Published in final edited form as:

J Magn Reson. 2020 February ; 311: 106671. doi:10.1016/j.jmr.2019.106671.

The Influence of the Probability Density Function on Spectral Quality in Nonuniformly Sampled Multidimensional NMR

Matthew A. Zambrello¹, D. Levi Craft², Jeffrey C. Hoch¹, David Rovnyak², Adam D. Schuyler^{1,*}

¹UConn Health, Department of Molecular Biology and Biophysics, Farmington, CT 06030 USA

²Bucknell University, Department of Chemistry, Lewisburg, PA 17837 USA

Abstract

The goal of nonuniform sampling (NUS) is to select a subset of free induction decays (FIDs) from the conventional, uniform grid in a manner that sufficiently samples short evolution times needed for improved sensitivity and long evolution times needed for enhanced resolution. In addition to specifying the *number* of FIDs to be collected from a uniform grid, NUS schemes also specify the *distribution* of the selected FIDs, which directly impacts sampling-induced artifacts. Sampling schemes typically address these heuristic guidelines by utilizing a probability density function (PDF) to bias the distribution of sampled evolution times. Given this common approach, schemes differentiate themselves by how the evolution times are distributed *within* the envelope of the PDF. Here, we employ maximum entropy reconstruction and utilize *in situ* receiver operating characteristic (IROC) to conduct a critical comparison of the sensitivity and resolution that can be achieved by three types of biased sampling schemes: exponential (PDF is exponentially decaying), Poisson-gap (PDF derived from a sine function), and quantile-directed (PDF defined by simple polynomial decay). This methodology reveals practical insights and trends regarding how the sampling schemes and bias can provide the highest sensitivity and resolution for two nonuniformly sampled dimensions in a three-dimensional biomolecular NMR experiment. The IROC analysis circumvents the limitations of common metrics when used with nonlinear spectral estimation (a characteristic of all methods used with NUS) by quantifying the spectral quality via synthetic signals that are added to the empirical dataset. Recovery of these synthetic signals provides a proxy for the quality of the empirical portion of the spectrum. The central finding is that differences in spectral quality are primarily driven by the strength of bias in the PDF. In addition, a sampling coverage threshold is observed that appears to be connected to the dependence of each NUS method on its random seed. The differences between sampling schemes and biases are most relevant below 20% coverage where seed-dependence is high, whereas at higher coverages, the performance metrics for all of the sampling schemes begin to converge and approach a seed-independent regime. The results presented here show that aggressive sampling at low coverage can produce high-quality spectra by employing a sampling scheme that adheres to a decaying PDF with a bias to a broad range of short evolution times and includes relatively few FIDs at long evolution times.

*Correspondence to schuyler@uchc.edu.

Keywords

nonuniform sampling (NUS); probability density function (PDF); in situ receiver operating characteristic (IROC); random seed dependence

Introduction

In order to study biomolecules by multidimensional NMR that are larger, transiently stable, and/or sparingly soluble while maintaining adequate sensitivity and resolution, the reductions in experiment time afforded by nonuniform sampling (NUS) are essential¹, especially at high magnetic field. Although NUS is increasingly used in routine backbone assignment experiments, NUS is still not widely employed by NMR spectroscopists. This lack of adoption is in part due to the diversity of sampling schemes^{2–8} and spectral analysis methods⁹, differences in the characteristics of NUS spectral estimates versus spectra derived from uniformly sampled (US) data processed with the discrete Fourier transform (DFT), and the absence of consensus best practices.

Nonuniform Sampling Approaches and Limitations

Despite the obstacles to widespread adoption of NUS methods, heuristics for constructing sample schedules are known. For example, higher coverages (i.e. the fraction of samples from a uniform Nyquist grid employed by a NUS scheme) generate smaller and often negligible sampling artifacts but obviously require more experiment time. In contrast, lowering the coverage by NUS to reduce experiment time can produce observable artifacts by violations of the Nyquist theorem^{10–13}. For experiments that have decaying signals, robust sampling schemes must meet at least three important criteria: avoiding aliasing of resonances, sampling long evolution times to ensure discrimination of similar resonance frequencies (i.e. resolution), and sampling short evolution times to ensure adequate signal detection (i.e. sensitivity). These criteria are often addressed by using a probability density function (PDF) to bias the selection² of sampled evolution times. The PDF is a function that defines the relative likelihood of selecting free induction decays (FIDs) at each tuple of evolution times (also referred to as points in sample space or as indels¹⁴) along the indirect dimensions of an experiment. Random sampling corresponds to a uniform, unbiased PDF, whereas any other PDF imposes a sampling bias. Even if a sampling scheme does not explicitly define a PDF, one may be recovered analytically or by simply generating a histogram from an ensemble of sample schedules.

The DFT-Convolution Theorem illustrates how artifacts are induced by NUS in the special case of zero-filling the missing FIDs and applying the Fourier transform¹⁵. Employing non-Fourier methods directly on the NUS data suppresses sampling-induced artifacts, but additional gains are possible by employing sample schedules that contain randomness^{15,16} and utilize higher dimensionality¹⁷, both of which introduce decoherence into the sampling¹⁸. However, incorporating randomness into a sampling scheme has the potential to produce varying sample schedules with a wide range of performance properties, especially at low coverages; we refer to such schemes as “**high-variance**”. In the absence of *a priori* methods that can identify the better sample schedules from a set of candidates, many

sampling schemes have sought to avoid this uncertainty by imposing constraints on how FIDs are selected from within the prescribed PDF. While this conservative approach yields “**low-variance**” schedules that fall within a narrow band of performance, it risks giving up some of the potential gains afforded by randomization.

The time saved by not sampling all FIDs on a Nyquist grid can be spent on collecting repeated scans for the FIDs that are collected. This time-equivalent approach to NUS has been shown to improve sensitivity^{2,19–24}. However, the optimal allocation of sampled evolution times is still the subject of investigation, and the importance of sampling long evolution times to promote resolution in the indirect dimensions of multidimensional experiments is one of the driving factors behind interest in NUS techniques, since excluding the intermediate evolution times from traditional 2D experiments provides the highest time savings. Several techniques for randomly sampling longer evolution times while forcing minimal gap size and reducing random seed dependence^{5,18,24,25} have been suggested. For uniformly sampled datasets, it has been shown that sampling FIDs at evolution times past $1.26T_2$ adds more noise than signal^{26,27}, whereas sampling to an evolution time of πT_2 resolves resonances with frequencies separated by the natural linewidth. Hyberts et al.²⁸ presented a strategy that yielded high sensitivity and resolution by sampling to $.5\pi T_2^*$ followed by extrapolation of the time-domain to $2\pi T_2^*$; their work focused on the Poisson-gap sampling strategy with a sine-weighted PDF⁵, and did not explore different schemes or PDFs.

Challenges for Comparison of Sampling Schemes

Several metrics have been used for studying NUS sampling schemes. The intrinsic signal-to-noise ratio^{6,22} (**iSNR**) provides a measure of the signal to noise ratio (**SNR**) based on the intensity of the signal envelope at the evolution times for each FID in a given sample schedule prior to any data manipulation. Other metrics are derived from the sampling function, a numeric representation of the sample schedule as an array with dimensions that match those of the indirect dimensions of the experiment that contains values of 1s and 0s, to represent FIDs that are collected and skipped, respectively. The point-spread function¹⁵ (**PSF**) is the DFT of the sampling function and consists of a central (zero-frequency) peak whose intensity scales with the number of evolution times sampled (the number of 1's in the sampling function), surrounded by side lobes of smaller intensity, which manifest themselves in a reconstructed spectra via the convolution operation^{15,29}. The peak-to-sidelobe ratio^{17,30} (**PSR**) is derived from the PSF and provides a quantitative measure of expected artifact intensity. These metrics reveal important characteristics of sampling-induced artifacts, but none completely describes the relative performance of sampling schemes. For example, none of the metrics can determine if an artifact falls in a blank region or obscures weaker peaks. In addition, the nonlinearities of NUS reconstruction methods complicate the transferability of properties observed in the time domain to properties expected in the frequency domain (e.g. the SNR in the frequency domain is ill-defined).

The challenges in comparing NUS approaches likely explain why many recent works explore the applications of NUS in specific experiments^{31,32} or seek to characterize only certain sampling schemes without examining other methods^{28,33,34}. In addition to the lack of

robust comparison metrics, the prohibitive combinatorics of such comparisons limits the feasibility of comparing all sampling schemes for all experiment types. Consider a hypothetical three-dimensional experiment based upon a Nyquist grid of 32 samples along each indirect dimension. Given the uniform grid of 1024 indirect points to select from, the number of ways to subsample just 5% of those points is astronomical, $\binom{1024}{51} = \frac{1024!}{973! \cdot 51!}$ or approximately 10^{86} .

Further, different sampling schemes may be constrained to distinct PDFs. Poisson gap sampling uses a PDF that adapts to the coverage level, is strongly biased to initial times, and also ensures that later times are still sampled⁵. Random exponentially biased sampling constrains samples to an exponential decay specified by the user. Finally, quantile sampling can be used with any PDF in principle, and is compatible with several different types of biased PDFs.

A Robust Approach to NUS Evaluation

Previously we showed that adding synthetic signals to an empirical dataset and subsequently testing for their recovery after spectral estimation can be used to assess the sensitivity and resolution of a particular sample schedule and reconstruction method³⁵. Termed *in situ* receiver operating characteristic (**IROC**) analysis, the technique involves adding synthetic signals to a resonance-free region of the dataset. The goal of IROC is to assess how well these synthetic signals can be recovered, thereby providing guidance on how to interpret the empirical signals for which ground truth is not available. While the IROC analysis is based on the synthetic peaks in the blank region, it is important that surrounding empirical peaks are present. Without the empirical peaks and the background of empirical noise, the reconstruction could unfairly benefit from unrealistically low signal sparsity and a lack of complexity (i.e. differentiation between weak peaks and noise); both of which could result in artificial improvements to the spectral quality metrics reported by IROC. The field of compressed sensing, and the work of Donoho in particular, has clearly addressed the relationships between signal sparsity, sampling sparsity, and the phase transition between successful and failed spectral reconstructions.³⁶

This work reports the use of IROC analysis to improve the understanding and design of efficient NUS strategies by conducting a head-to-head comparison of exponentially-biased random sampling² (**ES**), Poisson gap sampling⁷ (**PG**), and quantile sampling¹⁸ (**QS**). All of these sampling schemes employ a biased PDF to select more samples at shorter indirect evolution times. The PG and ES methods employ a random seed in all cases; the QS method is deterministic for one-dimensional schedules (2D NMR) and employs a random seed for higher dimensional schedules (3D-NMR and up). We use IROC to compare these three schemes applied to 3-dimensional protein backbone NMR assignment data (corresponding to NUS in two indirect dimensions). Since QS has a variety of input parameters (including various functions that can be used as a PDF), we conducted a thorough parameter search in order to inform our selection of QS parameters and ensure parity with the ES and PG methods (see Supporting Information). We then compare the three strategies (QS, ES, PG),

employing a weak and strong PDF bias for each scheme, and analyze their relative performance.

Although it is not practical to consider all experiment types, sampling schemes, or reconstruction methods, this work is the most extensive comparison of sampling methods to date, critically comparing three sampling schemes and revealing insights that should help guide spectroscopists considering the use of NUS. This work objectively supports the principle that using decaying PDFs in evolution periods that have decaying signals fundamentally improves the measurement of signal characteristics, which confirms the theoretical insight that any decaying PDF applied to any decaying signal has an intrinsic sensitivity advantage²². Significant differences in the performance of sampling schemes studied here are observed particularly for low coverage, whereas differences among the sampling schemes diminish as coverage increases. Indeed, for the protein 3D-NMR experiments considered in this work, a sampling coverage of 20% is identified as a threshold at which the comparative performance of the schemes begins to converge, and is also a threshold at which the performance of each individual scheme enters a low-variance regime (i.e. seed independent).

Methods

IROC Analysis

The interpretation of an NMR spectrum almost always begins with peak-picking. Distinguishing true weak peaks from noise or artifact is often not possible by inspection of line shape, phase or intensity. For example, it is rare that a single peak-picking threshold can separate all true peaks from the surrounding noise and artifacts. In addition, closely spaced frequency components may not be resolved, resulting in fewer peaks than expected. These factors complicate the selection of an appropriate threshold value for peak picking, the interpretation of resulting peak lists, and the comparison of peak lists derived from the same experiment (e.g. comparisons of NUS sampling schemes or reconstruction methods).

IROC analysis³⁵ is a robust and informative tool that utilizes synthetic peaks as a “ground truth” and continuously varies the peak-picking threshold from the most intense signal down to the noise to produce a quantitative description of spectral quality. The synthetic signals are injected into a resonance-free region identified in a preliminary spectral estimate generated from the empirical dataset. A spectral reconstruction is then performed on the augmented data. The points in an IROC plot are obtained by gradually lowering the peak-picking threshold and plotting the recovery rate (fraction of ground truth peaks detected) versus the false discovery rate (fraction of detected peaks that are not part of the ground truth). The characteristics of the synthetic peaks can be tailored to target specific properties of the empirical signals and the recovery of the synthetic signals informs the interpretation of the empirical portion of the spectrum. The resulting IROC plot is the basis for quantitative metrics that enable comparison of methods with different nonlinearities, and thus are more robust and transferable than SNR. The metrics used in this analysis are briefly defined here and the interpretation of IROC data is more thoroughly discussed in our previous work³⁵. The P_{min} metric (previously reported as the maximum recovery rate at minimum false discovery rate³⁵) gives the percentage of ground truth peaks that are recovered before

encountering the most intense false positive peak. The P_{opt} metric reports the recovery rate that is achieved at the point on the IROC curve that is closest to perfect classifier (i.e. the top left corner of the IROC curve where recovery rate is 100% and false discovery rate is 0%). The P_{min} value informs the “worst case scenario” and the P_{opt} value reveals the “optimal balance” between detection and false peak suppression.

While the IROC approach is robust to many variables, it is important to also comment on possible limitations of the analysis. First, the IROC analysis is embedded within a “carrier” experiment, where the characteristics of such empirical data has an impact on the recovery of the synthetic peaks. Donoho and Tanner^{36,37} have elegantly proved that there is a sharp transition between a successful spectral reconstruction and a failed one, based simply on the relationship between signal sparsity and sampling coverage. It is possible that other spectral features (i.e. dynamic range) may play a role, but these are secondary factors and outside the scope of this analysis. The Donoho and Tanner results provide a strict lower bound for the sampling coverage needed to recover peaks, but poor choices in sample schedule design or in parametrization of the spectral reconstruction technique can cause a failed reconstruction. This work aims to address sample schedule design, but does not address the impact that alternate reconstruction algorithms could have on the results. However, the robust and automated workflow developed here is easily transferable, so that alternate empirical data sets and reconstruction algorithms could be benchmarked.

Empirical Data

The 3D-HNCACB time-domain data of the 87 kDa homotrimer protein *proliferating cell nuclear antigen* (PCNA) was generously supplied³⁸ by Dmitry Korzhnev (UConn Health) for this analysis and was collected according to the standard BioPack ghn_cacb pulse sequence. PCNA is a trimer with each subunit having a molecular weight of 29kDa³⁹. The ^{13}C dimension was collected uniformly out to 48 complex increments at a carrier frequency of 150.6 MHz and a spectral width of 8435 Hz. The ^{15}N dimension was collected uniformly out to 30 complex increments at a carrier frequency of 60.7 MHz and spectral width of 1850 Hz. The ^1H N dimension was collected uniformly out to 1024 complex points at a carrier frequency of 599 MHz and a spectral width of 9058 Hz. The 0 Hz reference values were 44.73 ppm, 118.6 ppm and 4.772 ppm respectively. The ^{15}N dimension was collected as a semi-constant time experiment (the default for the pulse sequence noted above) producing exponentially decaying signals during the real-time period.

Synthetic Peak Lists

Two ground truth peak lists are utilized for the IROC analysis: one is designed to test detection of weak peaks (i.e. a proxy for spectral sensitivity) and one is designed to test detection of closely spaced peaks (i.e. a proxy for spectral resolution). These peak lists are referred to as the “sensitivity peak list” and the “resolution peak list”, respectively, and they occupy mutually exclusive regions of the HNCACB spectra that are free of empirical signals to facilitate independent analyses.

The sensitivity peak list is comprised of 50 single peaks. An estimate of the standard deviation of spectral noise in the preliminary spectrum using NMRPipe⁴⁰ was 0.62.

Synthetic peak intensities were varied randomly from a level lower than the noise estimate (0.27) to approximately 3.2x the noise estimate (1.97). Linewidths were set to values in the range observed for the empirical peaks: approximately 200 Hz on ^{13}C ($t_{\max}=3.6 \times T_2$), 80 Hz on ^{15}N ($t_{\max}=4.1 \times T_2$), and 20 Hz on proton ($t_{\max}=7.1 \times T_2$). The digital resolution along each dimension is at or slightly finer than the linewidths of the empirical and injected peaks. The chemical shifts are distributed in the three dimensions of the HNCACB spectrum according to a Sobol distribution⁴¹, which is quasi-random and results in an even spread of peak locations with minimal overlap of peaks and tails. Sobol distributions are often used to generate uniformly distributed values from a parameter space (e.g. parameter selection for Monte Carlo simulations⁴²). The sensitivity peak list probes spectral sensitivity since it spans peaks which should be easily recovered to peaks comparable in intensity to the noise.

The resolution peak list is comprised of 50 peak pairs. A Sobol distribution is again used to place one peak from each of the pairs with an intensity varied randomly between 11.4x the noise estimate (7.0) to 13x the noise estimate (8.0), which is the approximate intensity of the strongest native peaks in the spectrum. The second peak in each of the pairs is assigned with the same intensity as its partner and its location relative to its partner is randomly varied between 0.024–0.042 ppm in the proton dimension, 0.25–0.6 ppm in the carbon dimension, and 0.25–0.6 ppm in the nitrogen dimension. Linewidths are set as described above. Such separations, intensities, and linewidths were noted empirically as challenging to resolve and hence can be used to test the resolving power of each sampling scheme.

Exponentially-Biased Random Sampling

Exponentially-biased random sampling (**ES**) was first introduced by Barna et al.² with the goal of sampling a higher proportion of FIDs at early evolution times when the SNR is high and a lower proportion of FIDs at long evolution times when the SNR is low, thus improving sensitivity. There has been comparatively less exploration of resolution and the significance of distributing the same relative number of points over longer evolution times^{6,28}. Using ES with a PDF that has a sampling decay rate that matches the decay of the signal envelope is described as envelope-matched sampling. There have been compelling results in support for biasing the PDF at a decay rate faster than that of the signal envelope^{21,44}. In many cases, including in this work, the evolution period of the experiment is not long enough to observe significant signal decay²⁶ and if the ES was matched to the signal envelope, the resulting PDFs would have very little decay for the evolution period. In order to achieve PDFs that decay significantly over the evolution period, the schedules are chosen as if the signal has decayed to ca. 5% by $3 \times T_2$. A representative schedule following this convention for a matched PDF (referred to as **ES1**) is shown in Figure 1A and a schedule that follows a PDF that decays twice as quickly (referred to as **ES2**) is displayed in Figure 1B.

Poisson Gap Sampling

Poisson Gap sampling⁷ (**PG**) is an approach predicated on the observation that large gaps between sampled evolution times along any indirect dimension introduce substantial artifacts in resulting spectral estimates. The PG approach chooses gaps between samples according to a Poisson distribution. To construct multi-dimensional schedules, PG employs a series of 1D sampling vectors, which are then assembled into higher dimensional sampling

schedules. The PG method utilizes a sine function with a variable scaling parameter to bias the gap size selection. A sine parameter value of 1 (referred to here as **PG1**) results in a PDF that is biased to both early and late times, with less dense sampling in the middle of the schedule (Figure 1C). A sine parameter value of 2 (referred to here as **PG2**) results in a PDF that is biased to early times and less dense at later times (Figure 1D). Note that the method of using sine-weighted Poisson deviates in generating the PG schedules does not produce a sinusoidal PDF, but rather one that is a complex function of coverage⁵.

Quantile Sampling

Quantile sampling¹⁸ (**QS**) is an alternative to PG that also minimizes large gaps in sampling coverage. QS defines a PDF along each indirect dimension and then computes quantiles that partition the dimension into intervals of equal area under the PDF. QS generalizes to higher dimensions where the volume under the PDF is partitioned into regions of equal volume. Exactly one sample point is taken from each region, thereby constraining gap sizes and ensuring that some longer evolution time points are still captured as well as providing strong adherence to the user-specified PDF. Quantile sampling is controlled by the following parameters: *type*, which selects the shape of the PDF; *bias*, which can be tuned to compress or expand the distribution of selected points to early or later evolution time points; *jitter*, which controls the size of the random deviate applied to the centroid position of each quantile region (for example, a *jitter* value of 0.75 for two indirect dimensions indicates that each FID is selected from the central $0.75 \times 0.75 = 56\%$ of each quantile region); and *evolution time*, which selects a window from the PDF. The quantile parameter '*evolution time*', which is reported as a multiple of T_2 is a separate entity from conventional evolution. To avoid confusion, we refer to the quantile parameter '*evolution time*' as the evolution time parameter (**ETP**). Larger values of the **ETP** correspond to using the PDF out into the decaying tail; this introduces a strong bias towards low evolution times, similar to the *bias* parameter. Smaller values of **ETP** truncates the PDF to the early portion where it has not substantially decayed yet, thereby selecting the "flatter" part of the PDF and allowing for the selection of FIDs at relatively higher evolution times.

We selected quantile sampling with polynomial bias $(1-t)^2$, which combines favorable features of exponentially and Gaussian-biased PDFs¹⁸. The remaining parameters were selected from the following sets: *bias* $\in \{1.0, 1.5, 2.0\}$, *jitter* $\in \{0.25, 0.5, 0.75\}$, and **ETP** $\in \{1, 1.26, 1.52, 2, 3, 14\}$. All 45 possible combinations of these parameters, referred to as the "**QS Parameter Set**", were used to generate sample schedules. IROC analysis of these schedules is described in the Supplemental Information and the results are used to select specific parameter combinations for QS that produce mild and strong PDFs. Figure 1E displays an example of a QS schedule with a weakly biased PDF (*bias*=1.0, *jitter*=0.75, and **ETP**=1.00); we refer to this parameterization as **QS1**. Figure 1F shows an example of a QS schedule with a strongly biased PDF (*bias*=1.0, *jitter*=0.75, and **ETP**=3.14); we refer to this parameterization as **QS2**. It is worth noting that the survey in the Supplemental Information demonstrates that the *jitter* parameter has no effect on the overall distribution times that are sampled, the *bias* parameter has a mild effect on the distribution of FIDs, and the **ETP** parameter has a strong effect. These observations result in the QS1 and QS2 parameters both using the same values for *bias* and *jitter*, but different values for **ETP**.

Sample Schedule Construction

For each of the sampling approaches (ES1, ES2, PG1, PG2, QS1, and QS2), 100 sample schedules were generated for each of the 4 sampling coverages (7%, 10%, 15%, and 20%). QS and ES sample schedules were generated using *nus-tool*, a utility for making and analyzing sample schedules; *nus-tool* is provided on the NMRbox platform⁴⁵ (see <https://nmrbox.org/registry/nustool>). PG schedules were obtained using a web-based schedule generator⁴⁶. In total, we generated 19,600 sample schedules: for ES we generated 100 exponentially-biased random schedules at each coverage for both ES1 and ES2; for PG we generated 100 schedules at each coverage using PG1 and PG2; and finally for the quantile method we generated 100 schedules for each coverage for each of the 45 combinations in the QS Parameter Set.

Data Processing

The uniformly-sampled master dataset is subsampled according to each sample schedule by eliminating all FIDs not specified in the schedule. The acquisition dimension was apodized with a Gaussian window function of 20Hz, zero filled to 2048 points, processed with the DFT, and then reduced to 1024 complex points. The planes along the indirect dimensions were processed with Maximum Entropy reconstruction⁴⁷ (MaxEnt), to produce a final spectra size of $1024 \times 128 \times 64$ ($^1\text{HN} \times ^{13}\text{C} \times ^{15}\text{N}$). The reconstruction parameters DEF=0.809 and $\lambda=0.76$ were set according to automated procedures⁴³.

Data Analysis

Box and whisker plots are a standard visualization scheme for displaying a collection of data values and will be used to present the IROC results in the following sections. The central portion of each box and whisker plot is a box that spans from the first quartile to the third quartile (i.e. 50% of the observations), thus defining the interquartile range (**IQR**). Values exceeding $1.5 \times$ the length of the IQR above the third quartile or below the first quartile are considered outliers and are displayed as asterisks. The maximum and minimum values (excluding outliers) are displayed as the horizontal line on each whisker (i.e. dashed vertical lines). The median of the entire dataset is represented by a horizontal line through the IQR. Note that because of the properties of the IROC statistics, the values comprising each quartile can be limited to certain discrete values in some cases. Note that the term “quartiles” discussed here is a statistical term making reference to 25% of a population of ordered values, whereas “quantile”, which is used in quantile sampling, refers to equal areas/volumes under a PDF.

Results and Discussion

Relationship between Sensitivity and Resolution

The NUS tests performed here involve subsampling a uniformly-acquired data set and are therefore not time-equivalent to the original uniform acquisition. Such subsampling incurs losses in spectral sensitivity as coverage decreases, as previously described²⁶, which could potentially play a confounding role in this work in trying to understand how the NUS sampling schemes affect spectral resolution. For example, Figure 2 illustrates the recovery of

two synthetic peaks separated by 0.038ppm along the proton dimension. Without noise the pair can be discriminated at 2%, 7%, and 100% coverage (Panels A,B,C). If background noise from a signal free region is included, the peak pair is obscured by artifacts at 2% coverage (Panel D). At 7% coverage (Panel E), the separation between the peaks is apparent, yet noise and artifacts prevent reliable discrimination of the true peak locations. The two signals are easily recovered with high fidelity when uniform sampling is applied (Panel F). The demonstration in Figure 2 shows the inherent dependence of resolution on sensitivity. The intensity of the ground truth peaks used to test resolution must therefore be sufficient to mitigate the potential confounding role of inadequate sensitivity on measuring resolution.

Evaluating Sampling Schemes

A quantitative comparison of ES, PG, and QS sampling across a range of sampling coverages and PDF strengths is performed using two IROC analyses employing different sets of ground truth signals used to augment the empirical data: one to probe spectral sensitivity and one to probe spectral resolution. In the IROC sensitivity analysis, the master peak list contains signals that span from the noise up to the intensity of the weakest peaks in the empirical data. This intensity range is determined by peak picking the empirical data and is a highly robust parameter. The synthetic peak intensities simply need to span a sufficient range so that IROC recovers some of the peaks. If the synthetic peak intensities are set too high all peaks would be recovered and if synthetic peak intensities are set too low no peaks would be recovered; neither of these scenarios have diagnostic merit. In the IROC resolution analysis, the master peak list contains overlapping pairs of peaks. These comparisons are performed for each sampling scheme using both a mildly biased PDF (ES1, PG1, and QS1) and a strongly biased PDF (ES2, PG2, QS2). The following sections provide an analysis of the underlying PDFs, which serves as a foundation for interpreting the IROC results.

Features of the PDFs

The PDFs for ES and QS are defined inputs for each sampling scheme. In contrast, the PDF for PG is not directly defined, but must rather be determined *post hoc*, as shown in the careful analysis of gap sampling by Worley and Powers⁵. Figure 3 shows a comparison of the PDFs for ES, PG, and QS. Panels A and B of the top row are normalized to the highest sampling probability for each PDF and show how PG varies the *shape* of its PDF according to its sine parameter (PG1 for sine weight 1 vs PG2 for sine weight 2) and the *strength* of its PDF according to the sampling coverage (e.g. the bundle of PDFs for PG2). The area-normalized PDFs in the second row show that PG2 (Panel C) has the strongest bias at low evolution times and an almost flat bias at all other evolution times (an “L” shaped PDF). PG1 (Panel D) has extreme bias towards the very shortest *and* the very longest evolution times, but is nearly flat across the majority of the midrange (a “U” shaped PDF with a very wide base). In contrast, the ES and QS schemes, regardless of the strength of their bias, show a more consistent and gradual PDF across the entire evolution time range (Panels E and F). So, while the bias is stronger at the lowest evolution times for PG, it is restricted to a very short range of evolution times. The broader range of times that receive a sampling bias in QS and ES actually result in more FIDs being selected in the lowest quadrant of the resulting sample schedules. This is evidenced comparing the lower left corners of the plots

in Figure 1: ES2 (Fig 1B) and QS2 (Fig 1F) have far more FIDs selected at lowest evolution times than PG2 (Fig 1D).

QS guarantees very strong adherence to the PDF, even for small coverages and also has the largest variety of parameters for customizing the PDF. For these reasons, the quantile method was selected to enable testing of a large space of PDF shapes to characterize how the PDF shape informs the IROC analysis. The quantile method also provides a flexible platform to optimize the use of a random variate (i.e. *jitter*) in constructing NUS schedules, an approach which has been suggested by several investigations recently^{18,24,25}. The complete IROC survey of QS parameters is presented in the Supplemental Materials.

Sensitivity Analysis

The sensitivity analysis is performed on the HNCACB data augmented with the sensitivity ground truth peak list. The primary data are subsampled with 100 instances of each sampling scheme and reconstructed with MaxEnt, as previously described. Figure 4 shows the P_{opt} data generated by the corresponding IROC analysis. As expected, all sampling methods show improved sensitivity as sampling coverage is increased or as the strength of the PDF bias is increased. It is expected that stronger biases (ES2, PG2, and QS2) would show the best sensitivity performance, but only QS2 and ES2 separate themselves from the other sampling schemes. Despite the strong bias of PG2, its relatively narrow band of low evolution times that are selected (i.e. the long flat tail of the “L”-shaped PDF discussed in the previous section) is insufficient to provide sensitivity similar to the other strongly-biased methods. Even as sampling coverage increases to 20%, where the P_{opt} values begin to converge, ES2 and QS2 maintain their advantage because the IQRs begin to narrow. In other words, as the mean P_{opt} values get closer together, their distributions also narrow, thereby maintaining a performance difference. These trends are also observed in the P_{min} values in Figure S17.

Resolution Analysis

The resolution analysis is performed on the HNCACB data augmented with the resolution ground truth peak list. The same sample schedules utilized in the sensitivity analysis are also employed here, as well as the same MaxEnt reconstruction workflow. Figure 5 displays box-and-whisker plots of the P_{opt} values calculated from the IROC analysis. As expected, these resolution results are largely the inverse of those observed in the sensitivity analysis. Specifically, the sampling schemes that capture the most FIDs at higher evolution times do best. These top performers are all of the weakly-biased schemes (e.g. ES1, PG1, and QS1) as well as PG2, whose long flat PDF provides some collection of higher evolution-time FIDs. The variation among these methods is most pronounced at the lowest sampling coverages and begins to converge at the higher coverages. The same trends in P_{min} values are observed in Figure S18.

It is somewhat surprising that the PG2, QS1, and ES1 schemes reproduce nearly the same resolution as the PG1 scheme delivers, given that the PDF for the PG1 scheme was fundamentally different than all other methods with its spike at long evolution times. This result suggests that only a few long evolution times are needed to promote resolution,

whereas the sensitivity analysis shows that improvements come with accumulation of sampling at the low evolution times.

Seed Variance of Sampling Schemes

Does seed dependence matter? The IROC analysis is performed for sensitivity and for resolution and it is reported with P_{min} data and P_{opt} data (Figures 4, 5, S17, S18). In each of these figures, there are panels for each of the four sampling coverages, showing the box and whisker plots for each of the three sampling schemes at two bias strengths. Across the 16 panels that present this data, there is only one instance where the distribution of values from one sampling method completely outperforms the distribution of another sampling method at the same coverage (Figure 5, resolution analysis, P_{opt} metric: PG1 > ES2 at 7%). With the exception of this one case, there is overlap between the whiskers of all pairings of box plots for each sampling coverage. Even sampling schemes that are designed to be low-variance (i.e. PG and QS) are not without overlap. We conclude that even though some methods have been shown to perform better than others on average, there are random seeds that may produce schedules contrary to the expected trend. This observation does not eliminate the need for careful sample schedule design, but rather puts an emphasis on developing techniques that can make an *a priori* distinction between the sample schedules at the higher and lower ends of the performance distributions.

The number of ways to build a sample schedule of a particular sampling coverage reaches its maximum at 50% coverage. This suggests that sampling schemes should be low variance at very low (<10%) and very high (>90%) coverages, but show the most variation around 50% coverage. However, just because there are more ways to select a sample schedule at 50% coverage doesn't mean the ensemble of schedules demonstrates any greater variability than the smaller ensemble of schedules available at low or high coverages. In fact, observation shows that spectral quality is most variable at the lowest sampling coverages and continues to improve and decrease variability as coverage increases. This result can be attributed to the PDF. At very low sampling coverages, it is hard to strictly adhere to a continuous PDF; there simply aren't enough FIDs selected to represent the PDF. As sampling coverage increases, schedules begin to adhere to their underlying PDFs, but as schedules continue to increase coverage, there are exclusion effects (i.e. each additional FID selected cannot be placed where the sampling scheme dictates if that FID has already been selected). This results in uniform tracts in the sampling. While it has been shown that uniform tracts improve spectral quality^{11,18,48}, their presence indicates that sampling has likely deviated from the desired PDF. In fact, in the limit as coverage approaches 100%, all sampling schemes produce uniform tracts and approach a uniform PDF. All of the factors discussed here reveal a subtle dependence of PDF on coverage, which is at the heart of seed dependence.

Conclusions

This study establishes a rigorous approach based on IROC for quantifying the performance of NUS sampling schemes and employs the method to characterize the operational properties of exponentially biased sampling (ES), Poisson gap sampling (PG), and quantile sampling (QS). The NMR community has relied on several guiding heuristics for the design

of sampling schemes and these tenets are confirmed. Specifically, strong bias to shorter evolution times enhances sensitivity and sampling of longer evolution times improves resolution. These heuristics may be clearly derived in the context of uniform sampling, but are less concrete in the realm of nonuniform sampling. While sampling schemes often define themselves by how they choose to select FIDs through gap management and randomness, the results here show that the “microstructure” of specific sampling schemes is largely irrelevant, and that the underlying PDFs are the primary factors in determining performance. The results of this analysis are directly encoded in the area-normalized PDF plots of Figure 3. Specifically, only a few long evolution times are needed to enhance resolution, whereas gains in sensitivity come with accumulation of sampling over a broad range of short evolution times.

It is through an appreciation of the PDF that more subtle observations are given their foundation. In particular, we find that a PDF with a strong bias towards the shortest evolution times (i.e. the “L” shaped PDF produced by PG2 and discussed with Figure 3) can actually lose sensitivity if the bias is so strong as to only increase sampling at relatively few of the lowest evolution times. In contrast, biased PDFs with more gradual decay capture a broader window of short evolution times and produce better sensitivity. Long, flat sampling densities at long evolution times, as in PG2, offer modest benefits in resolution, but make a substantial sacrifice in sensitivity. These results demonstrate that optimal sensitivity is achieved when a certain threshold of sampling coverage is met at the short evolution time regime. Further, these results are all attenuated by increased sampling coverage. In other words, the principles observed here are most significant when NUS sampling is aggressively pushed to lower coverages. In fact, the concepts derived here are corroborated in the application of QS and PG sampling to 1,1-ADEQUATE experiments, which demonstrates that QS schedules can easily be designed to have more sensitivity than PG2 (manuscript from Rovnyak, Craft, et al. under review at *Magnetic Resonance in Chemistry*).

Low sampling coverage is the regime where most sampling methods struggle to adhere to their defining PDF and exhibit strong seed-dependence. This is why the community seems to favor low-variance methods. We believe this conservative approach misses an opportunity for improvement. With proper quantitative tools, such as the ones employed here, it is possible to analyze specific instances of sampling schedules and determine if they are at the high-performance end of the expected distribution or at the lower end of the performance distribution, *a priori*. In other words, variability can be utilized as another parameter that can be leveraged to improve performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grants R21-GM104517, P41-GM111135, R01-GM123249 and an NIDCR predoctoral fellowship to MZ (F30-DE026353). We thank Dmitry Korzhnev for supplying the PCNA time domain HNCACB data. During preparation and review of this manuscript, a Monte Carlo analysis demonstrating improved signal metrics in time domain analyses of exponentially weighted NUS with decaying signals has also been presented (Prof. L. Mueller, ENC 2018, Orlando, Florida and Private Communication).

References

1. Billeter M Non-uniform Sampling in Biomolecular NMR. 68, 65–66 (2017).
2. 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* 73, 69–77 (1987).
3. Barna JCJ & Laue ED Conventional and exponential sampling for 2D NMR experiments with application to a 2D NMR spectrum of a protein. *J. Magn. Reson* 75, 384–389 (1987).
4. Schmieder P, Stern AS, Wagner G & Hoch JC Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain. *J Biomol NMR* 4, 483–490 (1994). [PubMed: 8075537]
5. Worley B & Powers R Deterministic multidimensional nonuniform gap sampling. *J Magn Reson* 261, 19–26, doi:10.1016/j.jmr.2015.09.016 (2015). [PubMed: 26524650]
6. Palmer MR, Wenrich BR, Stahlfeld P & Rovnyak D Performance tuning non-uniform sampling for sensitivity enhancement of signal-limited biological NMR. *J Biomol NMR* 58, 303–314, doi:10.1007/s10858-014-9823-5 (2014). [PubMed: 24682944]
7. Hyberts SG, Takeuchi K & Wagner G Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* 132, 2145–2147, doi:10.1021/ja908004w (2010). [PubMed: 20121194]
8. Worley B Subrandom methods for multidimensional nonuniform sampling. *J Magn Reson* 269, 128–137, doi:10.1016/j.jmr.2016.06.007 (2016). [PubMed: 27301071]
9. Bostock M & Nietlispach D Compressed sensing: Reconstruction of non-uniformly sampled multidimensional NMR data. *Concepts in Magnetic Resonance Part A* 46A, e21438 (2017).
10. Mobli M & Hoch JC Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR. *Prog Nucl Magn Reson Spectrosc* 83C, 21–41, doi:10.1016/j.pnmrs.2014.09.002 (2014).
11. Maciejewski MW, Qui HZ, Rujan I, Mobli M & Hoch JC Nonuniform sampling and spectral aliasing. *J Magn Reson* (2009).
12. Bretthorst GL Nonuniform sampling: Bandwidth and Aliasing. *Concepts Magn. Reson* 32A, 417–435 (2008).
13. Manassen Y & Navon G Nonuniform sampling in NMR experiments. *Journal of Magnetic Resonance* 79, 291–298 (1988).
14. Monajemi H, Donoho DL, Hoch JC & Schuyler AD Incoherence of Partial-Component Sampling in multidimensional NMR. *arXiv* <https://arxiv.org/abs/1702.01830>(2017).
15. Mobli M, Maciejewski MW, Schuyler AD, Stern AS & Hoch JC Sparse sampling methods in multidimensional NMR. *Phys Chem Chem Phys* 14, 10835–10843, doi:10.1039/c2cp40174f (2012). [PubMed: 22481242]
16. Kazimierczuk K, Zawadzka A, Kozminski W & Zhukov I Random sampling of evolution time space and Fourier transform processing. *J Biomol NMR* 36, 157–168 (2006). [PubMed: 17031529]
17. Schuyler AD, Maciejewski MW, Stern AS & Hoch JC Nonuniform sampling of hypercomplex multidimensional NMR experiments: Dimensionality, quadrature phase and randomization. *J Magn Reson* 254, 121–130, doi:10.1016/j.jmr.2015.02.015 (2015). [PubMed: 25899289]
18. Craft DL, Sonstrom RE, Rovnyak VG & Rovnyak D Nonuniform sampling by quantiles. *J Magn Reson* 288, 109–121, doi:10.1016/j.jmr.2018.01.014 (2018). [PubMed: 29453083]
19. Barna JCJ, Laue ED, Mayger MR, Skilling J & Worrall SJP Reconstruction of phase sensitive 2D NMR spectra using maximum entropy. *Biochem. Soc. Trans* 14, 1262 (1986).
20. Kumar A, Brown SC, Donlan ME, Meier BU & Jeffs PW Optimization of two-dimensional NMR by matched accumulation. *J. Magn. Reson* 95, 1–9 (1991).
21. Rovnyak D, Sarcone M & Jiang Z Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. *Magn Reson Chem* 49, 483–491 (2011). [PubMed: 21751244]
22. Palmer MR et al. Sensitivity of Nonuniform Sampling NMR. *J. Phys. Chem. B* 119, 6502–6515 (2015). [PubMed: 25901905]

23. Hyberts SG, Robson SA & Wagner G Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J Biomol NMR* 55, 167–178, doi:10.1007/s10858-012-9698-2 (2013). [PubMed: 23274692]
24. Mobli M Reducing seed dependent variability of non-uniformly sampled multidimensional NMR data. *J Magn Reson* 256, 60–69, doi:10.1016/j.jmr.2015.04.003 (2015). [PubMed: 26004701]
25. Kazimierczuk K, Zawadzka A & Kominski W Optimization of random time domain sampling in multidimensional NMR. *Journal of Magnetic Resonance* 192, 123–130, doi:10.1016/j.jmr.2008.02.003 (2008). [PubMed: 18308599]
26. Rovnyak D, Hoch JC, Stern AS & Wagner G Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *J Biomol NMR* 30, 1–10 (2004). [PubMed: 15452430]
27. Rovnyak D The past, present, and future of 1.26T₂. *Concepts in Magnetic Resonance Part A* 47A, e21473 (2019).
28. Hyberts SG, Robson SA & Wagner G Interpolating and extrapolating with hmsIST; seeking a $t_{\max}$ for optimal sensitivity, resolution and frequency accuracy. 68, 139–154 (2017).
29. Hoch JC, Maciejewski MW, Mobli M, Schuyler AD & Stern AS Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR. *Acc Chem Res* 47, 708–717, doi:10.1021/ar400244v (2014). [PubMed: 24400700]
30. Lustig M, Donoho D & Pauly JM Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med* 58, 1182–1195, doi:10.1002/mrm.21391 (2007). [PubMed: 17969013]
31. Kakita VM & Hosur RV Non-Uniform-Sampling Ultrahigh Resolution TOCSY NMR: Analysis of Complex Mixtures at Microgram Levels. *Chemphyschem* 17, 2304–2308, doi:10.1002/cphc.201600255 (2016). [PubMed: 27124764]
32. Linnet TE & Teilum K Non-uniform sampling of NMR relaxation data. *J Biomol NMR* 64, 165–173, doi:10.1007/s10858-016-0020-6 (2016). [PubMed: 26847574]
33. Aoto PC, Fenwick RB, Kroon GJ & Wright PE Accurate scoring of non-uniform sampling schemes for quantitative NMR. *J Magn Reson* 246, 31–35, doi:10.1016/j.jmr.2014.06.020 (2014). [PubMed: 25063954]
34. Sidebottom PJ A new approach to the optimisation of non-uniform sampling schedules for use in the rapid acquisition of 2D NMR spectra of small molecules. *Magn Reson Chem* 54, 689–694, doi:10.1002/mrc.4444 (2016). [PubMed: 27160788]
35. Zambrello MA, Maciejewski MW, Schuyler AD, Weatherby G & Hoch JC Robust and transferable quantification of NMR spectral quality using IROC analysis. *J Magn Reson* 285, 37–46, doi:10.1016/j.jmr.2017.10.005 (2017). [PubMed: 29102819]
36. Monajemi H, Jafarpour S, Gavish M, Donoho DL & Collaboration SC Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. *Proc Natl Acad Sci U S A* 110, 1181–1186, doi:10.1073/pnas.1219540110 (2013). [PubMed: 23277588]
37. Donoho D & Tanner J Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos Trans A Math Phys Eng Sci* 367, 4273–4293, doi:10.1098/rsta.2009.0152 (2009). [PubMed: 19805445]
38. Pustovalova Y, Maciejewski MW & Korzhnev DM NMR mapping of PCNA interaction with translesion synthesis DNA polymerase Rev1 mediated by Rev1-BRCT domain. *J Mol Biol* 425, 3091–3105, doi:10.1016/j.jmb.2013.05.029 (2013). [PubMed: 23747975]
39. Moldovan GL, Pfander B & Jentsch S PCNA, the maestro of the replication fork. *Cell* 129, 665–679 (2007). [PubMed: 17512402]
40. Delaglio F et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277–293 (1995). [PubMed: 8520220]
41. Sobol I Vol. 7 86–112 (*USSR Computational Mathematics and Mathematical Physics*, 1967).
42. Burhenne S, Jacob D, Henze G 1816–1823 (*Proceedings of Building Simulation*, 2011).
43. Mobli M, Maciejewski MW, Gryk MP & Hoch JC Automatic maximum entropy spectral reconstruction in NMR. *J. Biomol. NMR* 39, 133–139 (2007). [PubMed: 17701276]
44. Schuyler AD, Maciejewski MW, Arthanari H & Hoch JC Knowledge-based nonuniform sampling in multidimensional NMR. *J Biomol NMR* 50, 247–262 (2011). [PubMed: 21626215]

45. Maciejewski MW et al. NMRbox: A Resource for Biomolecular NMR Computation. Biophys. J in press (2017).
46. Hyberts S & Wagner G Poisson Gap schedule generator, <http://gwagner.med.harvard.edu/intranet/hmsIST/gensched_new.html> (
47. Hoch JC, Stern AS, Donoho DL & Johnstone IM Maximum Entropy Reconstruction of Complex (Phase-Sensitive) Spectra. J. Magn. Reson 86, 236–246 (1990).
48. Mobli M & Miljenovi TM Framework for and evaluation of bursts in random sampling of multidimensional NMR experiments. J Magn Reson 300, 103–113, doi:10.1016/j.jmr.2019.01.014 (2019). [PubMed: 30738271]

Figure 1: Examples of each Sampling Scheme.

Sample schedules at 15% coverage showing the weakest and strongest bias employed by each scheme. (A) ES1, (B) ES2, (C) PG1, (D) PG2, (E) QS1, and (F) QS2.

Figure 2: Resolution is Intrinsically Dependent on Sensitivity.

An overlapping synthetic peak pair located at $\{^{13}\text{C}=50.4180\text{ppm} / ^{15}\text{N}=131.4800\text{ppm} / ^1\text{HN}=11.9420\text{ppm}\}$ and $\{^{13}\text{C}=50.4180\text{ppm} / ^{15}\text{N}=131.4800\text{ppm} / ^1\text{HN}=11.9040\text{ppm}\}$ are injected into either an empty dataset (left column) or into the background noise of a signal-free region of the HNCACB data (right column) to produce time data on a grid of size $48 \times 30 \times 1024$ ($^{13}\text{C} \times ^{15}\text{N} \times ^1\text{HN}$). These uniformly sampled data sets are subsampled with an exponentially biased sampling scheme at both 2% coverage (A and D) and 7% coverage (B and E); the 100% coverage uniform data is also considered (C and F). The proton dimension is zero-filled to 2048 points, processed by FFT, and reduced to 1024 points. The $^{13}\text{C} \times ^{15}\text{N}$ planes are reconstructed by MaxEnt to 128×64 points using 1000 iterations, with parameter values⁴³ $\text{DEF} = 0.809$, and $\lambda = 0.76$. The spectral reconstructions in each panel show a contour plot on the left side for the slice taken at a nitrogen chemical shift of 131.932ppm; the vertical axes (left edge of the Figure) show carbon chemical shift in ppm and the horizontal axes (bottom edge of the Figure) show proton chemical shift in ppm. A slice is taken through each peak pair (indicated with a solid line) and shown on the right side of

each panel, where vertical axes (right edge of the Figure) show peak intensity in a.u. and the horizontal axes (bottom edge of the Figure) show H chemical shift in ppm.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. PDFs for Sampling Schemes.

This figure shows plots of the functions used to bias the sampling of ES, PG, and QS. The horizontal axes show evolution time and the vertical axes show probability. Panels A and B are scaled so that the functions have a probability of 1 at $t=0$ (strictly speaking, panels A and B are thus not PDFs). The scaling in panels A and B helps to make differences in shape more apparent. Panels C-F are normalized by area and thus meet the definition of a PDF. These panels show the relative differences in sampling propensity at short, medium, and long evolution times. For example: PG1 @ 20% (Panel B) reaches a probability of 1 at both $t=0$ and $t=t_{max}$, whereas it only reaches a probability of 0.4 in Panel D, where the function is normalized for area. Panel B illustrates the sampling distribution for PG1 and facilitates comparison with the other sampling schemes shown in Panel A, whereas panel D shows how

PG1 is dependent on coverage (i.e. as coverage increases, sampling propensity for the lowest and highest evolution times is diminished as propensity for mid-range evolution times is increased).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4: Sensitivity Analysis by IROC.

Panels represent sampling coverages of 7%, 10%, 15%, and 20% and each one shows P_{opt} values as box and whisker plots for 100 repetitions of each sampling scheme at mild (ES1, PG1, QS1) and strong (ES2, PG2, QS2) biases.

Figure 5: Resolution Analysis by IROC.

Panels represent sampling coverages of 7%, 10%, 15%, and 20% and each one shows P_{opt} values as box and whisker plots for 100 repetitions of each sampling scheme at mild (ES1, PG1, QS1) and strong (ES2, PG2, QS2) biases.