coverage combination for a total of 80 total sample schedules. In total, we conducted IROC analysis on 4800 spectra (20 sample schedules, 20 ground truth realization, 4 coverages, and 3 sampling schemes).

RESULTS and DISCUSSION

Workflow for IROC analysis and test dataset construction

The central feature of our approach is that it is applicable to arbitrary NMR experiments and data processing methods, including both Fourier and non-Fourier methods of spectrum analysis, and it can be implemented using any extant NMR peak-detection software. A schematic workflow is shown in Fig. 2. The central step involves the addition of synthetic time-domain signals as ground truth to an empirical dataset and subsequent quantification of the synthetic signals. The empirical dataset (Fig. 2A) is first processed to generate a preliminary spectral estimate (Fig. 2B). The preliminary estimate is used to identify a region of interest (ROI) lacking empirical resonances and containing only noise (see Detailed Workflow for IROC analysis in Methods). The synthetic ground truth signals (Fig. 2C) are constructed to recapitulate characteristics of the empirical signals determined from the preliminary spectral estimate. The set of ground truth signals can be designed to test specific spectral characteristics, such as sensitivity or resolution. Adjustable parameters for the synthetic signals include properties such as linewidth, dynamic range, SNR, and frequency

separation. A DFT spectral estimate for a representative synthetic ground truth signal set is displayed in Figure 2D. The sum of the synthetic data and the empirical data produces an augmented time-domain dataset (Fig. 2E), which can be processed using any suitable method to yield a spectral estimate containing both empirical and ground truth peaks (Fig. 2F).

Next, peak detection is conducted with varying discrimination thresholds used to quantify the recovery of ground truth signals. Peak data is used to construct an IROC curve by computing the proportion of ground truth peaks correctly recovered (recovery rate) and the proportion of total number of peaks that are deemed false positives (false discovery rate) at each discrimination threshold (see Detailed Workflow for IROC analysis in Methods section). The quality of the spectrum can be quantified from the IROC curve in a number of ways. Metrics include (1) the recovery rate at the point with minimum distance to the perfect classification (MDPC; perfect classification corresponds to the upper left corner of the ROC curve, (0.0, 1.0)), (2) the area under the IROC curve (AUC), and (3) maximum recovery rate at minimum false discovery rate (MRMF)³⁰ (see Figure 3). In the context of ordinary receiver operating characteristic analysis, the point of the ROC curve that is closest to (0.0, 1.0) represents the tradeoff between accurately finding true positives and picking up more false positives. Here, the MDPC is an analogous metric that reflects the tradeoff between finding the most ground truth peaks and least false peaks (noise or artifacts). AUC is another numerical value for comparing the overall quality of different IROC curves/spectra without specifying a unique threshold for comparison. The user may opt to examine the absolute number of peaks that can be found without finding any false positives (MRMF metric). Finally, as a by-product of testing for the recovery of synthetic peaks, users can examine the RMSD between the recovered peaks and ground truth peak locations to obtain a metric for frequency accuracy.

To explore sensitivity, the ground truth set can be comprised of synthetic signals ranging in amplitude from the intensity of the noise to that of the empirical signals, based on the noise level and peak amplitudes found in the preliminary spectral estimate. Suitable spectral processing parameters, such as apodization functions or those controlling a non-Fourier method, can be determined from IROC metrics.

In the parameter comparison example below, the ground truth set is designed to measure resolution in an 800 MHz TROSY-HSQC spectrum for the 41.5 kDa catalytic domain of the protein USP7. Approximately 70% of the catalytic domain backbone resonances have been assigned (Alexandra Pozhidaeva and Irina Bezsonova, unpublished data). Assignment of poorly resolved peaks was achieved via triple resonance experiments conducted at 800 MHz on a deuterated sample. In several instances, 3D experiments at 800 MHz were essential to distinguish peaks with nearly identical nitrogen and hydrogen chemical shifts. The separation between these nearly-coincident empirical peaks was used to determine the proximity between the ground truth peaks (see below). The ground truth signal intensities are set within the limits of the dynamic range of empirical signals in the preliminary spectrum, however the synthetic linewidths were increased by approximately 10 Hz, resulting in additional overlap. The final augmented dataset included 100 ground truth signals. A first set of signal frequencies is selected using a two-dimensional pseudo-random

deviate (see Methods). A second, neighboring peak is added, with frequency separations varying randomly between 0.01 to 0.02 ppm in the acquisition dimension and 0.02 to 0.12 ppm in the indirect dimension from the frequencies of the initial signal.

Parameter analysis and comparison of spectral processing workflows

IROC analysis was used to compare DFT and MaxEnt processing workflows and to explore the influence of the adjustable parameters for each processing workflow for an 800 MHz TROSY-HSQC spectrum. The ground truth, as described in the previous section, is constructed to assess resolution. Using the augmented dataset, we conducted separate two-dimensional parameter sweeps of both the linewidth deconvolution³⁷ parameters (MaxEnt processing) and apodization (DFT processing) parameters. Linewidth deconvolution refers to removal of unwanted manifestations of a particular decay envelope from the spectrum. This is achieved by constructing a spectrum that lacks the effect of the undesirable decay envelope but is still consistent with the empirical data, termed the deconvolved spectrum. The DFT⁻¹ of the deconvolved spectrum would not match the empirical data and so is multiplied by the undesirable decay envelope (kernel) that has been removed from the spectrum in order to establish a matched between the empirical data and deconvolved spectrum¹⁶. See Methods for additional information about spectral processing and master dataset construction.

The IROC metrics described above were used to determine the best values within the range of adjustable parameters used in this survey. For DFT processing, linear prediction (LP) extrapolation was applied along the indirect dimension of the augmented dataset and shifted sinebell apodization was applied, with the shifts ranging from 0° to 90° in increments of 10° in both dimensions. In total, 100 DFT spectra were generated from the augmented dataset. A similar workflow was applied for MaxEnt deconvolution, with exponentially-decaying deconvolution kernels in each dimension corresponding to 0 to 20 Hz widths at half height in increments of 1 Hz, generating 441 spectra. The final output size of all spectra was 256×2048 points. Both parameter sweeps were repeated for 10 different realizations of the ground truth, using randomly-generated peak locations within the region of interest. Additional details regarding spectral processing and IROC curve construction can be found in the Methods section. After processing, we constructed IROC curves for each computed spectrum and calculated recovery rate at MDPC and MRMF metrics for each curve to ascertain which set of processing parameters provided the best recovery of the ground truth set. Figure 4 demonstrates the use of IROC analysis for identification of the spectrum with highest resolution and sensitivity for one realization of the ground truth set (and by proxy the empirical peaks). The spectrum processed with deconvolution parameters of 20 Hz in f_1 and 13 Hz in f_2 provides the most favorable IROC metrics (recovery rate at MDPC, MRMF). In this spectrum, 92% of the ground truth is recovered at the MDPC metric (recovered ground truth peaks are depicted in Fig. 4A). This is consistent with the MRMF metric as both identify the same spectrum as best. Deconvolution of 5 Hz in f_1 and 0 Hz in f_2 produced the least favorable IROC metrics. IROC curves for these spectra are compared in Figure 4C where it is clear that the 20Hz/13Hz deconvolved spectrum achieves the best recovery of the ground truth. The recovery rates obtained at the MDPC calculated for all spectra using one realization of the ground truth are depicted in Figure 4E, where “warmer”

colors (red) indicate higher recovery rate. Indeed, many of the deconvolved spectra demonstrate excellent recovery of the ground truth, as demonstrated by the warmer areas evident in the pseudo-color map.

Among the DFT spectra, the most favorable recovery rate at MDPC was achieved with 10° and 0° shifts in f_1 and f_2 , respectively. The recovered ground truth peaks from this spectrum, displayed in Figure 4B, are poorly resolved relative to the MaxEnt deconvolved spectrum. The highest achievable recovery rate in this MaxEnt spectrum was 76% of the ground truth peaks correctly identified at the MDPC. IROC curves generated from the most and least favorable spectra, as well as intermediate, are displayed in Figure 4D. Figure 4F shows the recovery rate achieved at MDPC for all DFT spectra. The values are generally lower than those from deconvolved spectra (4E), indicating that MaxEnt provides a better tradeoff between recovery rate and false discovery rate and is capable of recovering more ground truth peaks without misidentification. The trend apparent in 4F is consistent with the expected behavior of sinebell apodization (i.e. higher resolution enhancement with lower phase shift of the sinebell apodization). Boxplots illustrating the distribution of recovery rates at MDPC and at MRMF for all 10 realizations of the ground truth are provided in Figure 1S.

Comparison of IROC and other metrics

The available assignments for the USP7 catalytic domain were used to compare IROC-based metrics (calculated from the ground truth) with other metrics for spectral quality (calculated from the *empirical* peaks). Although this analysis uses 2D HSQC spectra, we focus on the peaks in spectra for the USP7 catalytic-domain that were only able to be resolved through conventional DFT processing of 3D triple resonance experiments. The separation of such peaks was used to set the spacing between pairs of ground truth peak pairs that were added to the HSQC data (examples of such peaks shown in Figure 5A–F). Having identified the most favorable set of processing parameters through IROC analysis (MaxEnt processed with 20 Hz deconvolution in f_1 and 13 Hz deconvolution in f_2 ; FT processed with 10° shift in f_1 and 0° shift in f_2), we compared the resolution of empirical peaks after processing by each method. The upper six panels of Figure 5 show instances of empirical peaks. The left panels show peaks taken from the IROC-optimized MaxEnt deconvolution spectrum and the right panels are taken from the IROC-optimized DFT spectrum. Excellent separation is apparent in the MaxEnt spectra, however these peaks are either poorly resolved or not resolved at all in any of the DFT spectra that we computed.

To determine if the resolution gains observed in the IROC-optimized deconvolution spectrum are consistent with conventional metrics, we measured the splitting parameter¹³, η , for the pair of peaks at (7.313 ppm, 116.224 ppm) and (7.310 ppm and 115.992 ppm) in each of the 441 MaxEnt deconvolved spectra. η can take on values ranging from 0.0 to 1.0, indicating no separation and separation to the baseline, respectively. Note that these two peaks were not separated at all in any of the DFT spectra. Figure 5G shows the recovery rates obtained at the MDPC for every MaxEnt deconvolved spectrum and 5H shows the η values for the peak pair displayed in 5A calculated from all deconvolved spectra. There is excellent correspondence between the patterns for η and recovery rate at MDPC. The

deconvolved spectrum processed with 19 Hz (f_2) and 14 Hz (f_1) achieves the highest η value (0.97) whereas the IROC-optimized spectrum was processed with 20 Hz and 13 Hz, respectively. Interestingly, the second highest η value was computed from the 20Hz/13Hz deconvolved spectrum (0.96). The distribution of η values calculated from the deconvolved spectra is broadly consistent with the IROC metric as shown in 5G and 5H, indicating that resolution of the empirical peaks is indeed consistent with the resolution of the ground truth peaks. Indeed, IROC metrics calculated from the ground truth peaks serve as an excellent proxy for resolution of the empirical peaks.

Sample schedule optimization using IROC analysis

The design of sampling schedules is a critical aspect of successful NUS experiments³⁸ as it impacts sensitivity, resolution, and the distribution and magnitude of sampling artifacts. Although there are a number of useful *a priori* metrics for characterizing NUS schedules, there is no formal theory available to guide the design of optimal schedules. Because IROC analysis can be applied to purely synthetic spectra, it provides a means of assessing the design of NUS schedules prior to their use. As an example of this approach, we compare the quality of spectra produced from three types of nonuniform sample schedules using a workflow similar to that described above. In this case, we generated purely synthetic data (synthetic time-domain signals added to random noise; see Methods). The ground truth was constructed with peak intensities ranging from well below the noise intensity to well above it. We sub-sampled the synthetic data according to each sampling schedule. In total, we generated 240 sample schedules (20 schedules at sampling coverages ranging from 10% to 22% by increments of 4% for random sampling, exponentially-biased random sampling with a matched decay rate, and exponentially-biased random sampling with a decay rate twice that of the time-domain data. We independently applied each schedule to 20 different synthetic time-domain datasets (realized using a two-dimensional sub-random deviate³⁹ for the ground truth peak locations), processed with MaxEnt reconstruction (without deconvolution) to produce 4800 spectral estimates, and subsequently generated IROC statistics for each spectrum. Randomization reduces the likelihood of bias from particularly fortuitous or deleterious signal frequencies or sampling schedules. Sample schedules were created with *nus-tool*, a utility for generating, converting, and analyzing sample schedules; *nus-tool* is distributed as part of the NMRbox platform.

Figure 6 panels A, B, and C depict the distributions of the *recovery rate at MDPC*, the distributions of AUC, and distribution of MRMF for spectra processed with each sampling scheme, as a function of sampling coverage (the percentage of indirect evolution time tuples sampled compared to the number of samples comprising the uniform Nyquist grid). The boxplots show that exponentially-biased random sampling (both matched decay and twice decay rate of the signal envelope) provides uniformly higher recovery rates and AUC values compared to purely random sampling. Biased sampling also provides lower variability (narrower boxes). These box plots show that IROC metrics (and therefore the ability to distinguish ground truth peaks from noise) improve with increasing coverage and increasing decay rate, a behavior that is consistent with previous reports on the properties of these sampling schemes⁴. Panel D shows the frequency accuracy of each spectrum as measured by the RMSD between the recovered peaks and ground truth peaks. The trend shows increasing

accuracy with increasing coverage and decay rate of the sampling schedule within the coverages and decay rates that were explored.

CONCLUDING REMARKS

The nonlinearity of non-Fourier methods present special challenges that render conventional metrics for spectral quality unreliable. SNR is not a useful proxy for sensitivity, and comparisons of spectra based on RMS difference or l_2 -norm fail to capture salient differences. Measures of spectral quality that are transferable, applicable to all classes of processing workflows, are essential for systematic comparisons. Separately, ROC and *in situ* approaches are only partial solutions. We have shown here that the metrics DPC, AUC, and MRMF derived from IROC curves enable comparison of processing algorithms, nonuniform sampling schedules, and optimize adjustable parameters in arbitrary signal processing workflows. This approach can be used to systematically identify best practices for experimental design and signal processing approaches in multidimensional NMR that previously were elusive or only identified anecdotally. Because IROC analysis is compatible with any data processing workflow it can be used as a routine test of spectral quality.

Acknowledgments

This work was supported by NIH grants R21GM104517 and P41GM111135, and an NIDCR predoctoral fellowship to MZ (F30DE026353). We thank Alexandra Pozhidaeva and Irina Bezsonova for supplying the USP7 catalytic domain data.

References

1. Bahrami A, Clos LJ 2nd, Markley JL, Butcher SE, Eghbalnia HR. RNA-PAIRS: RNA probabilistic assignment of imino resonance shifts. *Journal of Biomolecular NMR*. 2012; 52:289–302. DOI: 10.1007/s10858-012-9603-z. DOI: 10.1007/s10858-012-9603-z [PubMed: 22359049]
2. Mobli M, Hoch JC. Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR. *Prog Nucl Magn Reson Spectrosc*. 2014; 83C:21–41. DOI: 10.1016/j.pnmrs.2014.09.002
3. Hyberts SG, Robson SA, Wagner G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J Biomol NMR*. 2013; 55:167–178. DOI: 10.1007/s10858-012-9698-2 [PubMed: 23274692]
4. Wu Q, Coggins BE, Zhou P. Unbiased measurements of reconstruction fidelity of sparsely sampled magnetic resonance spectra. *Nat Commun*. 2016; 7:12281. [PubMed: 27459896]
5. Billeter M. Non-uniform sampling in biomolecular NMR. *J Biomol NMR*. 2017; 68:65–66. DOI: 10.1007/s10858-017-0116-7 [PubMed: 28620800]
6. Quinn CM, Polenova T. Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Q Rev Biophys*. 2017; 50:e1. [PubMed: 28093096]
7. Anglister J, Srivastava G, Naider F. Detection of intermolecular NOE interactions in large protein complexes. *Prog Nucl Magn Reson Spectrosc*. 2016; 97:40–56. DOI: 10.1016/j.pnmrs.2016.08.002 [PubMed: 27888839]
8. Cuniasse P, Tavares P, Orlova EV, Zinn-Justin S. Structures of biomolecular complexes by combination of NMR and cryoEM methods. *Curr Opin Struct Biol*. 2017; 43:104–113. DOI: 10.1016/j.sbi.2016.12.008 [PubMed: 28056362]
9. Fan TW, Lane AN. Applications of NMR spectroscopy to systems biochemistry. *Prog Nucl Magn Reson Spectrosc*. 2016; 92–93:18–53. DOI: 10.1016/j.pnmrs.2016.01.005
10. Luchinat E, Banci L. A Unique Tool for Cellular Structural Biology: In-cell NMR. *J Biol Chem*. 2016; 291:3776–3784. DOI: 10.1074/jbc.R115.643247 [PubMed: 26677229]

11. Worley B, Powers R. Deterministic multidimensional nonuniform gap sampling. *J Magn Reson.* 2015; 261:19–26. DOI: 10.1016/j.jmr.2015.09.016 [PubMed: 26524650]
12. Hyberts SG, Arthanari H, Robson SA, Wagner G. Perspectives in magnetic resonance: NMR in the post-FFT era. *Journal of Magnetic Resonance.* 2014; 241:60–73. doi:<http://dx.doi.org/10.1016/j.jmr.2013.11.014>. [PubMed: 24656081]
13. Stern AS, Li KB, Hoch JC. Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear-prediction extrapolation and maximum-entropy reconstruction. *J Am Chem Soc.* 2002; 124:1982–1993. [PubMed: 11866612]
14. Ying J, Delaglio F, Torchia DA, Bax A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR.* 2017; 68:101–118. DOI: 10.1007/s10858-016-0072-7 [PubMed: 27866371]
15. Schmieder P, Stern AS, Wagner G, Hoch JC. Quantification of maximum-entropy spectrum reconstructions. *J Magn Reson.* 1997; 125:332–339. [PubMed: 9144266]
16. Hoch, JC., Stern, AS. *NMR Data Processing.* Wiley-Liss; 1996.
17. Klukowski P, Walczak MJ, Gonczarek A, Boudet J, Wider G. Computer vision-based automated peak picking applied to protein NMR spectra. *Bioinformatics.* 2015; 31:2981–2988. DOI: 10.1093/bioinformatics/btv318 [PubMed: 25995228]
18. Bagaria A, Jaravine V, Huang YJ, Montelione GT, Güntert P. Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci.* 2012; 21:229–238. DOI: 10.1002/pro.2007 [PubMed: 22113924]
19. Bergt C, Reicher H, Malle E, Sattler W. Hypochlorite modification of high density lipoprotein: effects on cholesterol efflux from J774 macrophages. *FEBS Lett.* 1999; 452:295–300. [PubMed: 10386609]
20. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* 2015; 31:1325–1327. DOI: 10.1093/bioinformatics/btu830 [PubMed: 25505092]
21. Barnard EC, Brown G, Stow ND. Deletion mutants of the herpes simplex virus type 1 UL8 protein: effect on DNA synthesis and ability to interact with and influence the intracellular localization of the UL5 and UL52 proteins. *Virology.* 1997; 237:97–106. DOI: 10.1006/viro.1997.8763 [PubMed: 9344911]
22. Petersen WW, Birdsall TG, Fox WC. The Theory of Signal Detectability. *Transactions of the Institute of Radio Engineers Professional Group on Information Theory, PGIT.* 1954; 4:171–212.
23. Buder, R. *The Invention That Changed the World: How a Small Group of Radar Pioneers Won the Second World War and Launched a Technical Revolution.* Touchstone: 1997.
24. Mattson, J., Simon, M. *The Pioneers of NMR and Magnetic Resonance in Medicine.* Bar-Ilan University Press; 1996.
25. Collison P. Of bombers, radiologists, and cardiologists: time to ROC. *Heart.* 1998; 80:215–217. [PubMed: 9875074]
26. Fawcett T. An Introduction to ROC Analysis. *Patt Recog Lett.* 2006; 27:861–874.
27. Takeuchi K, Arthanari H, Wagner G. Perspective: revisiting the field dependence of TROSY sensitivity. *J Biomol NMR.* 2016; 66:221–225. DOI: 10.1007/s10858-016-0075-4 [PubMed: 27866370]
28. Schuyler AD, Maciejewski MW, Stern AS, Hoch JC. Formalism for hypercomplex multidimensional NMR employing partial-component subsampling. *J Magn Reson.* 2013; 227:20–24. DOI: 10.1016/j.jmr.2012.11.019 [PubMed: 23246651]
29. Maciejewski MW, et al. NMRbox: A Resource for Biomolecular NMR Computation. *Biophys J.* 2017 in press.
30. Yang C, He Z, Yu W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics.* 2009; 10:4. [PubMed: 19126200]
31. Paramasivam S, et al. Enhanced Sensitivity by Nonuniform Sampling Enables Multidimensional MAS NMR Spectroscopy of Protein Assemblies. *Journal of Physical Chemistry B.* 2012; 116:7416–7427. DOI: 10.1021/Jp3032786

32. Hyberts S, Frueh D, Arthanari H, Wagner G. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *Journal of Biomolecular NMR*. 2009; 45:283–294. [PubMed: 19705283]
33. Kumar R, Indrayan A. Receiver Operating Characteristic (ROC) Curve for Medical Researchers. *Indian Pediatrics*. 2011; 48:277–287. [PubMed: 21532099]
34. Swets JA, Dawes RM, Monahan J. *Scientific American*. 2000;82–87.
35. Mobli M, Maciejewski MW, Gryk MR, Hoch JC. An automated tool for maximum entropy reconstruction of biomolecular NMR spectra [3]. *Nature Methods*. 2007; 4:467–468. [PubMed: 17538627]
36. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP. Exponential sampling: an alternative method for sampling in two dimensional NMR experiments. *J Magn Reson*. 1987; 73:69–77.
37. Hoch JC, Stern AS. Maximum Entropy Reconstruction, Spectrum Analysis and Deconvolution in Multidimensional Nuclear Magnetic Resonance. *Methods Enzym*. 2001; 338:159–178.
38. Ziegler V, Suss E. The allergenic effect of rubber accelerators tetramethyl thiuram disulfide (TMTD) and mercaptobenzothiazole (MBT). *Allerg Immunol (Leipz)*. 1974;20–21. 281–285.
39. Brooks CL, Li M, Hu M, Shi Y, Gu W. The p53-Mdm2-HAUSP complex is involved in p53 stabilization by HAUSP. *Oncogene*. 2007; 26:7262–7266. [PubMed: 17525743]

Highlights

- Non-Fourier methods enable nonuniform sampling for multidimensional NMR
- Nearly all non-Fourier methods are nonlinear, making it difficult to quantify spectral quality (especially sensitivity)
- Receiver operator characteristic (ROC) is a decades-old technology for quantifying sensitivity
- Intrinsic ROC uses synthetic signals as ground truth needed for ROC
- IROC enables critical comparison of non-Fourier and NUS methods

Figure 1. The threshold problem, manifest in the choice of contour levels

The choice of contour levels for comparison of spectra is especially problematic for nonlinear methods of spectrum analysis, but is also difficult for linear methods such as the DFT when there is substantial dynamic range between the weakest and strongest signals. (A) HSQC spectrum of the protein prolactin at a high contour levels appropriate for discriminating high intensity overlapping peaks. (B) The same peaks are not discriminated at lower contour levels. (C) and (D) illustrate HSQC spectra of prolactin at identical contour levels, processed with MaxEnt deconvolution (C) and DFT (D).

Figure 2. Workflow for IROC analysis

A through F depict the state of the data during the IROC workflow. (A) Real/real channel of the empirical time-domain data. (B) A preliminary spectral estimate is used to identify an empty region of the spectrum. (C) Synthetic signals with frequencies falling in the identified empty region (spectrum shown in (D)) are added to the measured time-domain data. (E) The augmented data set. (F) The spectrum of the augmented data is analyzed to determine the discovery rates for constructing the IROC curve.

Figure 3. The IROC curve and derived metrics

An annotated IROC curve depicts three scalar metrics for spectral quality. The point (0.0,1.0) represents perfect detection, recovery of all ground truth peaks without detection of false positives. The point on the IROC curve that is closest to this point is denoted *DPC*, and the recovery rate achieved at this point can be interpreted the most favorable tradeoff between recovery rate and false discovery rate. The point denoted by *MRMF* signifies the highest achievable recovery rate without detection of any false positives. The metric “area under the curve” (AUC) gives a measure of how much better signal detection is compared to a detector that randomly assigns peaks (a random detector has AUC of 0.5). See Methods for details on peak detection and correspondence.

Figure 4. IROC-optimized spectral processing

Results for the region of interest (ROI) indicated in Fig. 2(F). (A) MaxEnt deconvolution and (B) DFT spectra, using values for the deconvolution kernel decay rates (MaxEnt) and sine-bell shifts (DFT) determined to be optimal from the MRMF in the IROC curves. IROC curves for MaxEnt (C) and DFT (D) for three different sets of parameter values. The fraction of ground truth peaks recovered in the ROI as a function of the adjustable parameters for (E) MaxEnt deconvolution (kernel decay rate) and (F) DFT (sine-bell shift).

Figure 5. Quantifying resolution

Three examples of empirical peaks in the optimized MaxEnt (A, C, E) and DFT (B, D, F) spectra. Vertical (A, B, C, D) and horizontal (E, F) one dimensional cross-sections through the peaks are shown in red, demonstrating the superior resolution of MaxEnt spectra. (G) and (H) depict the recovery rate (fraction of ground truth peaks recovered, using closely-spaced signals) and splitting parameter η (see text; 0 corresponds to peaks not resolved), as a function of the values of the adjustable parameters for MaxEnt reconstruction. The black regions indicate parameter values for which the MaxEnt reconstruction did not converge. The concordance between RR and η shows that IROC analysis can be used to report on resolution using an appropriately-designed ground truth set.

Figure 6. Distribution of recovery rates for different NUS schedules

(A), (B) and (C) Box plots depicting the distributions of recovery rate at the MDPC (A), AUC (B) and the MRMF (C) for purely synthetic data containing random noise, sampled according to exponentially decaying with the sampling density decay rate matched to the signal envelope decay rate (cyan), decay rate of twice the signal envelope decay rate (red), or purely random sampling (blue). Results are depicted as a function of the sampling coverage (the number of samples divided by the number of elements in the corresponding uniform sampling grid). Exponentially-biased schedules yield more favorable IROC metrics at each sampling coverage relative to random sampling. (D) RMSD values computed between recovered peak and ground truth peak locations. There is a negative correlation between IROC metrics and RMSD. Outliers are represented by “+”.