Journal of Magnetic Resonance **137,** 161–176 (1999)
Article ID jmre.1998.1639, available online at http://www.idealibrary.com on IDEAL®

# A New Method for Spectral Decomposition Using a Bilinear Bayesian Approach

M. F. Ochs, R. S. Stoyanova, F. Arias-Mendoza, and T. R. Brown[1]

*NMR and Medical Spectroscopy, Fox Chase Cancer Center, Philadelphia, PA*

Received June 7, 1998; revised October 2, 1998

**A frequent problem in analysis is the need to find two matrices, closely related to the underlying measurement process, which when multiplied together reproduce the matrix of data points. Such problems arise throughout science, for example, in imaging where both the calibration of the sensor and the true scene may be unknown and in localized spectroscopy where multiple components may be present in varying amounts in any spectrum. Since both matrices are unknown, such a decomposition is a bilinear problem. We report here a solution to this problem for the case in which the decomposition results in matrices with elements drawn from positive additive distributions. We demonstrate the power of the methodology on chemical shift images (CSI). The new method, Bayesian spectral decomposition (BSD), reduces the CSI data to a small number of basis spectra together with their localized amplitudes. We apply this new algorithm to a $^{19}$F nonlocalized study of the catabolism of 5-fluorouracil in human liver, $^{31}$P CSI studies of a human head and calf muscle, and simulations which show its strengths and limitations. In all cases, the dataset, viewed as a matrix with rows containing the individual NMR spectra, results from the multiplication of a matrix of generally nonorthogonal basis spectra (the spectral matrix) by a matrix of the amplitudes of each basis spectrum in the the individual voxels (the amplitude matrix). The results show that BSD can simultaneously determine both the basis spectra and their distribution. In principle, BSD should solve this bilinear problem for any dataset which results from multiplication of matrices representing positive additive distributions if the data overdetermine the solutions.** © 1999 Academic Press

*Key Words:* **Bayesian methods; spectral analysis; bilinear forms; NMR spectroscopy; mixture analysis; positive additive distributions.**

## INTRODUCTION

A common need in the analysis of the large datasets found in chemical shift imaging (CSI) and many other fields is the reduction of the very large amount of information contained in the data to a manageable size. For example, in a CSI examination 512 spectra of 512 points are usually acquired. While many of these spectra contain nothing but noise, typically there are still hundreds of spectra to analyze. These spectra are rarely completely independent of one another but rather are a mixture of a handful of spectra coming from different tissue types making varying contributions to individual voxels. We are interested in how the CSI dataset can be decomposed into the spatial distributions of the spectra of the different tissue types. Since we know neither the spectra nor their spatial distributions, we must solve a bilinear problem in order to determine them simultaneously. Most traditional methods of data analysis (e.g., Fourier transformation, least squares fitting) cannot decompose the data in this way, but simply estimate the individual spectra (or their properties) in each voxel with no attempt to determine their interrelationship.

In a general bilinear problem, the data matrix, $D$, can be considered as a series of $M$ vectors taken from $R^N$, yielding an $M \times N$ matrix. The problem is to obtain both the matrix of $K(K \ll M, N)$ often nonorthogonal, basis vectors, $F(K \times N)$ (here the spectral shapes), and a mixing matrix, $A(M \times K)$, which gives the amount of each basis vector in the actual data. The data is then related to the model through a matrix multiplication,

$$D = AF. \qquad [1]$$

This is similar to a standard "inverse" problem except that in the "inverse" case one of the matrices is known, and thus least squares methods can be used to find the matrix which minimizes the residuals between the reconstruction and the data. With neither $A$ nor $F$ known (even if $K$ is only 2 or 3), the problem is much more difficult. Since the number of possible solutions is very large and there is no analytical method to identify them, we use the Markov chain Monte Carlo procedure (MCMC) to sample the space of possible solutions to determine its properties. MCMC is a technique derived from statistical mechanics, where it has been used for more than 50 years to explore the solution spaces associated with distributions of interacting molecules or spins. Since MCMC algorithms directly sample the solution space, uncertainty estimates are determined simultaneously with a "best" solution. Further, if the data support them, multiple solutions are possible. Their application to stochastic image processes was initially demonstrated by Geman and Geman (*1*), leading to exploration of a wide variety of sampling procedures (*2–4*) for solution of imaging problems, reviewed by Besag *et al.* (*5*).

[1] To whom correspondence should be addressed.



1090-7807/99 $30.00

MCMC techniques require relative probability measurements at each sampled point in the solution space, which we provide through a Bayesian approach. In the past decade Bayesian methods using MCMC techniques have been used in a wide variety of problems in data analysis, e.g., medical imaging, agricultural field studies, population studies, and economic forecasting (*6–10*). Bayesian statistical analysis starts with the apparently trivial statement,

$$p(M, D) = p(M|D)\,p(D) = p(D|M)\,p(M), \qquad [2]$$

where $p(M, D)$ is the probability of both the model and the data (the *joint probability distribution*), $p(M|D)$ is the conditional probability of the model given the data (the *posterior*), $p(D)$ is the probability of the data (the *evidence*), $p(D|M)$ is the conditional probability of the data given the model (the *likelihood*), and $p(M)$ is the probability of the model (the *prior*). The posterior distribution is the solution space for our problem, since it measures the probability of the present model (sample) in light of the data. Rearrangement of Eq. [2] yields the posterior,

$$p(M|D) = \frac{p(D|M)\,p(M)}{p(D)}, \qquad [3]$$

which provides the MCMC algorithm with the needed probabilities in the solution space for the problem. Since the evidence, $p(D)$, usually acts as a scaling parameter, it can be ignored in this case since MCMC only needs relative probabilities. This means that the relative probability at a point in the solution space is determined completely by the likelihood, which is easily determined by comparing the model to the data, and the prior, which is the probability of the model independent of the data. The prior encodes any knowledge of the solution independent of the data. For example, a prior for a system reconstructing spectra might give higher probability to a narrow spike than to a flat offset.

Putting in the matrices $A$ and $F$ for the model leads to the specific form of Bayes' equation (Eq. [3]) for the bilinear problem,

$$p(A, F|D) \propto p(D|A, F)\,p(A, F). \qquad [4]$$

The sampling from the posterior distribution and the encoding of the prior are done using a heavily modified version of the Massive Inference Gibbs sampler from MaxEnt Solutions Ltd., Cambridge, England, which also enforces positivity on the solutions. The primary modifications revolve around how the likelihood changes as the MCMC samples the solution space. The original Massive Inference system handled systems where $A$ in Eq. [1] is a known constant matrix, which makes the change in the likelihood dependent only on a change in $F$, $\delta F$. When $A$ is treated as a variable matrix on the same footing as $F$, the calculations of the change in the likelihood with a change in flux in either $A$ or $F$ requires that the other matrix be constantly updated, which is discussed in detail below.

Some discussion about the choice of prior is necessary, since this is the area in which most of the controversies about Bayesian techniques arise. The objective is to use the most general possible arguments to determine a form of the prior that is appropriate to the general character of the problem under consideration with only a few adjustable parameters. Since BSD is to reconstruct spectral shapes (known to contain fairly sharp lines) and spatial distributions (essentially images), using the atomic prior from Massive Inference is appropriate. An atomic prior represents the model as a few point fluxes (atoms) with the highest probability assigned to the distribution with the fewest atoms. It contains only two adjustable parameters, the average strength (flux) of the atoms and the probability of finding an atom. Both are adjusted by the program to match the data. This prior follows naturally from general divisibility arguments (*11*), and thus is widely applicable. For example, it should also be effective in describing systems where the signals arise from discrete objects (e.g., photons striking a photographic plate, nuclei undergoing spin flips).

Once the prior is chosen, the remainder of the problem is straightforward, although a number of features have been added to BSD to improve efficiency. BSD starts the Markov chain at a point in the posterior distribution representing a completely flat model containing a reconstructed flux equal to the flux in the data. In this way the sampler starts nearer the region of high probability while avoiding any initial bias on expected spectral shapes or distributions. The likelihood is calculated using the sum of the squares of the residuals normalized by the standard deviation, $\sigma$, of the noise in the spectral data, i.e., a normalized $\chi^2$ distribution. Rather than perform a full likelihood calculation for each movement of the Markov chain, the change in the likelihood is calculated for the specific change in the model, so that the likelihood can be updated incrementally. The likelihood, $L$, can be written in matrix notation as

$$L = \frac{1}{2\sigma^2}\,\mathrm{Tr}[(AF - D)^{\mathrm{T}}(AF - D)]. \qquad [5]$$

where $A^{\mathrm{T}}$ represents the transpose of $A$ and Tr indicates the trace of the quantity in the brackets. Then the behavior of the change in the likelihood, $\Delta L$, can be derived by looking at the effect of adding a small amount of flux, $\delta F$, to the model. By inserting $F + \delta F$ for $F$ in Eq. [5] and subtracting Eq. [5] from the result, the change in likelihood for a change in $F$ is

$$\Delta L(\delta F) = \frac{1}{2\sigma^2} \times \mathrm{Tr}\left[\begin{array}{l}(A\delta F)^{\mathrm{T}}(AF - D) \\ \quad + (AF - D)^{\mathrm{T}}A\delta F + (A\delta F)^{\mathrm{T}}(A\delta F)\end{array}\right], \qquad [6]$$

where it is assumed that only changes to $F$ are made. The

coding is made more efficient by maintaining a mismatch vector which measures the misfit between the data and the reconstruction from the model, i.e.,

$$M = D - AF. \quad [7]$$

A great increase in calculational efficiency is gained by updating the mismatch vector incrementally after each Markov step just as the likelihood is incremented. For added flux $\delta F$, $M$ changes by

$$\Delta M = D - A\delta F, \quad [8]$$

where only the affected components of $M$ must be updated. Equations [6] and [8] have similar forms for changes in the model for $A$. In order to simplify the calculations, we do not allow simultaneous changes in $A$ and $F$, since allowing such changes would require evaluation of terms involving $\delta A \delta F$. Note that barring such changes does not prevent the system from reaching any state and should have no effect on the final result, since the sampler can move $\delta F$ followed by $\delta A$ and reach the same point. As long as detailed balance is maintained, the sampler still samples the space correctly. At each step of the Markov chain, the program calculates the change in the likelihood using Eq. [6] and determines whether to move by comparing this with a randomly generated value. If the step is taken, the likelihood and the mismatch vector are updated. MCMC samplers require a "burn-in" time to reach an area of high probability which is suitable for sampling. The sampler runs for an operator-specified time without recording samples and then continues while recording for a further number of steps specified by the operator.

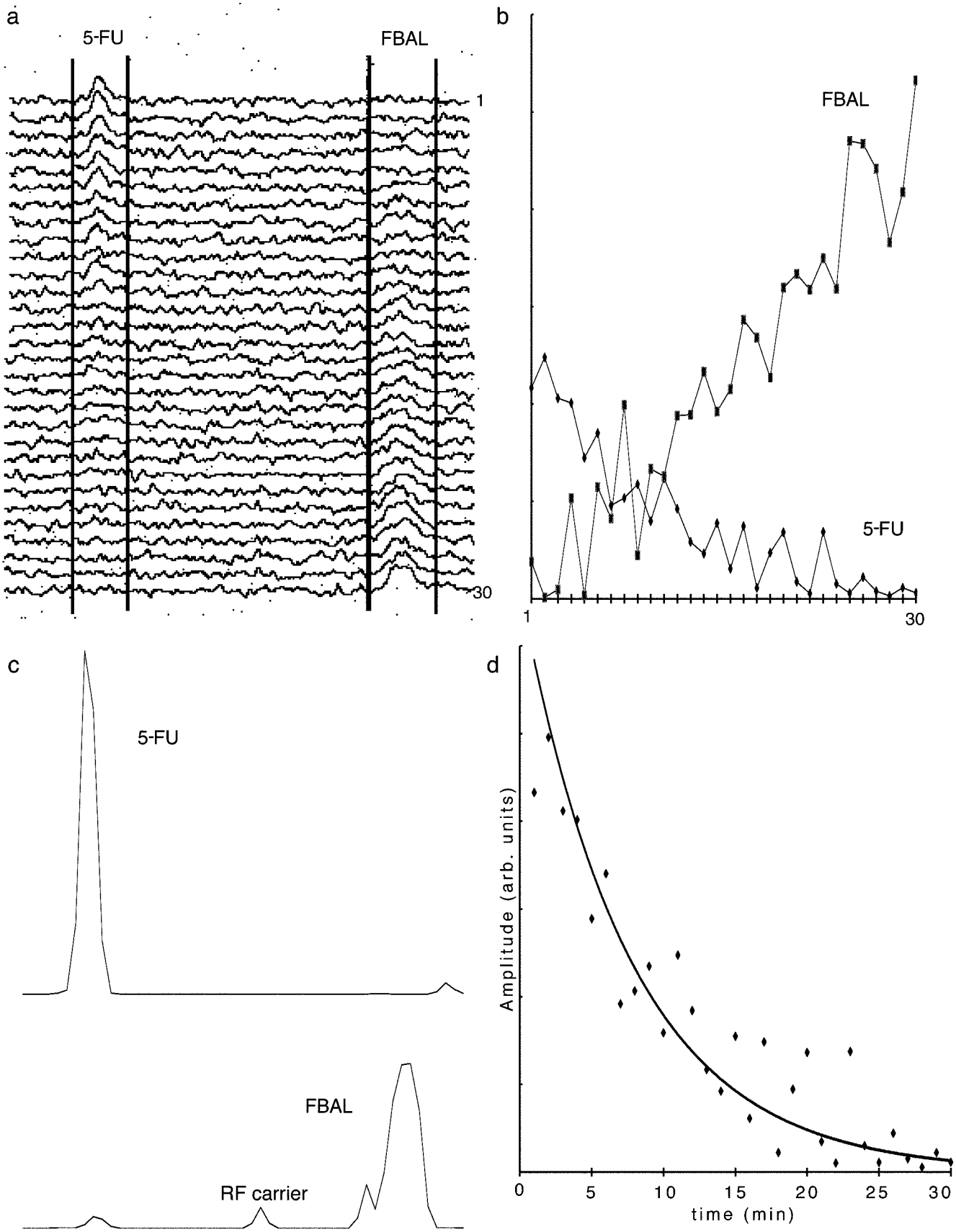
A final modification was made to BSD in order to more fully represent the physical world in the models. Atoms in $F$ are given a Gaussian lineshape with a width defined by the operator, which is generally the natural width of the problem, usually directly measurable from the narrowest line in the spectrum. For the mixing matrix, $A$, *a priori* knowledge of the absence of material is sometimes available, so the operator also has the option of specifying a certain number of zeros in one solution component in the $A$ matrix. For strongly overlapping spectra, especially when a single line is dominate in one of the underlying spectra, as in the CSI study of the human head presented below, it greatly improves efficiency to add such *a priori* knowledge of the distribution of signals.

The operation of BSD on CSI and multispectral datasets is straightforward. First, principal component analysis (PCA) is used to correct the data for instrumentally induced frequency and phase shifts as described previously (*12, 13*). PCA is then applied to the corrected data to determine the number of independent spectral shapes, $K$ in Eq. [1], needed in the model to reconstruct the data. Generally it is obvious from the PCA results how many independent shapes are present in the data. However, if there is any uncertainty, BSD can be run with several different $K$ values. The data, the number of shapes, the standard deviation of the noise, and the linewidth are fed into BSD together with the number of iterations desired. These are the only inputs that BSD requires to operate. During sampling BSD is free to exchange flux between the $A$ and $F$ domains, so the individual samples are scaled prior to averaging. BSD is generally run using several different Markov chains in order to verify the results, as MCMC techniques have no established convergence criteria. Since BSD samples the solution space, the output includes not only a mean solution but also uncertainty estimates at each spectral point as well as at each amplitude in the mixing matrix. If there are multiple possible solutions, BSD will find these as well. The power of the BSD algorithm is demonstrated on a series of increasingly complex datasets below.

## RESULTS

A straightforward example of the operation of BSD is presented in Fig. 1 with a study of the catabolism of 5-fluorouracil (5-FU) to $\alpha$-fluoro-$\beta$-alanine (FBAL) in human liver during chemotherapeutic treatment (*14*). PCA was used to remove small frequency offsets in the individual spectra (Fig. 1a) and to determine that two orthogonal components adequately described the data. The BSD algorithm was told to search for two spectral shapes. These shapes and their amplitudes are shown in Figs. 1b and 1c. Repeating the analysis with four different seeds and thus four different Markov chains generated identical results (not shown). Note that the fluctuations in amplitude in Fig. 1b are not due to the MCMC procedure but reflect the actual variations in the data in Fig. 1a. The time constant of the exponential fit shown in Fig. 1d is $7.61^{+1.90}_{-1.27}$ minutes (95% confidence levels) for the decline of 5-FU, in agreement with previously published results obtained using PCA (*14*). In this case BSD took less than 7 min on an Apple PowerBook 3400/200 running the application compiled using Metrowerks Codewarrior C and sampling 10,000 points from the posterior distribution. Increasing the sampling to 20,000 points did not change the result, demonstrating that the sampling had achieved equilibrium. Note that while the previous analysis by Li *et al.* required *ad hoc* transformation of the PCA components to reconstruct the 5-FU and FBAL spectra, these spectral shapes were determined automatically by BSD. The reconstructed spectral shapes clearly show the power of the atomic prior, which encourages noise in the spectra to be reduced to the baseline, while maintaining features which are slightly above the noise. The small peak on the shoulder of FBAL in Fig. 1c can be seen in the data in Fig. 1a; however, a dataset with better SNR would be required to confirm its presence.

A more complex decomposition problem is shown in Fig. 2. This is a dataset comprising 256 $^{1}$H decoupled $^{31}$P spectra of typical peak signal-to-noise ratio (SNR) of approximately 6. These spectra were selected by choosing axial slices with signal from 512 spectra ($8 \times 8 \times 8$ voxels) obtained by spatial and time FFT of CSI data acquired from a volunteer's head as


a
5-FU
FBAL
1
30
b
FBAL
5-FU
1
30
c
5-FU
FBAL
RF carrier
d
Amplitude (arb. units)
0
5
10
15
20
25
30
time (min)

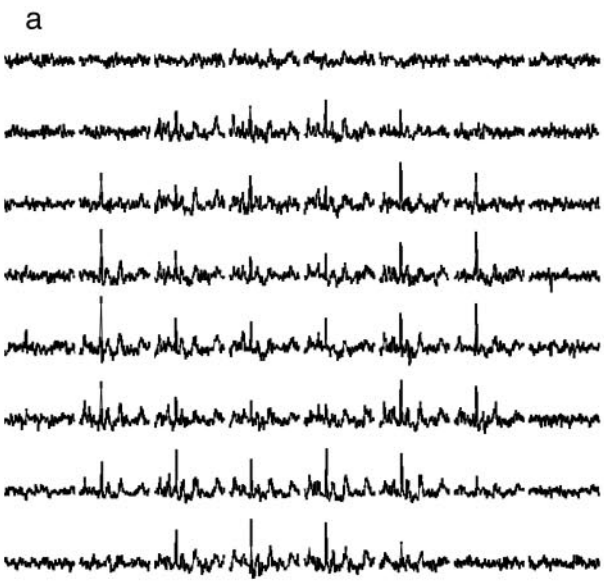
described elsewhere (*15*). The low SNR of the spectra (typical for such studies), 64 of which are shown in Fig. 2a together with the corresponding proton image in Fig. 2b, make it virtually impossible to study individual peaks. PCA was again used to align the spectra on the PCr peak. The PCA analysis determined that two components adequately described the dataset so BSD was run looking for two spectral shapes and their distributions. Figure 2c shows the resulting reconstructed amplitude distributions on the same scale for comparison, while Fig. 2d shows the underlying spectral shapes, which were reconstructed using Gaussian lineshapes with widths of 5.7 points. The reconstructed spectral shapes are clearly identifiable as characteristic of muscle tissue and of brain tissue. The brain spectrum shows large phosphodiester and phosphomonoester (PME) peaks, the expected broad $\beta$ATP resonance arising from exchange between free and ATP-bound magnesium, and the typical $\beta$ATP frequency shift indicating a lower free $Mg^{2+}$ concentration than in muscle (*16*). The amplitudes show the muscle localized on the edge of the skull and at the occipital lobes as expected, while the brain is internal to the muscle signal. Since the reconstructed spectra result from fitting the model to 256 data spectra, there is a dramatic improvement in the SNR over the unprocessed data.

This case demonstrates some of the complexity of this procedure since the solution in Fig. 2 was only one of the possible solutions found by the BSD procedure. This solution resulted when 12 zeros were set in the amplitude of one spectral shape deep inside the head, which had the effect of forcing that region to be represented by only the "brain" spectral shape. In addition to this solution, BSD found solutions with a "brain" spectrum with either half or almost no PCr when run with no forced zeros. The fit to the data was preserved by adding a fraction (typically 10%) of the "muscle" spectral shape into the brain region (see Fig. 3 for an extreme example). In fact, Fig. 4 shows plots of the data, reconstructions from the models, and residuals for both cases. There is no perceivable difference in the residuals, indicating that there is no support for one solution over the other in the data itself. Since BSD samples the solution space directly, it finds such mathematically possible solutions, which can be helpful when the physical situation is not as well determined as here. This second, nonphysical solution could be excluded *a posteriori* by noting that the brain does not contain muscle tissue or *a priori* by forcing a solution to have zero amplitude deep in the brain. The *a priori* approach is computationally more efficient, since it does not require many different Markov chains to obtain physically significant results. Both BSD analyses on the 256 spectra of 369 points took roughly 9 h on a 600 MHz Digital alpha workstation with sampling of 50,000 points from the posterior distribution following 24,000 iterations to allow equilibration.

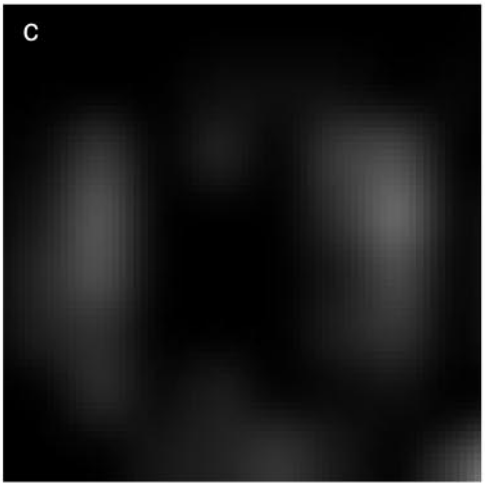
In order to explore the meaning of these multiple solutions more fully, we generated a dataset composed of 100 data spectra of 300 points each with strongly overlapping peaks. Each spectrum in the data was a mixture of three basis spectra, which were modeled on typical muscle spectra containing small pH differences and small *J* coupling and ATP shift differences. The basis spectra together with their distributions are shown in Figs. 5a and 5b, respectively. The individual basis spectra contain 10 spectral lines, each with a Gaussian lineshape of width 2.2 points. Random Gaussian noise was then added to each data spectrum at different levels and BSD was told to search for three basis spectra using a number of different Markov chains.

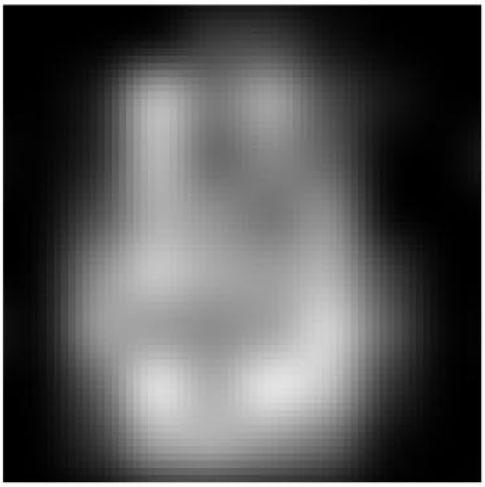
The picture which emerges from these simulations is one where BSD reliably finds the expected solution in cases where the SNR is high, but as the noise level increases it finds this solution only part of the time. In Fig. 6, sample spectra of the data for each noise level are shown. The differences between the simulated basis spectral shapes are primarily in the Pi and ATP peaks. The maximum SNRs used in the simulations for these peaks in the data are 8, 6, 4, and 2 for ATP and 16, 12, 8, and 4 for Pi. Figure 7 shows the two solution types found in the case of the highest SNR. As can be seen, they are almost identical. The spectral shapes shown in Fig. 7a show some minor crosstalk between the basis spectra in the ATP regions of the second and third spectra leading to small peaks around the expected larger peaks. The uncertainties calculated by BSD for these peaks are roughly half their peak amplitude, indicating that they are not well supported by the data (typical peak uncertainties claimed by BSD in these spectra are at the 15% level while they are at the 5% level in the Fig. 7b solution). Both solutions (Figs. 7a and 7b) show the correct larger peaks compared to the true basis spectra with the correct relationships between Pi and ATP shifts and *J* couplings. As the noise level is increased, BSD begins to find other possible solutions. At the second highest noise level, 20% of the time (2 out of 10 Markov chains), BSD returns a solution (Fig. 8) which strongly mixes the three basis spectral shapes to form a solution which has fewer atoms (thus a higher prior probability) while having a higher $\chi^2$ (thus a lower likelihood). The reconstructed model's fit to the data is poorer as measured by a root mean square residual misfit in the amplitudes, which is more than twice the size of the correct case. However, as Fig. 9 shows, the residuals of the reconstruction compared to the data appear equally uniform. As the noise increases this solution is seen more

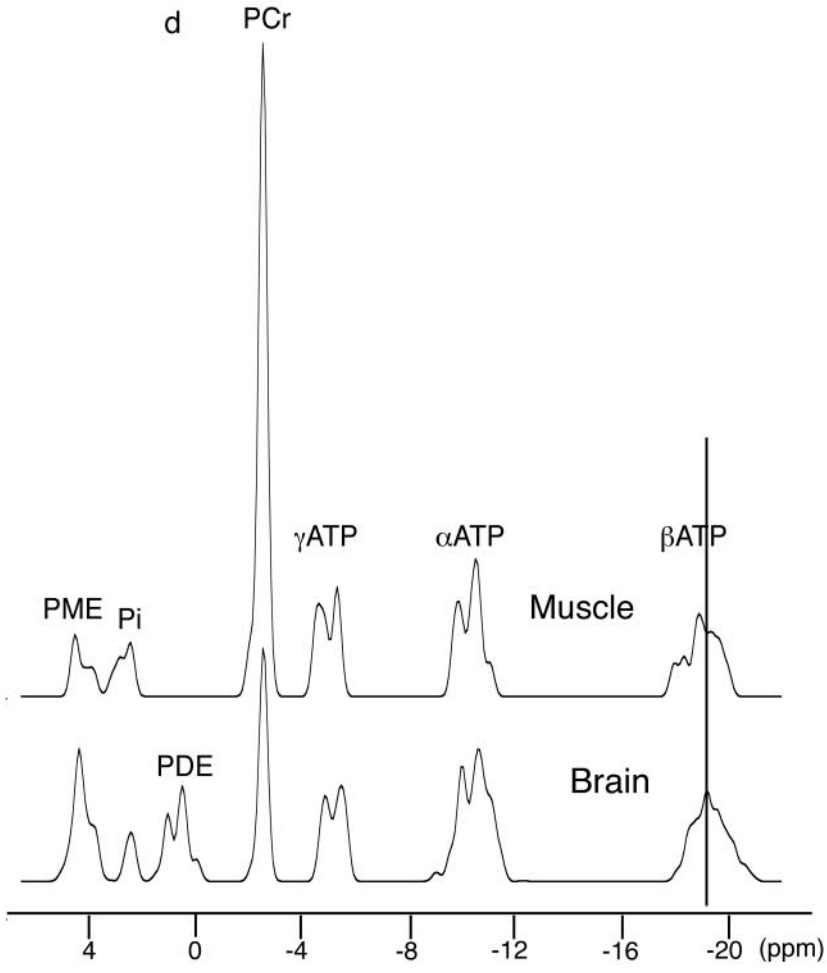
**FIG. 1.** Time series data for the catabolism of 5-FU to FBAL: (a) The data contain 51 points extracted from 30 1-min, nonlocalized $^{19}F$ NMR spectra with peak SNR $\sim$ 7 acquired over 30 min following rapid (1-min) bolus injection of 5-FU. The time series runs from top to bottom and the locations of the 5-FU and FBAL peaks are shown. (b) The amplitudes of the two underlying spectral shapes within the data determined by BSD. The time axis runs from 1 to 30 min. (c) The two underlying spectral shapes determined by BSD. At top is the 5-FU spectrum; at bottom is the FBAL spectrum with the RF carrier showing in the middle. (d) An exponential fit by regression analysis to the decay of the amplitude of the 5-FU signal. A time constant of 7.61 min ($+1.90/-1.27$ min at 95 confidence) was determined.


a


b


c




d
PCr
γATP
αATP
βATP
PME
Pi
Muscle
PDE
Brain
4
0
-4
-8
-12
-16
-20
(ppm)

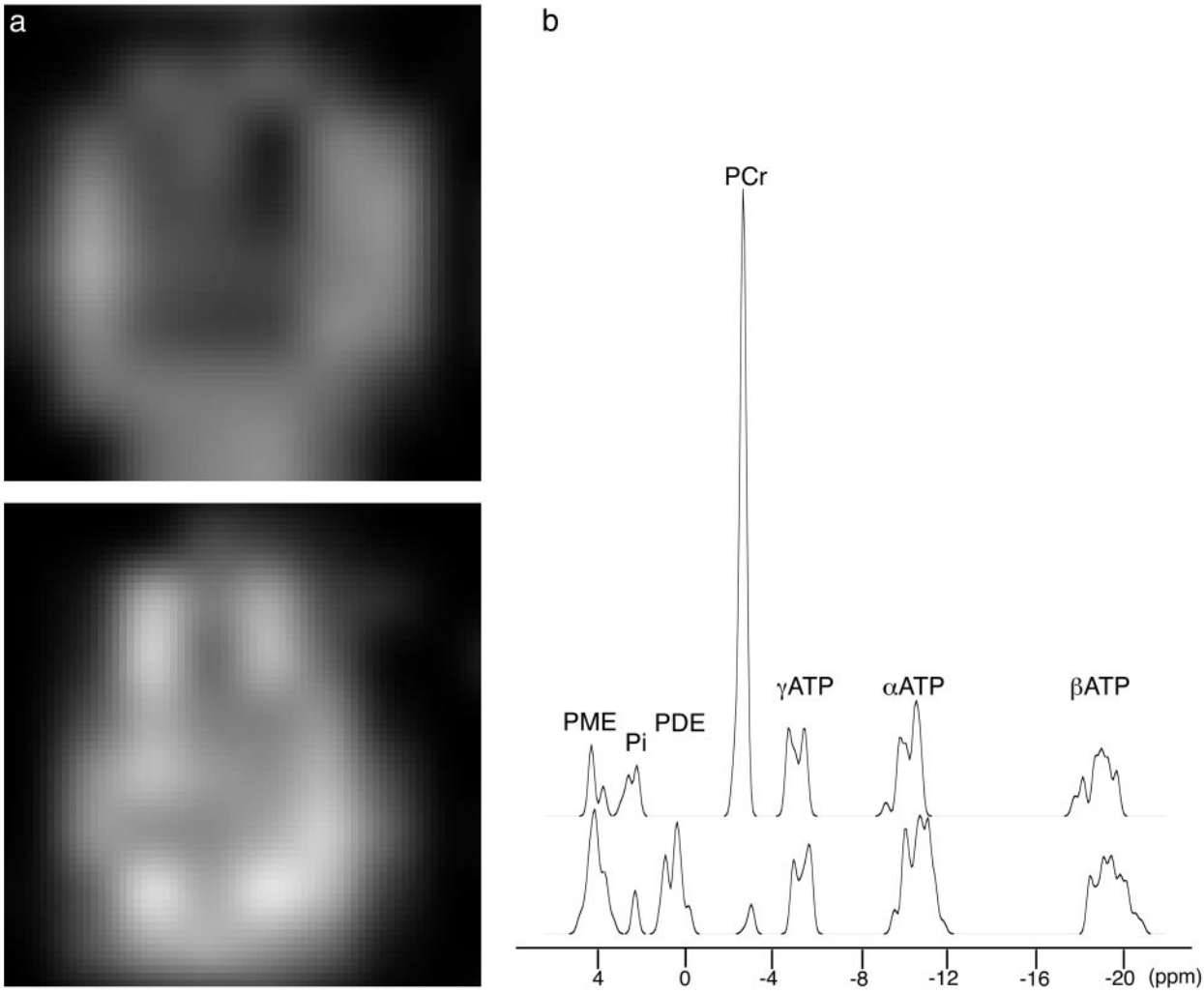



**FIG. 3.** Second solution for head data with no zeros set in the brain tissue. (a) Amplitude distributions for two solutions, again on the same scale. Top solution is "muscle-like"; however, it now encroaches into the brain tissue region. (b) Spectra corresponding to the distributions with linewidths of 5.7 points as in Fig. 2. The spectra are similar except that the "brain-like" spectrum has lost its PCr, while the $\beta$ATP and PME regions are mixed between the two solutions.

often, so that at the third level of SNR (ATP maximum SNR of 4), only one-third (7 out of 20 Markov chains) of solutions are the correct solution. Finally, in the analysis of the dataset with the lowest SNR, the prior probability dominates the solution space, and BSD prefers to fit the data with two basis spectral shapes to reduce the number of atoms. These results reflect the general pattern seen in this type of Bayesian analysis, in which the prior becomes more and more dominant as the information content in the data diminishes.

These datasets also allowed a test of the effect of increased equilibration and sampling times on the residuals. For the highest SNR case, we took the two solution types and ran the data through BSD for a variety of sampling periods. The data are summarized in Table 1, which shows that the sampler settles down at different rates in different situations. The model with the smaller misfit essentially is close to its best possible fit within only 20,000 samples (after 11,000 equilibration steps), while the second model requires roughly 50,000 samples (after

**FIG. 2.** CSI dataset ($8 \times 8 \times 4$ voxels of 22 $cm^3$ each) from the human head of a normal volunteer: (a) $^{31}P$ spectra from a single axial slice; 64 of 256 total spectra shown. (b) The corresponding $^{1}H$ image centered axially on the region of the voxels. (c) The $8 \times 8$ amplitude distributions are shown with slight Gaussian blurs applied to make the distributions easier to see. The intensity scale is the same in both distributions to aid in comparison. At top is the distribution for the spectral shape characteristic of muscle tissue, while at the bottom is the distribution for the spectral shape characteristic of brain tissue. (d) At top is the reconstructed spectral shape associated with muscle which shows sharp $\beta$ATP lines centered at $-18.62$ ppm with PCr set at $-2.52$ ppm. At bottom is the reconstructed spectral shape associated with brain which shows $\beta$ATP centered at $-18.92$ ppm with PCr set at $-2.52$ ppm. The lineshape used by BSD was Gaussian with width 5.7 points.

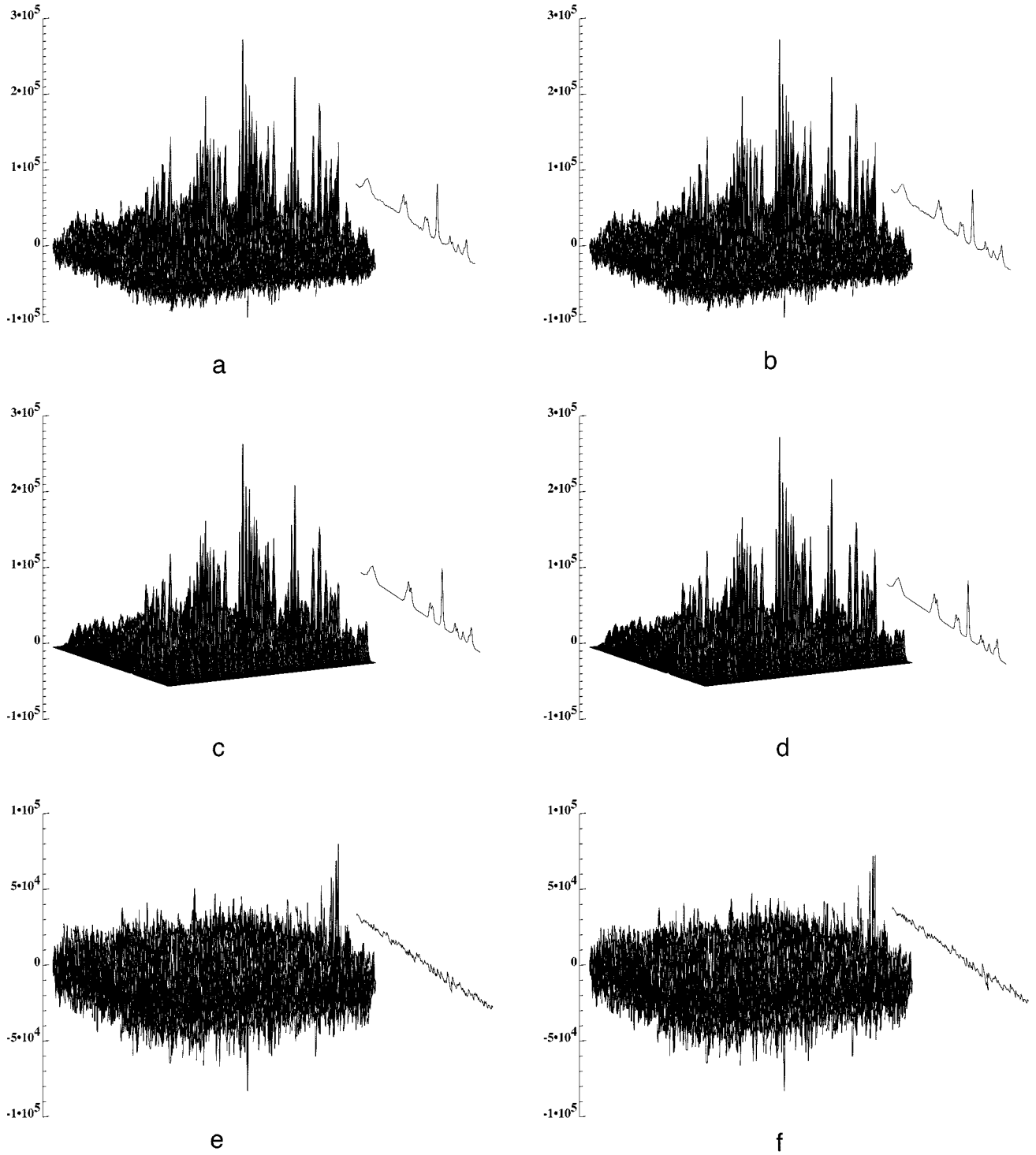



**FIG. 4.** Plots of full datasets, reconstructions, and residuals for head data: Each plot contains 256 spectra of 369 points each shown from left to right together with an average across all the spectra at the far right. (a) and (b) The input datasets (identical to each other) with the low SNR and large variations apparent. (c) Reconstruction from the solution shown in Fig. 2. (d) Reconstruction from the solution shown in Fig. 3. (e) Residuals, (a) − (b), for the reconstruction in (c). (f) Residuals, (b) − (d), for reconstruction in (d).

24,000 equilibration steps) to reach its best fit. The use of multiple Markov chains also aids in determining if adequate equilibration and sampling time was allowed, since if the solutions are not equilibrated, they will show different results for different Markov chains.

We also wished to look at failure modes in BSD. We

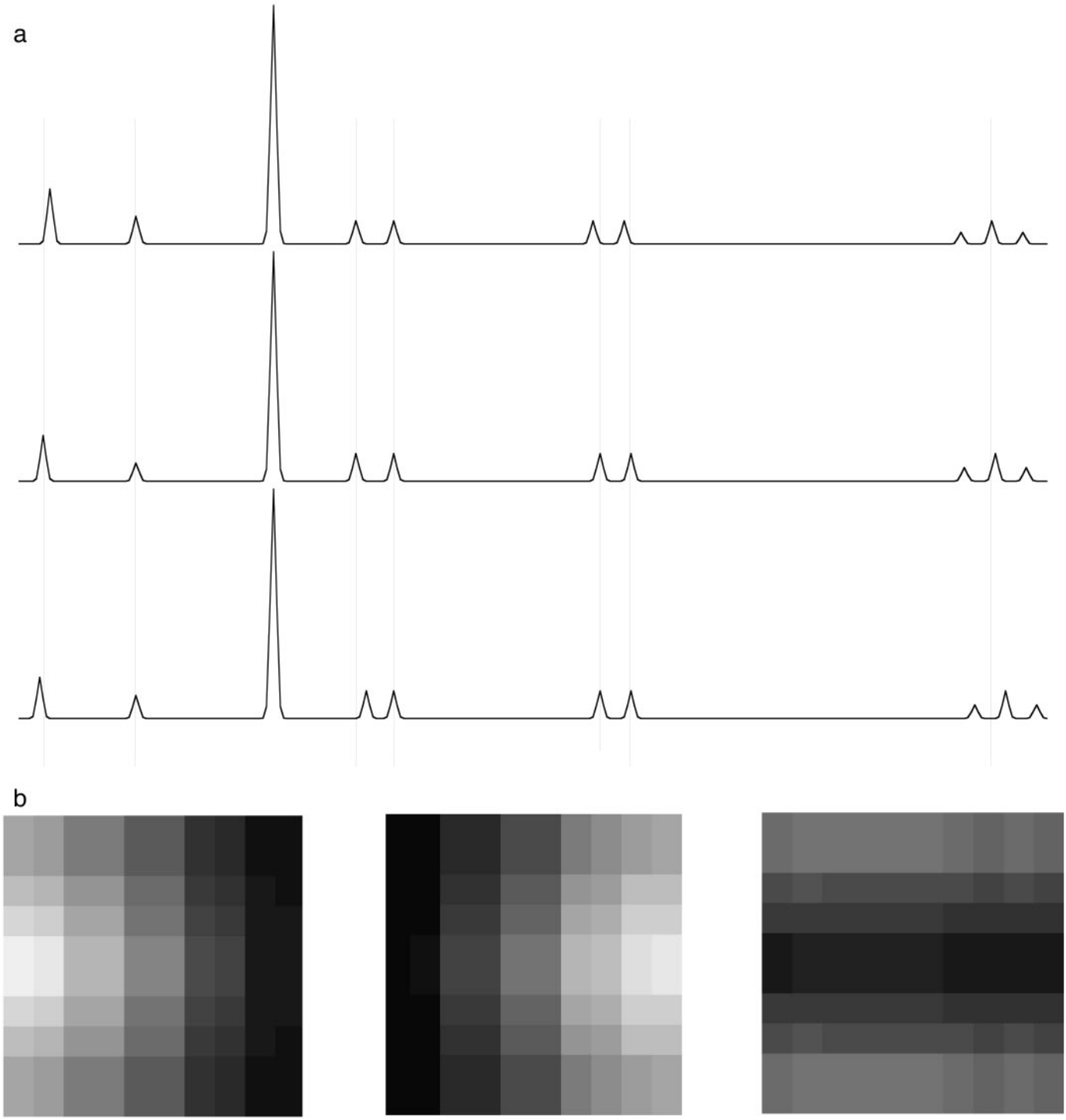



**FIG. 5.** Model spectra and distributions: (a) Three model spectra showing shifts of peaks between them. (b) Amplitude distributions for spectra, top spectrum goes with left distribution, middle spectrum with middle distribution, bottom spectrum with right distribution.

discovered that if we put in incorrect linewidths, BSD still finds the correct solution as often as with the correct linewidth, as long as the real linewidth is larger than the linewidth used by BSD; however, the normalized $\chi^2$ value is higher than expected. If the linewidth is too wide, however, BSD returns spectra which are not realistic (incorrect frequency relationships) so that BSD indicates to the user that there is a problem in the model. If BSD is told to seek more than the correct number of solutions, the case is more complicated. Many times we have seen BSD find one solution of the set with either a zero solution (flat spectral shape with no amplitude) or a nonphysical solution with a few random spikes in the spectral

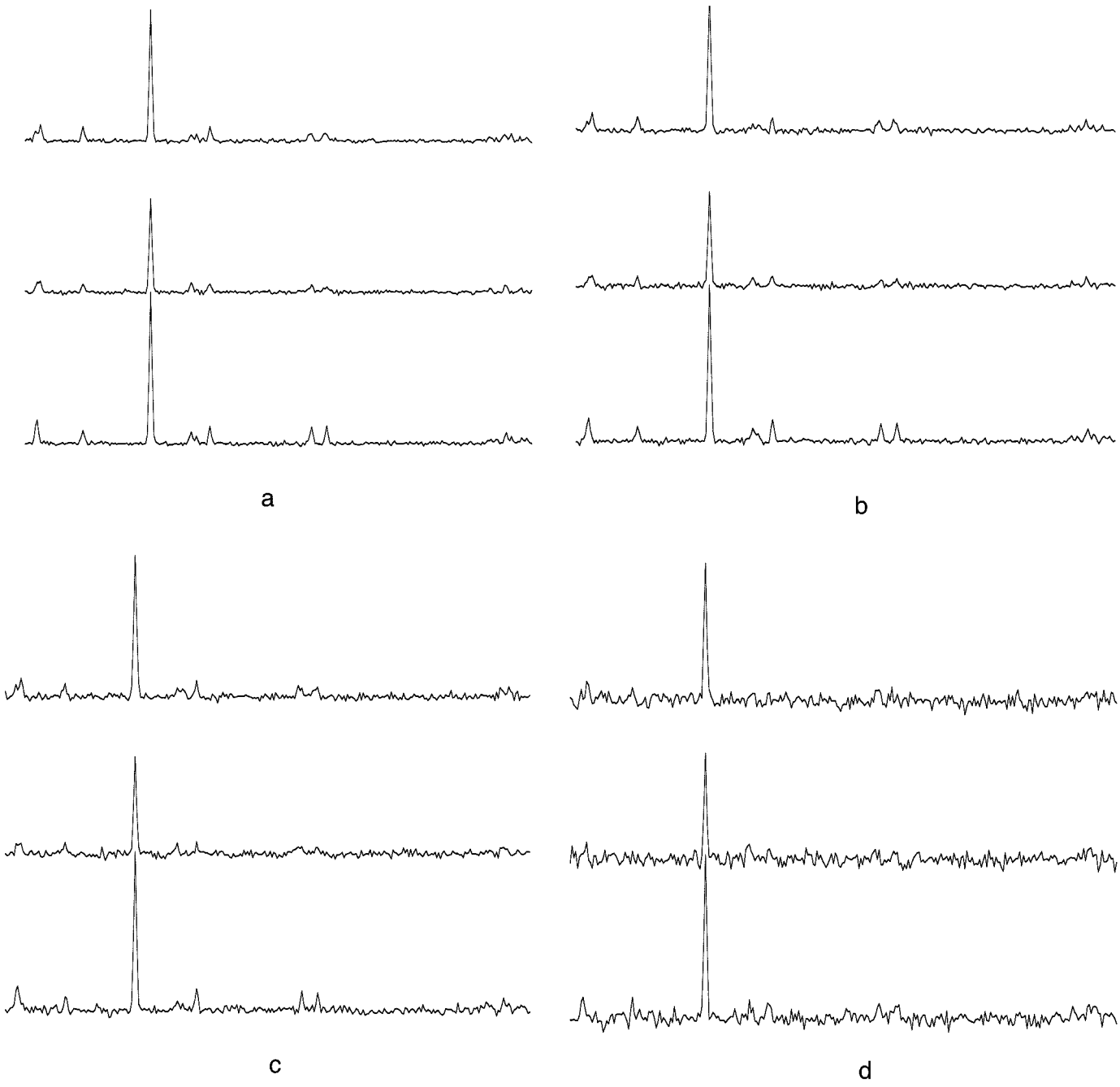



**FIG. 6.** Sample data spectra from model in Fig. 5 together with Gaussian noise: (a) Peak SNR of 8 for ATP peaks. (b) Peak SNR of 6 for ATP peaks. (c) Peak SNR of 4 for ATP peaks. (d) Peak SNR of 2 for ATP peaks.

shape and very low amplitude. However, in cases of high SNR and strongly overlapping spectral shapes, it can create an additional solution. This shows the importance of having an independent procedure (PCA) which can be used in cases where there is extreme overlap in the basis vectors. If too few solutions are sought, BSD is unable to return a reasonable normalized $\chi^2$ value.

During sampling, BSD also gathers statistical data on the distribution of the possible models, which allows it to give both the mean model and the standard deviations of the points in the model. In the bilinear case, these uncertainties are more complex than for a Markov chain in a linear system. In a bilinear system there is the possibility of correlated uncertainties between the two domains, $A$ and $F$. In our specific instance this is compounded by the treatment of an atom in $F$ as a spectral line, which effectively means an atom in $F$ is distributed over many points while an atom in $A$ is not. In order to test the uncertainties, we ran the high SNR dataset first with the correct

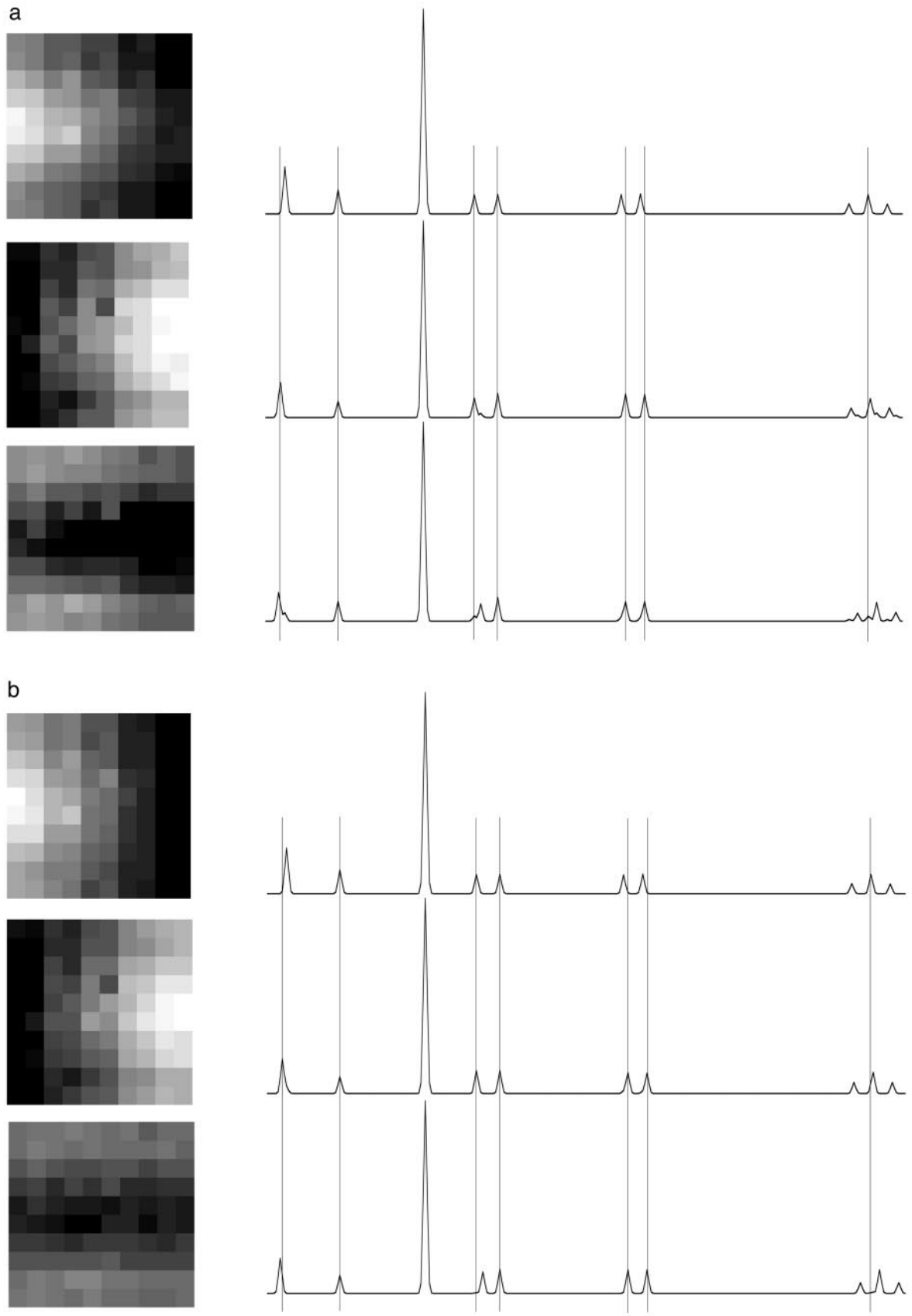



**FIG. 7.** Two solutions found by BSD for highest SNR case: (a) First solution with higher root mean square misfit to the known distributions. (b) Second solution with lower root mean square misfit to the known distributions.

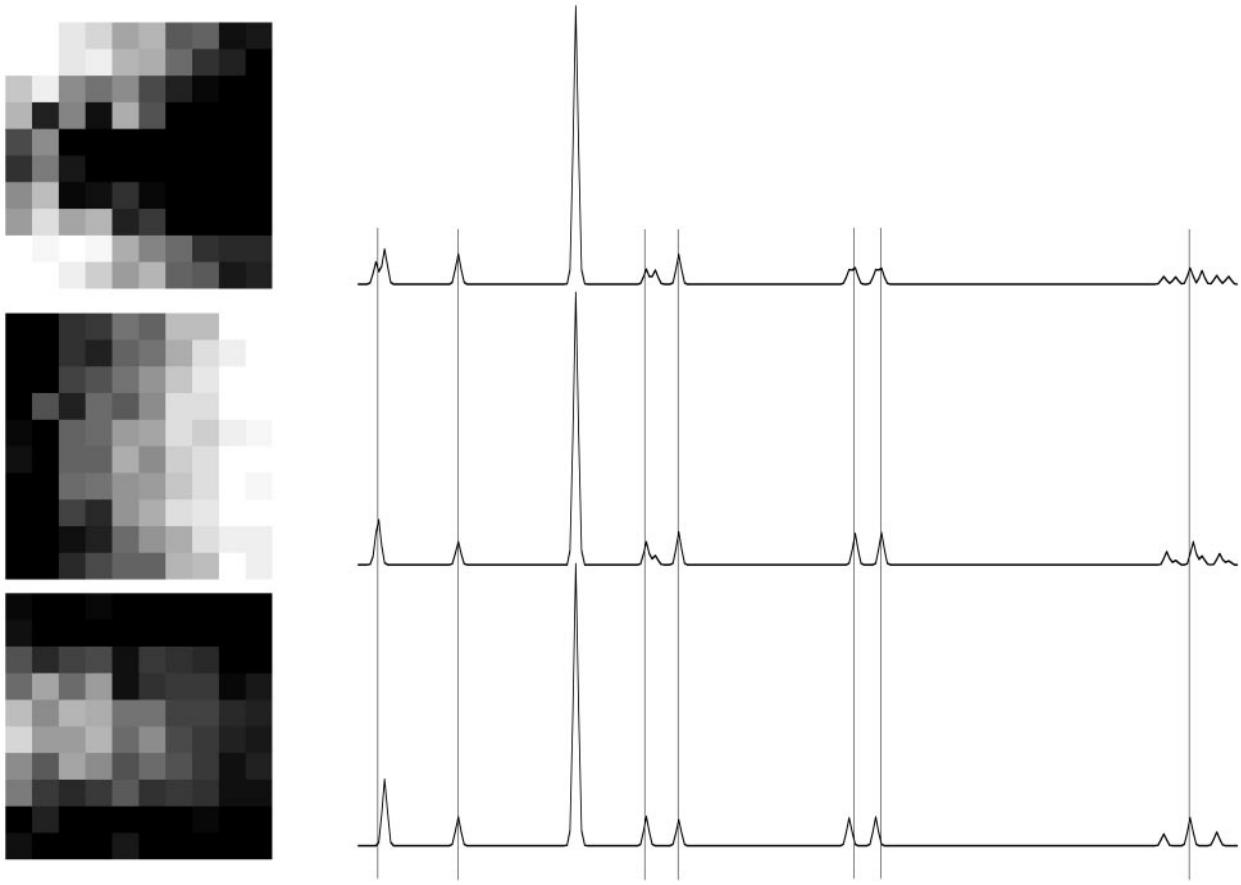


**FIG. 8.** "Bad" solution found in second highest SNR case: The root mean square misfit to the known distribution is roughly twice that for the "good" solution.

linewidth and then with no linewidth (effectively treating each point in the spectra independently). The uncertainties summed over all points in $A$ and $F$ are summarized in Table 2. Here there is a clear better overall fit to the spectra when atoms in $F$ are given a lineshape, but this results in slightly poorer fit in $A$. Also, the calculated standard deviations show that the sampler is more tightly locked into the spectral shapes when an atom is converted to a linewidth than to a single point (standard deviation of $8.5 \times 10^{-5}$ vs $1.4 \times 10^{-4}$). This leads to the sampler possibly underestimating uncertainties for the peak heights in the spectral shapes and overestimating them for the amplitudes in the mixing matrix, which indicates that running multiple Markov chains may be a better way to estimate uncertainties in the bilinear case.

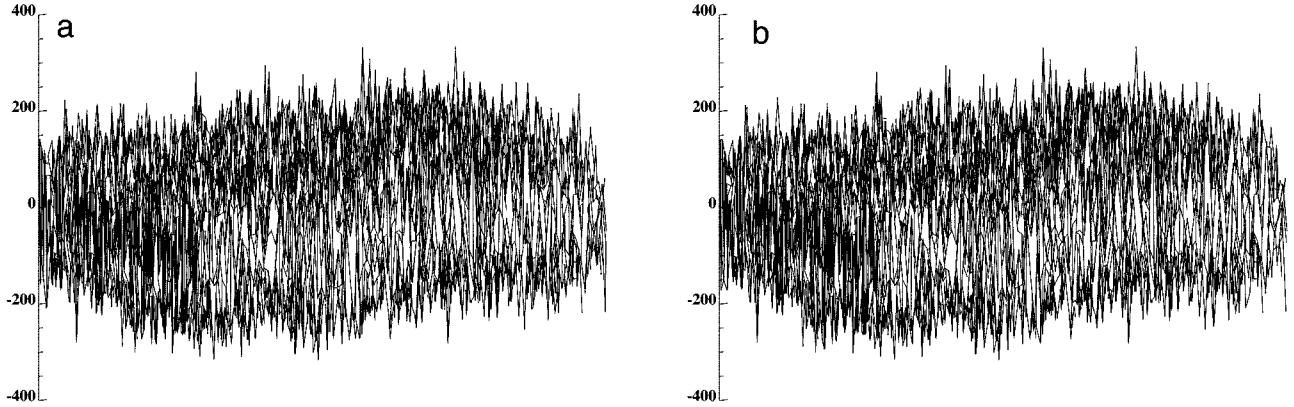



**FIG. 9.** Residual plots between the data and the reconstruction for solutions at second highest SNR ratio: (a) Plot for "good" solution. (b) Plot for "bad" solution.

**TABLE 1**
**Root Mean Square Deviations from the Known Amplitudes for the Solutions Shown in Fig. 7**

| Number of samples | First solution amplitude misfit (Fig. 7a) | Second solution amplitude misfit (Fig. 7b) |
|---|---|---|
| 5,000 | | 79 |
| 10,000 | 187 | 77 |
| 20,000 | 139 | 71 |
| 30,000 | 133 | 67 |
| 50,000 | 114 | 67 |
| 75,000 | 119 | — |

One final example is a CSI dataset from human calf muscle. The dataset was gathered as a $12 \times 12 \times 8$ set, zerofilled, and Fourier transformed to $16 \times 16 \times 8$ voxels as descibed for $8 \times 8 \times 8$ datasets previously (*17*). Using the proton image, 156 spectra out of 2048 were selected for being within the leg in the two axial slices showing the largest cross-section of calf muscle in the proton image. PCA was used to align the 156 data spectra on the PCr frequency. Further PCA showed that there were three components in the data with very large frequency overlap among them. In Fig. 10a, one of the two axial slices from the calf muscle is shown, with a sample of the $^{31}$P CSI data in figure 10b. The data are of high SNR; however, there are no clear differences between them on initial inspection. Figures 10c and 10d show the BSD results for amplitude distribution and spectral shape, respectively. The results are an average of 50,000 samples from the posterior distribution following 25,000 steps of equilibration, which required approximately 14 h on a Digital UNIX workstation with a 600 MHz alpha processor.

A summary of the differences between the spectra is given in Table 3 and shows that there are three distinct signals arising from the calf muscle. The first and second spectral shapes are similar, except for differences in pH. The third spectral shape shows a smaller $\gamma$ATP splitting due to *J* coupling and a higher $\beta$ATP shift. In addition to their spectral differences, the components have different spatial distributions within the calf muscle as shown in Fig. 10c. The third shape is stronger in the posterior of the calf, while the first is stronger in the anterior.

**TABLE 2**
**The Misfit to the Known Input for the Highest SNR Simulation Averaged over the Entire Dataset of Mean Amplitude 7754 and Mean Spectral Peak Height $3.33 \times 10^{-3}$, Including the Estimates from BSD for the Standard Deviations**

| Linewidth (points) | Amplitude RMS misfit actual | Amplitude avg. std. dev. BSD | Spectra RMS misfit actual | Spectra avg. std. dev. BSD |
|---|---|---|---|---|
| 1 | 56 | 564 | $1.79 \times 10^{-5}$ | $1.43 \times 10^{-4}$ |
| 7 ($\sigma = 1.1$) | 66 | 690 | $1.51 \times 10^{-5}$ | $8.47 \times 10^{-5}$ |

**TABLE 3**
**The *J* Couplings, Shifts, and pH Values (from Pi Peak Position) for the Three Reconstructed Spectral Shapes in Human Calf Muscle**[a]

| | Shape 1 | Shape 2 | Shape 3 | |
|---|---|---|---|---|
| $\gamma$ATP *J* coupling | 18.3 Hz | 18.6 Hz | **15.1 Hz** | |
| $\alpha$ATP *J* coupling | 16.1 Hz | 16.4 Hz | 15.9 Hz | ±0.5 Hz |
| $\beta$ATP *J* coupling | 17.3 Hz | 17.3 Hz | **16.0 Hz** | |
| $\gamma$ATP shift | −4.87 ppm | −4.85 ppm | −4.91 ppm | ±0.04 ppm |
| $\alpha$ATP shift | −9.94 ppm | −9.96 ppm | **−10.06 ppm** | PCR at |
| $\beta$ATP shift | −18.40 ppm | −18.44 ppm | **−18.57 ppm** | −2.52 ppm |
| pH | **7.03** | 7.11 | 7.09 | ±0.02 |

[a] Key differences are shown in bold text.

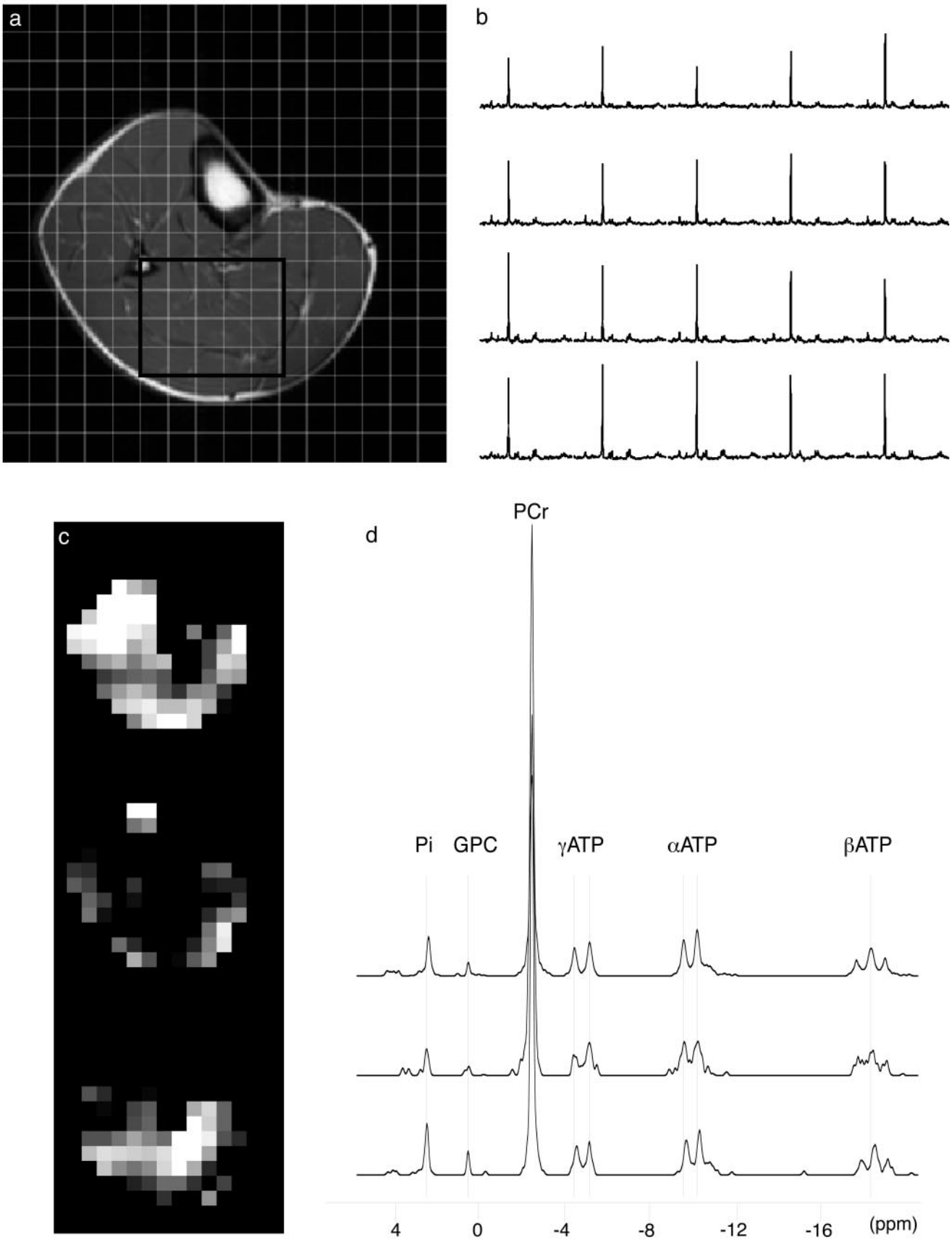
The second shape is strongest in a ring along the outer edge of the calf muscle. Initial results on other individuals indicate that the spectral shapes are consistent across individuals while their distributions show some variations. The origin of these differences is not clear; however, it seems plausible that they may be due to variations in fiber type between the muscle groups.

## DISCUSSION

We find it encouraging that in the wide variety of spectral shapes and distributions studied, BSD was able to find good solutions while using only minor constraints. For the 5-FU catabolism, PCA was used previously to obtain the same results; however, the PCA basis shapes are orthogonal and generally require *ad hoc* transformations to reconstruct the spectral shapes. These spectral shapes are then used to determine the amplitude distribution. In contrast, BSD automatically determines the spectral shape and the amplitude for 5-FU and FBAL, removing the time necessary to reconstruct the spectral shapes and removing the uncertainty involved in the final result.

While the efficiency of automatic recovery of basis spectra is useful, BSD demonstrates its real power on the larger and more complex datasets. In the case of the head data, the PCA analysis becomes more difficult. There is a problem of uniqueness in the transformation of the orthogonal shapes back into spectral shapes which is not present for BSD, which is able to determine the spectral shapes and their distribution directly. Furthermore, in the case of the head, there is an additional, *mathematically* possible solution which can be discarded based on our detailed physiological knowledge. The fact that BSD finds this solution demonstrates one of its great strengths: BSD is not constrained by our preconceived ideas on what it should find, which allows us to more fully explore the realm of possible solutions, discarding those we can, but retaining the others for further exploration.

For the calf musle, BSD offers the only method for recovering the strongly overlapping spectral shapes. In this case PCA calculates three orthogonal shapes which permit too many


a
b
c
d
PCr
Pi
GPC
γATP
αATP
βATP
4
0
-4
-8
-12
-16
(ppm)

possible reconstructions into spectral shapes. Although the three orthogonal components clearly indicate the presence of differences within the muscle spectra at a level of a few percent of the total signal, interpretation of these differences without the unique reconstruction provided only by BSD is virtually impossible. Since BSD reconstructs the actual spectral shapes as well as their amplitudes, it becomes possible to interpret the results in terms of different physical conditions. From the spectral shapes and distributions, it is clear that the calf muscle contains distinct spectral signatures, roughly aligned with the muscle groups. These signatures are present as mixtures within the individual muscles, with some types stronger within a given muscle than other types. For such a case with variations at only a few percent, BSD is the only method that we have found with a demonstrated ability to reconstruct true spectral shapes and distributions, thereby allowing analysis of their physical quantities.

These results show several of the strengths of BSD. First, through the direct sampling of the actual posterior distribution, BSD determines not only the mean results, but also the true uncertainties at each spectral point and amplitude. Some methods give uncertainty estimates by treating the distribution of solutions as Gaussian. This is highly unlikely to be true, making such estimates inaccurate and potentially misleading. Second, methods which find solutions by inversion (such as FFT procedures) are prone to artifacts in sparsely sampled sets such as those shown. BSD, on the other hand, creates possible solutions out of the vacuum and tests them against the data, eliminating such artifacts. Third, BSD identifies mathematically possible solutions. Thus, when real multiple solutions are possible, they are found. Often these additional solutions can be ruled out *a posteriori*, as in the case of the head data. However, if the multiple solutions were all physically possible, then it is really not possible to decide on a "best" solution. If a single solution in a case like this were, in fact, determined by any method, it would be extremely misleading. In contrast, by providing these multiple possible solutions BSD can guide further experimentation, allowing the discovery of correct, unique solutions when further constraints or data become available. Fourth, by determining both the spectral shapes and their fractional distribution within the voxels, BSD allows a much purer reconstruction of the spectra associated with underlying tissue which is not spatially resolved than any other method. Finally, BSD avoids biasing the results in any way. BSD only "knows" the number of underlying spectra to look for and has no preference for one spectral shape over another.

In order to constrain the solution space adequately for BSD to find acceptable solutions, certain limitations must be applied. First, the model must be a positive additive distribution, which limits the solution space to the positive hyperquadrant. Fortunately, this restriction is not terribly onerous, since this type of distribution can represent many physical problems. Second, it is necessary for the data to overdetermine the solutions, since Eq. [1] is degenerate in general. The degree of overdetermination necessary is likely to depend on the frequency overlap of the spectral shapes in the problem, since the spectra in solution space can then easily exchange flux. The calf muscle and simulation results show that for reconstruction of 3 strongly overlapping spectral shapes and their amplitude distributions, 100 spectra are adequate and probably even excessive at reasonable SNR.

## CONCLUSIONS

While a number of Bayesian methods, usually coupled with single value decomposition procedures, have been introduced to solve various bilinear problems, the results have not proven the usefulness of adding the computationally intensive Bayesian algorithms. The work presented here dramatically demonstrates the power of Bayesian methods to improve analysis of bilinear systems. Bayesian spectral decomposition operates on the simple principle that by exploring the space of all possible solutions, equivalent to the phase space of statistical mechanics, while remaining cognizant of additional prior knowledge, the "best" answer together with its uncertainties must be the result. Powerful computers have finally made possible the realization of Bayes' original idea—by simply "doing sums" over a probability distribution, you obtain the correct answer. Mathematically, we are simply asking at each point what is the likely step to take in order to find the maximum of the nearby points. Then we sum up the results as we wander in this space and we obtain the desired result. People seeing this method for the first time are often astounded by the results, since they flow from an apparently trivial statement (Eq. [2]). This is the beauty of the mathematics and the method—do the sums correctly and the correct answer *must* result.

The analyses of the three datasets presented here demonstrate that this simple concept has great power and offers a new tool for analysis of bilinear problems in a number of fields. In spectral analysis, time series analysis, and econometrics, the data often take the form of a bilinear distribution which may be amenable to analysis by BSD. Presently, we are exploring a number of extensions to our original work, including modification of the prior for the mixing matrix to take into account local correlations, methods to fix the flux scale in the two domains independent of scaling

**FIG. 10.** CSI dataset (156 spectra from 2 axial slices of human calf muscle): (a) One $^{1}$H image is shown with the 16 × 16 overlay for the CSI voxels together with a box indicating the location of the spectra shown in (b). (b) $^{31}$P spectra from the region outlined in (a) are shown (25 of 156 spectra shown). (c) The 16 × 16 amplitude distribution is shown for the three reconstructed spectral shapes. At the top is shape 1, in the middle is shape 2, and at the bottom is shape 3. Although 16 × 16 voxels are shown, the dataset does not include voxels outside the leg, so these are automatically set to zero amplitude. (d) The three reconstructed spectral shapes are shown, numbered from top to bottom. Gaussian lineshapes with widths of 5 points were used by BSD. See Table III for a summary of the peak locations and the differences between spectra.

during averaging, modifications to spectral lineshapes to allow the width to vary throughout the dataset which should be of use in mass spectrometry, and the application of BSD to other fields.

## ACKNOWLEDGMENTS

We thank John Skilling for his helpful discussions and for providing us with the Massive Inference source code. This work was funded under NIH grants CA41078 and CA62556.

## REFERENCES

1. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intelligence* **6,** 721–741 (1984).
2. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57,** 97–109 (1970).
3. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Physics* **21,** 1087–1091 (1953).
4. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science* **220,** 671–680 (1983).
5. J. Besag, P. Green, D. Higdon, and K. Mengersen, Bayesian computation and stochastic systems, *Statist. Sci.* **10,** 3–66 (1995).
6. J. Besag and P. J. Green, Spatial statistics and Bayesian computation, *J. R. Statist. Soc. B* **55,** 25–37 (1993).
7. U. Grenander and M. I. Miller, Representations of knowledge in complex systems, *J. R. Statist. Soc. B* **56,** 549–603 (1994).
8. J. Besag, On the statistical analysis of dirty pictures, *J. R. Statist. Soc. B* **48,** 259–302 (1986).
9. B. M. Hill, Bayesian forecasting of economic time series, *Econometric Theory* **10,** 483–513 (1994).
10. G. J. Marseille, R. DeBeer, A. F. Mehlkopf, and D. van Ormondt, Bayesian estimation of MR images from incomplete raw data, *in* "Maximum Entropy and Bayesian Methods" (J. Skilling and S. Sibisi, Eds.), pp. 13–24, Kluwer, Dordrecht (1996).
11. S. Sibisi and J. Skilling, Prior distributions on measure space, *J. R. Statist. Soc. B* **59,** 217–235 (1997).
12. R. Stoyanova, A. C. Kuesel, and T. R. Brown, Application of principal component analysis for NMR spectral quantitation, *J. Magn. Reson. A* **115,** 265–269 (1995).
13. T. R. Brown and R. Stoyanova, NMR spectral quantitation by principal-component analysis. II. Determination of frequency and phase shifts, *J. Magn. Reson.* **112,** 32–43 (1996).
14. C.-W. Li, W. G. Negendank, K. A. Padavic-Shaller, P. J. O'Dwyer, J. Murphy-Boesch, and T. R. Brown, Quantitation of 5-fluorouracil catabolism in human liver *in vivo* by three-dimensional localized $^{19}$F magnetic resonance spectroscopy, *Clin. Canc. Res.* **2,** 339–345 (1996).
15. J. Murphy-Boesch, R. Stoyanova, R. Srinivasan, T. Willard, D. Vigneron, S. Nelson, J. S. Taylor, and T. R. Brown, Proton-decoupled $^{31}$P chemical shift imaging of the human brain in normal volunteers, *NMR Biomed.* **6,** 173–180 (1993).
16. J. S. Taylor, D. B. Vigneron, J. Murphy-Boesch, S. J. Nelson, H. B. Kessler, L. Coia, W. Curran, and T. R. Brown, Free magnesium levels in normal human brain and brain tumors: $^{31}$P chemical-shift imaging measurements at 1.5 T, *Proc. Natl. Acad. Sci. USA* **88,** 6810–6814 (1991).
17. T. R. Brown, R. Stoyanova, T. Greenberg, R. Srinivasan, and J. Murphy-Boesch, NOE enhancements and T1 relaxation times of phosphorylated metabolites in human calf muscle at 1.5 Tesla, *Magn. Reson. Med.* **33,** 417–421 (1995).