SAN plot: A graphical representation of the signal, noise, and artifacts content of spectra

Kirill F. Sheberstov2 | Eduard Sistaré Guardiola1 | Marion Pupier1 | Damien Jeannerat1

1 Department of Organic Chemistry, University of Geneva, Geneva, Switzerland
2 Institute of Physics, Johannes Gutenberg-Universität, Mainz, Germany

Correspondence
Damien Jeannerat, Department of Organic Chemistry, University of Geneva, Geneva 1211, Switzerland.
Email: damien.jeannerat@unige.ch

Funding information
Swiss National Science Foundation (Bern), Grant/Award Number: 165725; Chemistry section; Swiss National Science Foundation, Grant/Award Number: 200020-126525

Abstract
The signal-to-noise ratio is an important property of NMR spectra. It allows to compare the sensitivity of experiments, the performance of hardware, etc. Its measurement is usually done in a rudimentary manner involving manual operation of selecting separately a region of the spectrum with signal and noise, respectively, applying some operation and returning the signal-to-noise ratio. We introduce here a simple method based on the analysis of the distribution of point intensities in one- and two-dimensional spectra. The signal/artifact/noise plots, (SAN plots) allows one to present in a graphical manner qualitative and quantitative information about spectra. It will be shown that besides measuring signal and noise levels, SAN plots are also quite useful to visualize and compare artifacts within a series of spectra. Some basic properties of the SAN plots are illustrated with simple application.

KEYWORDS
comparison of spectra, signal-to-noise ratio, SINO, SNR

INTRODUCTION

Measuring the signal-to-noise ratio (SNR)$^1$ of a spectrum is quite useful when evaluating the performances of a spectrometer, comparing the sensitivity of experiments, setting automatically the contour levels or peak-picking threshold,$^{[2,3]}$ etc. The standard procedure to determine the SNR of a spectrum consists in selecting a region with signals, another one with noise only, and to call an algorithm calculating

$$SNR = \frac{S^{\text{max}}}{2\sigma_N},$$

where $S^{\text{max}}$ is the intensity of the highest signal, and $\sigma_N$ is root mean-square of the intensities in the noise region.

This measure can be automatized provided one can determine with certainty that a region of the spectrum will contain exclusively noise. The SNR is commonly measured in 1D spectra, but for 2D spectra, the measure in the direct and indirect dimensions may differ if, for example, $t_i$ noise is observed in the indirect dimension.

Beside signal and noise, spectra may also show so-called “artifacts.” This porte-manteau word can refer to anything unwelcomed. They often refer to signals caused by spectrometer imperfections such as $t_i$ noise in 2D spectra,$^{[4-6]}$ or quadrature imperfection,$^{[7]}$ etc. But sometimes, the source is an intrinsic properties of the spin systems—typically, strong coupling distortions or isotropic distributions adding “satellites” peaks to a main signal ($^{13}$C satellites, for example). Spectral artifacts can also be due to signal processing, typically when attempting to
increase spectral resolution with Lorentz-to-Gauss transformation. For a chemist, the signals caused by a degradation products of the main component of a sample may be called “artifacts” even if their presence is quite natural and unavoidable. It is therefore very difficult to make general claims about artifacts. We will therefore define them in a rather vague terms only to partially distinguish them from the noise and “proper” signals. The key property we shall exploit here is that artifacts are either single or multiple minor duplicates of the main signals with intensities that are intermediate between the real signals and the noise.

2 | SIGNAL-ARTIFACTS-NOISE ANALYSIS

We present here the so-called “signal-artifacts-noise (SAN) plot” as a visualization tool facilitating the comparison of series of spectra with respect to signal, artifacts, and noise. Combined with a simple statistical analysis, it provides an estimate of the SNR of a one-dimensional, or a two-dimensional spectrum—the latter being taken as a whole, rather than just a specific F1 or F2 cross-section. We will also show how SAN plots can be used to characterize artifacts and compare their levels within series of spectra.

2.1 | The SAN plot

The SAN plots are obtained by sorting either the positive and negative points of a spectrum according to their decreasing absolute values—we will call these series $S^+$ or $S^−$, respectively—and plotting, either one of them using logarithmic scales on both axes (see $S^+$ in blue in Figure 1). The logarithmic scale of the ordinate is used to facilitate the observation of changes in intensities over a broad dynamic range: It allows to easily observe variations at the level of the noise in the presence of much larger signals. The logarithmic scale of the abscissa is used to give more importance to the signals (usually covering a small percentage of the spectrum) relative to the noisy part of the spectrum, which is presented in a compact manner.

The SAN plot of the positive points ($S^+$) typically shows two main regions. The first corresponds to the distribution of the points of the signals, the second corresponds to the noise of the spectrum. The dashed gray line separates the two regions in Figure 1. Broadly speaking, artifacts—when present—will appear between these two regions. When spectra include only positive signals, the SAN plot of the negative points ($S^−$) will show only noise and artifacts when present (see Figure S2).

In the “Signal” region of Figure 1, two main bumps can be observed. They correspond to the central and the outer peaks of the triplet, respectively, whereas the small singlet (see inset) produces no visible feature on the SAN plot (see Supporting information for a more detailed analysis of the shape of the SAN plot of a pure Lorentzian peak.). The measure of $S_{\text{max}}$ simply consists in taking the level of the left-most point of the SAN plot. The right part of the SAN plot shows the distribution of the noise, which can be used to determine the noise level of the spectrum and, consequently calculate the SNR according to Equation (1).

2.2 | Measuring the noise level in SAN plots

Assuming that the spectral noise corresponds to a white Gaussian distribution, it has a probability density function

$$p_G(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}},$$

where $\mu$ is the mean value, which is close or equal to zero when the baseline offset is corrected, and $\sigma$ is the standard deviation, that is, the noise level.

A very good approximate of the noise level can be obtained by adjusting the $\sigma$ value of white Gaussian noise until its SAN plot matches the one of the experimental spectrum. The details of the algorithm are described in the Supporting information.

This approach was tested on diverse types of spectra and provided values that are extremely close to the ones calculated with SINO function of the Bruker software. When applied to the synthetic data of Figure 1, the SNR...
deviates very slightly from 100, the expected value, only because of the intrinsic random nature of noise.

Note that when the spectra are processed in the “magnitude mode” (or “absolute value” mode), our program matches the SAN plot to the Rayleigh instead of Gaussian distribution. In this case, the value of the SNR returned by our function has a different meaning because $\sigma_N$ does not correspond to the standard deviation of the noise as for the Gaussian distribution. (see Supporting information for more details).

One obvious advantage of this method to measure the SNR is that it does not require a prior knowledge of the location of signal and noise, respectively, making it quite appropriate for a fully automated workflow. Hybrid methods using our approach as a first step to only locate spectral regions with noise before applying other techniques for the calculation of the SNR represent a possible alternative.

### 3 | PROPERTIES OF SAN PLOT IN ONE- AND TWO-DIMENSIONAL NMR

Before exploiting SAN plot in Section 4, we should present in this section the effect of signal averaging and apodization on the SAN plots.

#### 3.1 | Impact of increasing the number of acquired transients on signal and noise

It is well known that when accumulating FIDs, the SNR in the resulting spectra increases with the square root of the number of scans (NS). This is because the signal increases linearly with the NS, whereas the noise increases by a factor equal to the square root of the NS. The net effect is an increase by a factor

$$\frac{NS}{\sqrt{NS}} = \sqrt{NS}. \quad (3)$$

This can be verified using the SAN plots of spectra acquired using one and two scans per time increment (compare the blue and orange lines in Figure 2). When multiplying the level of the SAN function of the spectrum acquired with $NS = 1$ by $\sqrt{2}$ and 2 (respectively, the dotted and dashed black lines in Figure 2), one can observe that the level of the signal of the spectrum acquired with $NS = 2$ reaches the expected levels: On the left part of the plot where signals are densely packed, the SAN function reaches the level of the dashed line, that is, double in intensity, whereas the right part, corresponding to the noise, reaches the level of the dotted line, that is, increased by a factor $\sqrt{2}$.

The phase cycling obtained with number of scans $= 2$ also reduces the quadrature artifacts as shows the SAN plot of Figure S1 in the Supporting information.

#### 3.2 | Influence of apodization on the SAN plots

The effect of window functions (apodization) applied prior to the Fourier transform can also be studied with the SAN plots. Figure 3 shows how the exponential multiplication of the FID of a DEPT-135 spectrum with increasing line broadening factors decreases the signal

---

**FIGURE 2** Comparison of the SAN plots of 2D spectra recorded with one (blue) and two (orange) scans per FID. The black dotted and dashed lines corresponding to $\sqrt{2}$ and 2 times $S+$ (number of scans = 1) facilitate the comparison with the $S+$ (number of scans = 2). The increase in signal and noise levels were, respectively, 2.04 and 1.51 when going from one to two scans per increment.

**FIGURE 3** Influence of apodization on the SAN plots of a DEPT-135 spectrum. The SNR were 233, 257, 246, 209, 158, and 115 with line broadening factors $LB = 0, 0.5, 1, 2, 4,$ and $8$ Hz, respectively. For spectra obtained using regular and squared sine bells shifted by $\pi/2$, the SNR was 257.
amplitude. This can be observed by a lowering of the “signal” part of the SAN plots. The broadening of the signals results in a shift toward the right of the steepest part of the plot (see the horizontal arrow). This reflects the fact that broad signals cover a relatively larger part of the spectrum. The noise level also decreases. For values of the line-broadening factor LB set below the “match” condition—that is, for LB smaller than the signal’s width—the noise decreases faster than signal resulting in the expected increase in SNR level. Figure 3 shows that the best SNR is observed for LB = 0.5 Hz, which is consistent with the natural linewidth of about 0.7 Hz.

4 | ANALYSIS OF SPECTRAL ARTIFACTS

We should consider here application of SAN plots to benchmark pulse sequences, processing technique and start with an example of use to identify $t_1$ noise in 2D spectra.

4.1 | Artifacts in 2D NOESY spectra caused by temperature instabilities

Variation of chemical shifts or signal intensity during the acquisition of 2D experiments produce artifacts. These variations may be due to temperature instabilities, problems with the lock system, or perturbation of the automatic shimming process caused by pulse-field gradients, etc. The consequence of these instabilities is that they may distort the lineshape and cause $t_1$-noise, which consists in sets of spurious peaks appearing at seemingly random positions along the F1 dimension. When spectra require the observation of small signals among much larger ones, such as for 2D NOESY spectra, where cross-peaks have just a few percent of the intensity of the diagonal peaks, these instabilities can render the spectra completely unusable.

Figure 4 shows that SAN plot can discriminate spectra suffering from temperature instabilities from normal ones. The SAN plot of a NOESY spectrum acquired using relatively stable temperature conditions is shown in blue and serves as a reference. In the two others (measured at slightly different temperature, but this is not relevant here because the artifacts were due to erratic temperature control), the noise and the signals were quite similar in intensities (observe the perfect overlap at the left and right parts of the SAN plots), but the middle part of the SAN plots—where artifacts are expected to appear—the level reflects the intensity of the $t_1$-noise observed in the spectra.

4.2 | Comparison of pulse sequences

Comparing the sensitivity and the quality of NMR spectra is quite important in NMR methodology developments. The focus is often set on the sensitivity, which is easy to compare within series of experiments. But in some cases, the true measure of the quality of a spectrum depends on the ratio between the level of the signals and the artifacts. An example is provided by the F1-decoupled CLIP-COSY experiment. The spectra it produces are quite powerful to analyze complex mixtures of compounds such as carbohydrates because they eliminate scalar coupling structures in the vertical F1 dimension (see Figure 5). But they suffer not only from diverse type of artifacts caused by the intrinsic limitations of the homodecoupling scheme but also from second-order effects and other properties of the spin systems. In such a case, one should not only compare signal amplitude of the desired peaks but also measure the level of the artifacts present in these spectra.

A simple visual inspection of Figure 5 shows that the experiment using the PSYCHE method for the F1 decoupling is more sensitive. But even after a careful adjustment of the contour level of the two spectra to compensate this difference gives no clear answer as to which experiment produces the least artifacts.

The SAN plot of Figure 6, on the other hand, provides a direct quantitative answer. The fact that the vertical distance between the two curves is smaller in the “artifact” region (center) than in the “signal” regions on the left shows that artifacts are relatively lower in the PSYCHE spectrum. The $S^+$ of the noise region is nearly equal ($\times 0.94$), but while the signal region increases by a factor...
2.5, the artifact regions also increase by a factor close to 1.5. The signal-to-artifact ratio therefore increases, but only by 67%. Reporting only the 250% increase in signal would be misleading with respect to the "true" quality of the experiment. Note that these results should not be used to draw any conclusion about the respective merits of these two homodecoupling methods, which will be the object of a separate study—this is just providing an example of possible application of the SAN plots. But the similarity in the performances of these two families of homodecoupling schemes was already pointed out for the DIAG experiment[16] and the applications of $^{13}$C decoupling to carbon enriched compounds.[17]

4.3 | Noise generated by reference deconvolution

This last example of SAN plot is used to compare different processing methods applied to a given experiment. This example should show how SAN plot could be used to automatize the optimization of parameters for processing techniques such as linear prediction or reference deconvolution, etc. In most cases, the default methods and parameters used in routine NMR are selected according to their robustness rather than for their true potential, which would need to be optimized in terms of sensitivity, resolution, and generation of artifacts.

This need of robustness is limiting the application of reference deconvolution from the routine processing of spectra even if it is quite powerful to correct lineshape, balance sensitivity and resolution, etc.[18] Usually, adjusting a lineshape deconvolution parameter toward an increased resolution initially improves the spectrum but suddenly produces spectra with dramatic levels of artifacts because of problems due to division by zero. In a recent development of deconvolution specifically tailored to the effect of field inhomogeneities in 2D correlation spectra, we could increase its robustness, but this only moved the problem of the breaking threshold a little bit further.[19] The choice of this example of application is justified by the potential of using SAN plot to facilitate the automatic selection of processing methods and parameters according to the quality of the spectra they produce.
Figure 7 shows the SAN plots of a F1-decoupled DIAG spectrum obtained after the application of lineshape deconvolution to correct the tilted lineshape. This elongation is due to the unavoidable small $B_0$ inhomogeneity observed in top-resolution spectra (see the left-most inset in Figure 7), that is, spectra recorded with long evolution times in $t_1$ and $t_2$. The respective SAN plots show the expected cost of correcting the lineshape of the peaks as a moderate increase in the level of noise and artifacts (compare the blue and purple lines in Figure 7).

Within the series, the best SNR of deconvoluted spectra was obtained using an intermediate level of broadening (SNR = 7'348 for 1.3 x 1.3 Hz Gaussian shape—not shown). The most extreme deconvolution provided extremely fine signals (0.5 x 0.5 Hz in F1 and F2, respectively, in the right most inset in Figure 7) at the cost of an increase in the level of the noise by almost one order of magnitude (compare the right parts of the blue and orange SAN plots). For a reliable peak picking, this spectrum is certainly not appropriate. But it may be quite suitable to determine or refine the values of coupling constants by best fit with simulated spectra. Producing such a “best for fitting” spectrum could be generated automatically by running a series of lineshape deconvolutions refining the target lineshape until the noise level increases by a maximal factor (say) 10 relative to the noise of a “neutral” spectrum. Such a controlled application of lineshape deconvolution combined with the ability of current computer to apply a full 2D deconvolution within a few seconds may contribute to increase the popularity of these deconvolution techniques.

4.4 | Implementation

The SAN plots presented in this article were obtained using the Matlab/Octave package of functions available on https://github.com/Gr-Jeannerat-unige/san-plot. A compiled version of the Matlab program running on MacOS and Windows systems allowing one to generate SAN plots for single or series of spectra without the need of licensed software will also be available. A very minimal code (c.a. 10 lines), which can be used as a starting point for programmers or to adapt the code to other programming languages, is given in the Supporting information.

5 | CONCLUSION

The SAN plots shall be quite useful to evaluate the SNR of spectra and facilitate the comparison of series a spectra both qualitatively and quantitatively. They are quickly and easily obtained and provide a reliable source of SNR and estimate of the level of artifacts. This makes the use of SAN plot compatible with a fully automatize analysis at diverse steps of the generation and use of NMR spectra. They can be used to verify the good “health” of a spectrometer by revealing stability problems, optimized the processing parameters of a spectrum, automatically define the contour level used to plot 2D spectra, and determine the best within a series of spectra when the choice of the optimal experiment or acquisition parameters cannot be determined prior to their acquisition.

ACKNOWLEDGEMENTS

We thank the State of Geneva, the Swiss National Science Foundation (grants 165725), and the Chemistry section for funding. We thank Marta Brucka for the homodecoupled spectra and other contributions to the manuscript and Kenji Caprice from the group of prof. Cougnon for the sample used to record the NOESY spectra.

ORCID

Kirill F. Sheberstov  
https://orcid.org/0000-0002-3520-6258

Eduard Sistaré Guardiola  
https://orcid.org/0000-0002-2631-0803

Marion Papier  
https://orcid.org/0000-0001-6033-2910

Damien Jeannerat  
https://orcid.org/0000-0001-7018-4288

REFERENCES

[1] D. I. Hoult, R. E. Richards, J. Magn. Reson. 1976, 24, 71. https://doi.org/10.1016/0022-2364(76)90233-X
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Sheberstov KF, Sistaré Guardiola E, Pupier M, Jeannerat D. SAN plot: A graphical representation of the signal, noise, and artifacts content of spectra. Magn Reson Chem. 2020;58:466–472. https://doi.org/10.1002/mrc.4882