Statistical Analysis of ESCA Spectra

Christian Ritter
CORE and Insitut de Statistique
Université Catholique de Louvain
34, voie du Roman Pays
1348 Louvain-la-Neuve
Belgium

1 General Introduction

This is a study of a practical problem for which prior information arises naturally and must be included in the analysis to provide sensible results. In the first part, we introduce ESCA and show how the problem of analyzing ESCA spectra can be formulated in a Bayesian framework. In the second part, we conduct a Bayesian analysis of a specific data set. We employ a wide range of techniques, beginning with simple normal approximations, and ending with Markov chain simulations. As a result, we arrive at a much deeper understanding of the ESCA problem.

In many research problems, data acquisition and evaluation proceeds in stages. The results of the previous stage carry on to the next stage, and knowledge accrues over time. There is a great need for statistical concepts which are consistent with this structure. The Bayesian approach has this capability, provided that the corresponding computations can be conducted in a generic fashion, allowing wide flexibility in the choice of the statistical model.

We demonstrate, in the context of the ESCA problem, that such a Bayesian analysis environment can be assembled from available building blocks. Simple and approximate methods, such as normal approximations about the posterior mode, can be augmented by tiers of more sophisticated, more computing-intensive, and nearly exact techniques, such as profile methods, marginalization using the Laplacian approximation, and Markov chain simulations. In combination, these methods allow problem-driven statistical analyses, not only for the ESCA problem, but for a much wider range of applications.

2 Introduction to ESCA for Polymers

ESCA (Electron Spectroscopy for Chemical Analysis) is an important tool for studying the chemistry of material surfaces. In the ESCA process, these surfaces are irradiated with low-energy X-rays, causing emission of photo-electrons from the surface, which can be detected and classified by their kinetic energies. ESCA spectra give the counts y_i^* of photo-electrons at each of the energies x_i, $i = 1, ..., n$, over a desired energy range $[x_a, x_b]$ for a pre-specified duration of irradiation.

The energy retained by the photo-electrons is the difference between the energy of the X-ray photons and the energy needed for freeing the electrons from their orbits. Chemical compositions and bonding structures correspond to characteristic spectra. Therefore, ESCA can be applied to *identify* elements and chemical structures.

Furthermore, in different bonding configurations, orbits of the same element, such as carbon-1s in carbon-carbon, carbon-oxygen, or carbon-nitrogen bonds, have slightly different binding energies, thereby splitting the ESCA peak into sub-peaks. The relative areas of those sub-peaks are proportional to the frequencies of the corresponding bonds in the molecule. Therefore, ESCA spectroscopy can also be used in *quantitative* analyses of complex polymer surfaces.

3 Structure of the Data and a Statistical Model

Provided that one chooses sufficiently long scanning times, the distribution of the y_i^* can be approximated by a normal distribution with energy dependent mean $\eta(x_i)$ and variance $w_i^{-1} \approx y_i^*$. Often, a background b_i is subtracted from the raw signal y_i^* to form the adjusted counts $y_i = y_i^* - b_i$. A parametric model $\eta(x_i) = \eta(x_i; \boldsymbol{\theta})$ leads to the likelihood

$$L(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}, \mathbf{w}) \propto \exp\left[-\frac{1}{2} S_{\mathbf{x},\mathbf{y},\mathbf{w}}(\boldsymbol{\theta})\right]$$

with the weighted sum of squares $S_{\mathbf{x},\mathbf{y},\mathbf{w}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[y_i - \eta(x_i; \boldsymbol{\theta})\right]^2 \cdot w_i$, which we simply call $S(\boldsymbol{\theta})$, and therefore to weighted nonlinear least squares as a fitting method.

A commonly used template for ESCA sub-peaks of simple non-metallic elements, such as carbon, is the additive Gaussian-Lorentzian function (Wertheim and DiCenzo, 1986) of the form

$$\begin{aligned} f(x; \beta, \gamma, \rho) &= \rho \cdot \exp\left\{ -\ln 2 \left[\frac{2(x-\beta)}{\gamma}\right]^2 \right\} + \\ &\quad +(1-\rho) \cdot \left\{ 1 + \left[\frac{2(x-\beta)}{\gamma}\right]^2 \right\}^{-1} . \end{aligned}$$

Here, β is the peak location, $\gamma > 0$ the bandwidth (full width at half maximum), and $0 < \rho < 1$ the mixture ratio. The maximum height of this function is one; its integral is

$$\gamma \left[\rho \cdot \sqrt{\frac{\pi}{4 \ln 2}} + (1-\rho) \cdot \frac{\pi}{2}\right].$$

Assuming linear superposition for a known number m of sub-peaks, we obtain the model

$$\eta(x; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho}) = \sum_{j=1}^{m} \alpha_j f(x; \beta_j, \gamma_j, \rho_j); \quad j = 1, \ldots, m.$$

The choice of parameters α_j, γ_j, and ρ_j subject to the constraints $0 < \alpha_j$, $0 < \gamma_j$, and $0 < \rho_j < 1$ makes the minimization of $S(\boldsymbol{\theta})$ and subsequent inference difficult. This can be avoided by substituting $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho})$ by $(\mathbf{a}, \mathbf{b}, \mathbf{g}, \mathbf{r})$ defined as $\alpha_j = \exp(a_j)$, $\beta_j = b_j$, $\gamma_j = \exp(g_j)$, and $\rho_j = \exp(r_j)/\left[1 + \exp(r_j)\right]$.

4 Difficulties Related to Peak Decomposition and Model Specification

There are two main complications in the analysis of ESCA spectra. First, some of the electrons are scattered and lose energy, due to collisions in the top atomic layers of the surface. These form a background intensity in the observed spectra which is hard to model or to adjust for. Second, peaks of orbits of similar bonds overlap, producing composite peaks which require curve fitting for analysis. The difficult problems involving peak decomposition and background removal have been under investigation for many years. Some of the key references are Perram (1968), Vanderginste and De Galan (1978), Baruya and Maddams (1978), Maddams (1980), and Powell and Seah (1990).

A typical example of overlapping sub-peaks is presented in Figure 1.

Figure 1 about here

The two plots show a complex peak, which originates from the carbon-1s orbit in plasma-polymerized methyl-methacrylate (PPMMA), a polymer used in micro-electronics manufacturing, and its decomposition into a background and four single peaks. This decomposition is very good, but not unique.

Even for a known number of sub-peaks, it can be difficult or impossible to obtain a unique weighted least squares fit. In fact, in our example, a large class of decompostions into four sub-peaks provides fits of almost indistinguishable quality. Figure 2 displays four such decompositions, and Figure 3 shows that they produce nearly equivalent fits.

Figures 2 and 3 about here

The decomposition in the lower right corner of Figure 2 is the most realistic, having sub-peaks with similar widths, while the decomposition in the upper left, corresponding to the smallest weighted sum of squares among the displayed fits, has sub-peaks of greatly differing widths and does not lead to a sensible chemical interpretation.

If a unique weighted least squares fit can be found, the nonlinearity of the model makes it difficult to assess the precision of the estimates of model parameters and derived quantities such as the ratios of peak-areas.

Finally, there is also uncertainty about the methods for background removal. Therfore, the model may have to be augmented by another parametric or nonparametric component. Several scenarios for background removal may have to be studied and compared. In our analysis, we adjust the raw counts y_i^* by an integrated background estimate b_i (Powell and Seah, 1990). Furthermore, we augment the model by the linear adjustment term $i + s \cdot (x_i - x_a)$.

5 Prior Information

Researchers working with ESCA know from their experience with standard samples that close peaks involving the same element have similar, but not necessarily identical, widths and mixture ratios. Moreover, the approximate peak-positions for the spectra of certain model compounds are known from previous experiments and have been tabulated. If no unique fit can be achieved, the experienced researcher uses this auxiliary information informally to discriminate between different fits of equal quality. Here, we show that by using Bayesian techniques, such auxiliary information can be formally incorporated into the estimation procedure. In doing so, one can not only arrive at plausible fits, but may also be better able to assess the precision of the subsequent parameter estimates.

5.1 Simple Quadratic Penalties

As we just mentioned, one usually has information about some of the parameters. Typically, much of this information concerns spacings of parameters, but not their precise values. For example, we know quite precisely that the peak widths and mixture ratios have to be similar but we know much less about their values. These values depend partly on properties of the measuring instrument, which can vary, and which cannot all be controlled. Similarly, we know a lot about the spacing of the possible sub-peaks, but not about their absolute positions, since the entire spectrum can be shifted due to surface charging. Here, we discuss how simple quadratic penalties can be constructed which are mild with respect to global shifts and strict with respect to changes of the parameters relative to each other.

Suppose that $\boldsymbol{\theta}^{(0)}$ is an *a priori* estimate of the parameter vector $\boldsymbol{\theta}$ and suppose that the deviations of the components θ_i from $\theta_i^{(0)}$ can be represented by a standard deviation σ while the deviations of the differences $\tau_{ij} = \theta_i - \theta_j$ from the $\tau_{ij}^{(0)} = \theta_i^{(0)} - \theta_j^{(0)}$ can be represented by a standard deviation ω (usually smaller than σ). Then, the information about the precisions can be summarized by a quadratic penalty

$$\left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\right)' \boldsymbol{\Sigma}_P^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\right)$$

to the sum of squares with penalty matrix

$$\boldsymbol{\Sigma}_P = \begin{pmatrix} \sigma^2 & t & \cdots & t \\ t & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t \\ t & \cdots & t & \sigma^2 \end{pmatrix},$$

where $t = \sigma^2 - \omega^2/2$.

In Bayesian terms, this penalty corresponds to a multivariate normal prior distribution with mean vector $\boldsymbol{\theta}^{(0)}$ and covariance matrix $\boldsymbol{\Sigma}$. Usually, one requires that parameters for which a multivariate normal prior is to be chosen are unconstrained. This holds for the transformed parameters $\mathbf{a}$, $\mathbf{b}$, $\mathbf{g}$, and $\mathbf{r}$.

5.2 Constructing a Prior Distribution for Our Example

5.2.1 Peak Heights

Little is known beforehand about the peak heights and therefore about the parameters a_j and we assume a flat prior for them.

5.2.2 Peak Locations

For our polymer, we can expect sub-peaks corresponding to C–C, C–O, C=O, O–C–O, and O–C=O bonds, and therefore construct a model with four peaks (O–C–O and C=O create sub-peaks of almost identical locations). Using the tabulated peak positions from Clark and Thomas (1976a) and Clark and Thomas (1976b), and our judgement of the precision of the absolute peak positions and their spacing, we can construct a quadratic penalty of the previously introduced type with $\mathbf{b}^{(0)} = (289.0, 287.8, 286.5, 285.0)'$, $\sigma_b = 2$, and $\omega_b = 1$. In our example, it is apparent that the whole spectrum is shifted by about $1.2eV$ due to surface charging, and we adjust by this amount.

5.2.3 Peak Widths and Shapes

In our measurements, carbon 1s peaks are commonly between $1eV$ and $2eV$ wide and can be approximated by Gaussian–Lorentzian mixtures of 60% to 95% Gaussian content. From previous experiments we derive an average peak width of $1.8eV$ and an average mixture ratio of 88% Gaussian. Furthermore, we know that typically the peak widths and mixture ratios do not differ by more than $0.1eV$ and 5%, respectively. Therefore, we choose priors with $\mathbf{g}^{(0)} = (0.6, 0.6, 0.6, 0.6)'$, $\sigma_g = 0.5$, and $\omega_g = 0.1$ for the widths and $\mathbf{r}^{(0)} = (2, 2, 2, 2)'$, $\sigma_r = 2$, and $\omega_r = 0.5$ for $\mathbf{r}$, the vector of the transformed mixture ratios.

5.3 Combined Penalty

Together, we obtain the combined penalty

$$\begin{aligned} P(\mathbf{a},\mathbf{b},\mathbf{g},\mathbf{r}) &= \left(\mathbf{b}-\mathbf{b}^{(0)}\right)' \boldsymbol{\Sigma}_b^{-1} \left(\mathbf{b}-\mathbf{b}^{(0)}\right) \\ &+ \left(\mathbf{g}-\mathbf{g}^{(0)}\right)' \boldsymbol{\Sigma}_g^{-1} \left(\mathbf{g}-\mathbf{g}^{(0)}\right) \\ &+ \left(\mathbf{r}-\mathbf{r}^{(0)}\right)' \boldsymbol{\Sigma}_r^{-1} \left(\mathbf{r}-\mathbf{r}^{(0)}\right) \end{aligned}$$

with

$$\mathbf{b}^{(0)} = \begin{pmatrix} 290.2 \\ 289.0 \\ 287.7 \\ 286.2 \end{pmatrix}, \boldsymbol{\Sigma}_b = \begin{pmatrix} 4.0 & 3.5 & \cdots & 3.5 \\ 3.5 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 3.5 \\ 3.5 & \cdots & 3.5 & 4.0 \end{pmatrix};$$

$$\mathbf{g}^{(0)} = \begin{pmatrix} .6 \\ .6 \\ .6 \\ .6 \end{pmatrix}, \boldsymbol{\Sigma}_g = \begin{pmatrix} .250 & .245 & \cdots & .245 \\ .245 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & .245 \\ .245 & \cdots & .245 & .250 \end{pmatrix};$$

and

$$\mathbf{r}^{(0)} = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{\Sigma}_r = \begin{pmatrix} 4.000 & 3.875 & \cdots & 3.875 \\ 3.875 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 3.875 \\ 3.875 & \cdots & 3.875 & 4.000 \end{pmatrix}.$$

The combined penalty corresponds to independent multivariate normal priors for the parameters $\mathbf{b}$, $\mathbf{g}$, and $\mathbf{r}$, while a flat prior is chosen for $\mathbf{a}$. The resulting posterior can be written as a nonlinear least squares likelihood for an augmented model with pseudo-data. Thus, the usual nonlinear least squares algorithms can be used to find the posterior mode, if it exists.

6 Finding the Posterior Mode and the Approximate Standard Errors

The posterior density, corresponding to the previously introduced prior, has a unique mode which can be found by nonlinear least squares.

Table 1 about here

Table 1 shows the resulting parameter estimates and their approximate standard errors. These standard errors were computed from the normal approximation to the posterior distribution at the mode. The corresponding fit is shown in Figure 4. The fit is good and the residuals do not reveal systematic deviations. The first order autocorrelation coefficient of the residuals is $r = -0.09$, which is not significantly different from zero.

Figure 4 about here

7 Using Posterior Profiles for Assessing the Quality of the Normal Approximation to the Posterior

Profile methods (Bates and Watts, 1988; Ritter, 1992) are very general tools for summarizing a function F of several variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ in a neighborhood of a maximum $\hat{\boldsymbol{\theta}}$. In our context this function is the posterior density. In the profiling operation each parameter component θ_i is varied in turn, and for each value $\theta_i = t$ the function F is maximized with respect to the other components. The resulting conditional maximizer $\tilde{\boldsymbol{\theta}}_i(t)$, taken as a function of t, is called the ith profile trace, and $\tilde{F}_i(t) = F(\tilde{\boldsymbol{\theta}}_i(t))$ is called the i-th profile value. Accordingly, we call the structure $\left\{\tilde{\boldsymbol{\theta}}_i(\cdot), \tilde{F}_i(\cdot)\right\}$ the i-th profile. The profile value can be interpreted as the shadow of F when seen from the θ_i axis. Posterior profiles can be used to investigate whether a normal approximation of the posterior is appropriate; in addition, more elaborate inference techniques can be based on posterior profiles.

However, the plain posterior profile $\left\{\tilde{\boldsymbol{\theta}}_i(\cdot), \tilde{\pi}_i(\cdot)\right\}$ is not always a good way of summarizing the posterior along the θ_i axis, since it ignores that for some values of θ_i the conditional maximum $\tilde{\boldsymbol{\theta}}_i(\theta_i)$ may be narrower than for others. If the conditional maximum is wide, many other settings of the remaining parameters correspond to fits of almost equal quality. If the conditional maximum is narrow, only few other settings are as good as $\tilde{\boldsymbol{\theta}}_{i}(\theta_i)$. The *modified* profile adjusts for such imbalances by multiplying $\tilde{\pi}_{i}(\cdot)$ by a width adjustment factor of the form $|\tilde{\mathbf{R}}_{-i}(\theta_i)|^{-1/2}$ where

$$\tilde{\mathbf{R}}_{-i}(t) = \frac{\partial^2}{\partial \boldsymbol{\theta}_{-i} \partial \boldsymbol{\theta}'_{-i}} \log \pi(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_{i}(t)},$$

and $\boldsymbol{\theta}_{-i}$ is the vector $\boldsymbol{\theta}$ without its i-th component. The determinant of the Hessian matrix $\tilde{\mathbf{R}}_{-i}(\cdot)$ is a measure of the local width of the conditional maximum $\tilde{\boldsymbol{\theta}}_{i}(\cdot)$; it is small when the maximum is wide. The modified profile values are proportional to the Laplacian approximations to the component-wise marginals (Leonard, 1982; Tierney, Kass and Kadane, 1989; Leonard, Hsu and Tsui, 1989).

In our example, the modified posterior profiles can be computed for all parameters over large intervals. Figure 5 shows the resulting modified profile values (solid) contrasted with the marginals of the normal approximation corresponding of the posterior at the posterior mode (dotted).

Figure 6 shows the pairwise plots of the profile traces of the log height parameters a_1, a_2, a_3, and a_4. If the posterior is approximately multivariate normal, the posterior profile traces are nearly straight lines.

Figures 5 and 6 about here

From Figure 5, one can conclude that the modified profile values are not well-approximated by the normal approximations. Moreover, the profile traces corresponding to the height parameters have sudden bends and even jumps. The pairwise diagnostic plots provide a clear warning that inference about linear or nonlinear functionals of the parameters derived by the delta method from the approximate standard errors and correlations may be strongly misleading.

8 Using an Interesting Posterior Profile to Constuct Scenarios for Decompositions

As is evident from the modified profile of the parameter a_2, there are three distinct likely configurations, the first at about $a_2 = 5.3$, the second at about 6.0, and the third at about 6.6. Figure 7 shows the corresponding fits. The upper left plot, for $a_2 = 5.3$, shows a decomposition into four distinct sub-peaks which is consistent with our auxiliary knowledge. The upper right plot, for $a_2 = 6.0$, shows a decomposition into three sub-peaks where the right (carbon-carbon) sub-peak is split. The third plot, for $a_2 = 6.6$, shows a decomposition into three sub-peaks with a very small fourth peak in the right tail of the spectrum.

Figure 7 about here

Since there are two minor modes corresponding to three peak fits and a dominating mode corresponding to a four peak fit, we conclude that the spectrum together with our auxiliary knowledge provides strong but not sure evidence of four peaks and therefore of the existence of the O–C–O or C=O bonds.

9 Using Markov Chains to Deepen the Analysis

9.1 General Comments

The posterior can also be investigated by sampling parameter vectors $\left\{\boldsymbol{\theta}^{(i)}\right\}_{i=1}^{N}$ from it and studying the corresponding fits. If a large number of such parameter vectors has been drawn, pairwise scatterplots of the $\boldsymbol{\theta}^{(k)}$ can be used to explore the important part of the parameter space which contains fits which are nearly equivalent to the optimal one. Such displays are convenient tools for detecting unusual features of the posterior such as multiple modes. However, it is usually difficult to sample from the posterior directly, unless it is of a simple form. Here, we use two methods specifically designed for complicated posteriors, the Griddy Gibbs sampler (Ritter and Tanner, 1992), and the Griddy Hastings algorithm (Ritter, 1992), also known as the sliding Lattice approach (Tierney, 1991). Both are members of the large family of algorithms for creating ergodic Markov chains with desired stationary distributions. For additional details on such algorithms, we refer to Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953), Hastings (1970), Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), and Müller (1991).

9.2 Preparing our Example for Markov Chain Exploration

To be able to use Markov chain sampling techniques in our example, we need to assure that we are indeed working with a proper posterior. In our example we do this by restricting the parameter space to a finite domain containing all reasonable parameter vectors. For the log height parameters, we choose intervals from 2 to 8.5. Peaks of log height less than 2 are invisible in the noise, wherease a peak of log height 8.5 is clearly higher than any part of the observed spectrum. For the locations we impose the intervals $[284, 291.5]$. The first value is clearly below the carbon-carbon peak locations, and the second value is clearly higher than the expected position of the O–C=O peak.

For log peak widths and for the logistic transforms of the mixture rations we choose the intervals $[-0.5, 1.5]$ and $[0, 8]$ respectively and for the intercept and the slope we consider the intervals $[-200, 100]$ and $[-20, 20]$ as sufficiently large.

9.3 Conducting a Markov Chain Analysis of the Posterior

In our example, we construct 100 parallel chains. This approach is convenient but not very efficient; more efficient implementations use only one long chain. We choose multiple chains since we want be able to easily monitor the progress of the sample from iteration to iteration.

We start with 100 parameter vectors uniformly distributed over the parameter domain and use the Griddy Gibbs sampler with 15 grid points until for all parameters the plots of the sample quantiles versus the iterations indicate that the sample has stabilized. In our case this occurs after about 200 iterations. We continue the sampling using the Griddy Hastings algorithm with five grid points. Since the posterior is quite complicated, convergence is very slow, and even after 5000 iterations the width parameters g_1 and g_2 have not completely stabilized.

In Figure 8 the traces of the 0th, 25th, 50th, 75th, and 100th percentiles of the location parameters indicate that the first peak quickly assumes a position at about $290eV$ while the other three peaks initially share the position of the main peak of the spectrum at about $286.5eV$. Later, the second and the third peak begin to move away from the fourth peak. Eventually, the second peak assumes positions between $289eV$ and $288eV$, the third peak $288eV$ and $286.5eV$, and the fourth peak locks in at about $286.2eV$.

Figure 8 about here

In order to compare the results from the Markov chain run with the results obtained by profiling we generate estimates of marginal densities of all parameters from the Markov chain sample after the 5000th iteration. Figure 9 shows the modified profiles and the marginal histograms from the Markov chain sample of size 100 (see Figure 11 for a comparison of the modified profiles (solid) and smooth density estimates of the Markov chain sample (dotted) obtained by the method of Kooperberg and Stone, 1991).

Figure 9 about here

We conclude that the overall agreement is quite good. Possible reasons for the remaining differences in some of the parameters are that either the Laplacian approximation is not good for these parameters or that the Markov chain had been stopped prematurely.

Figure 10 shows the pairwise plots of the Markov chain sample of the height parameters at the last iteration.

Figure 10 about here

This plot is in agreement with the profile traces shown in Figure 6.

9.4 Assessing the Effect of Ignoring the Prior Information

Using the Markov chain approach, we can also investigate what happens when the prior information is ignored. Since without prior no unique fit can be obtained, the profile methods cannot be used here. Figure 11 shows smooth density estimates derived from a Markov chain sample of the posterior (dotted), and smooth density estimates derived from a Markov chain sample of the plain likelihood (dashed). The Markov chain sample of the plain likelihood (restricted to the same finite domain as the posterior and therefore made integrable) was generated similarly to the Markov chain sample of the posterior by starting with 100 uniformly distributed parameter vectors, and then running the Griddy Gibbs sampler for several hundred iterations followed by 5000 iterations of the Griddy Hastings algorithm. Also here, although most of the parameters had stabilized, full convergence could not be achieved. The smooth density estimates were computed using log-spline density estimation (Kooperberg and Stone, 1991).

The striking features of Figure 11 are the complete loss of precision regarding the mixture parameters and a preference for mixtures with a widened second peak, where the second peak is moved to the right, and peaks 3 and 4 are either forming a split peak, or peak 3 takes over the position of peak 4, pushing it into the right tail of the spectrum. Such decompositions do not correspond to sensible chemical interpretations.

We therefore conclude that the observed spectrum contains very little information about the mixture ratios and by itself does not suggest a sensible peak structure. Only by introducing all available auxiliary information can the decomposition produce useful results.

Figure 11 about here

10 Conclusions

Using modern statistical techniques, we can investigate decompositions of ESCA spectra in great detail. In order to obtain safe and sensible interpretations, it is essential to incorporate as much expert information as possible into the fitting process. In well-conditioned cases simple summaries consisting of parameter estimates and standard errors are sufficient. If there is doubt, the profile methods can be used to assess the adequacy of these summaries. If there is discussion about the prior information in the fitting, Markov chain simulations can be used to investigate what happens, if this information is weakened. Here, it is not necessary that a unique fit still exists. In this context, even if the Markov chain algorithms do not converge completely, the intermediate results can provide useful information to the researcher.

A major drawback of using Markov chain methods to study ESCA decompostions is the large amount of computing involved. In our example 5000 iterations in the Markov chain were conducted, corresponding to roughly $100,000,000$ evaluations of the posterior. Profile methods are much less computing intensive, but they only work if a unique fit can be achieved, and if the profiles can be computed sufficiently far away from the posterior mode.

11 Acknowledgements

The author would like to thank Søren Bisgaard, Milan Buncick, and Jeff Tobin from the Engineering Research Center of Plasma Aided Manufacturing, and Ngoc Tran from the Material Sciences Department at the University of Wisconsin-Madison for their help and encouragement. The author would also like to thank Douglas Bates, Martin Tanner, and Tom Leonard for introducing him to some of the new statistical methods used in this research. This work was supported by the National Science Foundation under grant ECD-8721545 and by the European Commission under grant ERBCHBGCT920167.

References

Baruya, A. and Maddams, W. F. (1978). An examination of the uniqueness of Gaussian and Lorentzian profiles, *Applied Spectroscopy* **32**: 563–566.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, John Wiley, New York.

Clark, D. T. and Thomas, H. R. (1976a). Applications of ESCA to polymer chemistry. X. Core and valence energy levels of a series of polyacrylates, *Journal of Polymer Science: Polymer Chemisty Edition* **14**: 1671–1700.

Clark, D. T. and Thomas, H. R. (1976b). Applications of ESCA to polymer chemistry. XI. Core and valence energy levels of a series of polymethacrylates, *Journal of Polymer Science: Polymer Chemisty Edition* **14**: 1671–1700.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**: 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Kooperberg, C. and Stone, C. J. (1991). Logspline density estimation for censored data, *Technical Report 226*, Department of Statistics, University of Washington.

Leonard, T. (1982). Comments on "A simple predictive density function", *Journal of the American Statistical Association* **77**: 657–658.

Leonard, T., Hsu, J. S. J. and Tsui, K. W. (1989). Bayesian marginal inference, *Journal of the American Statistical Association* **84**: 1051–1058.

Maddams, W. (1980). The scope and limitations of curve fitting, *Applied Spectroscopy* **34**: 245–267.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *J. of Chem. Physics* **21(6)**: 1087–1092.

Müller, P. (1991). A generic approach to posterior integration and Gibbs sampling, *Technical Report 91-09*, Department of Statistics, Purdue University.

Perram, J. W. (1968). Interpretation of spectra, *Journal of Chemical Physics* **49**: 4245.

Powell, C. J. and Seah, M. P. (1990). Precision, accuracy, and uncertainty in quantitative surface analyses by auger-electron spectroscopy and x-ray photoelectron spectroscopy, *Journal of Vacuum Science and Technology (A)* **8(2)**: 735–763.

Ritter, C. (1992). *Modern Inference for Nonlinear Least Squares Regression*, PhD thesis, Department of Statistics, University of Wisconsin-Madison.

Ritter, C. and Tanner, M. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy Gibbs sampler, *Journal of the American Statistical Association* **87**: 861–868.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association* **82**: 528–541.

Tierney, L. (1991). Markov chains for exploring posterior distributions, *Technical Report 560*, School of Statisitics, University of Minnesota.

Tierney, L., Kass, R. and Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions, *Biometrika* **76**: 425–433.

Vanderginste, V. G. M. and De Galan, L. (1978). Critical evaluation of curve fitting in infrared spectroscopy, *Annals of Chemistry* **47**: 2124–2132.

Wertheim, G. K. and DiCenzo, S. B. (1986). Least squares analysis of photoemission data, *Journal of Electron Spectroscopy and Related Phenomena* **40**: 57–67.

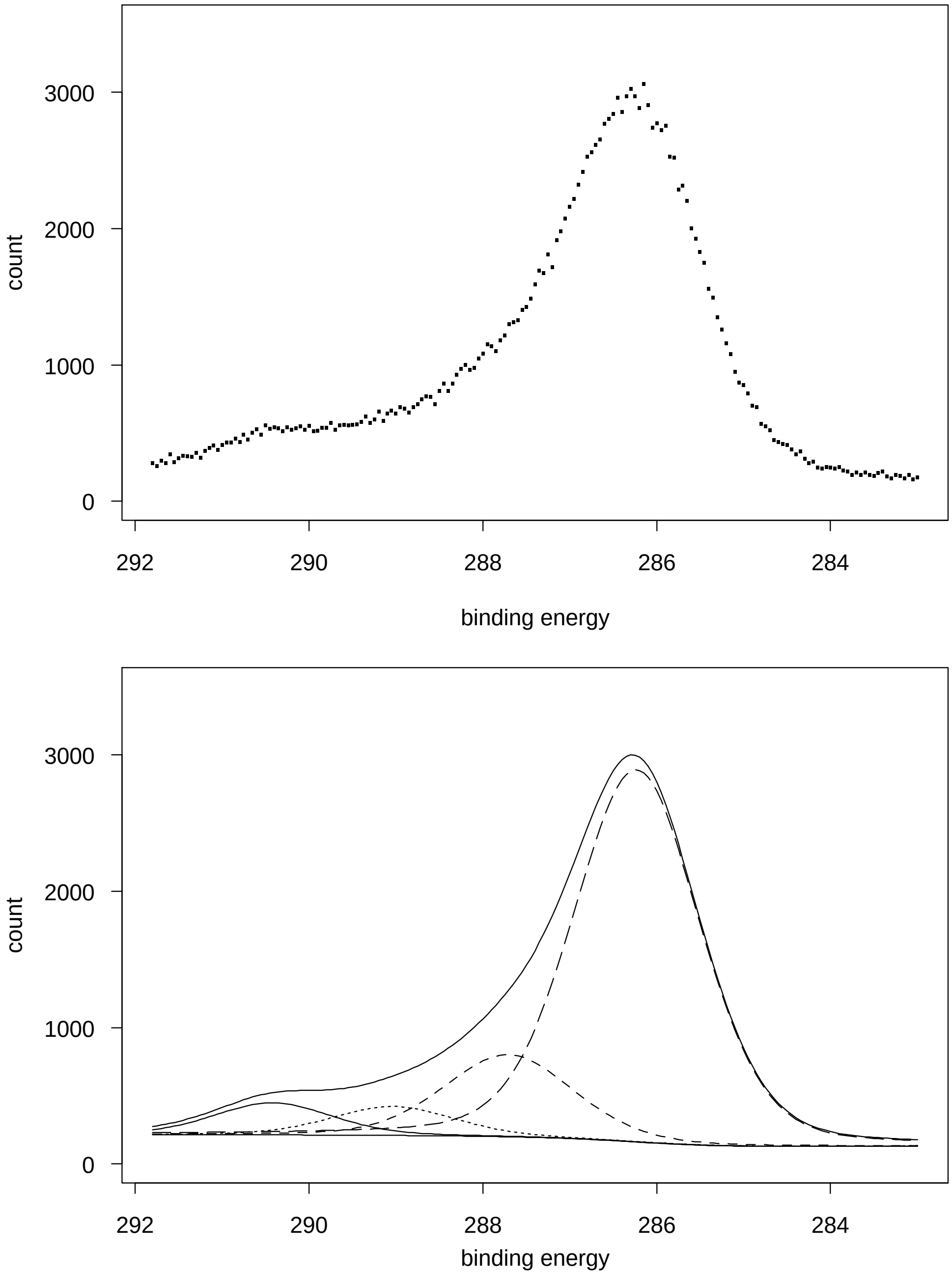

Figure 1: C1s Peak, Poly-Methyl-Methacrylate. Upper plot: Raw data. Lower plot: Decomposed peak with background. Commonly in plots of ESCA spectra the energy (horizontal) axis is reversed (decreasing) since higher binding energies correspond to lower observed energies in the electron detector.

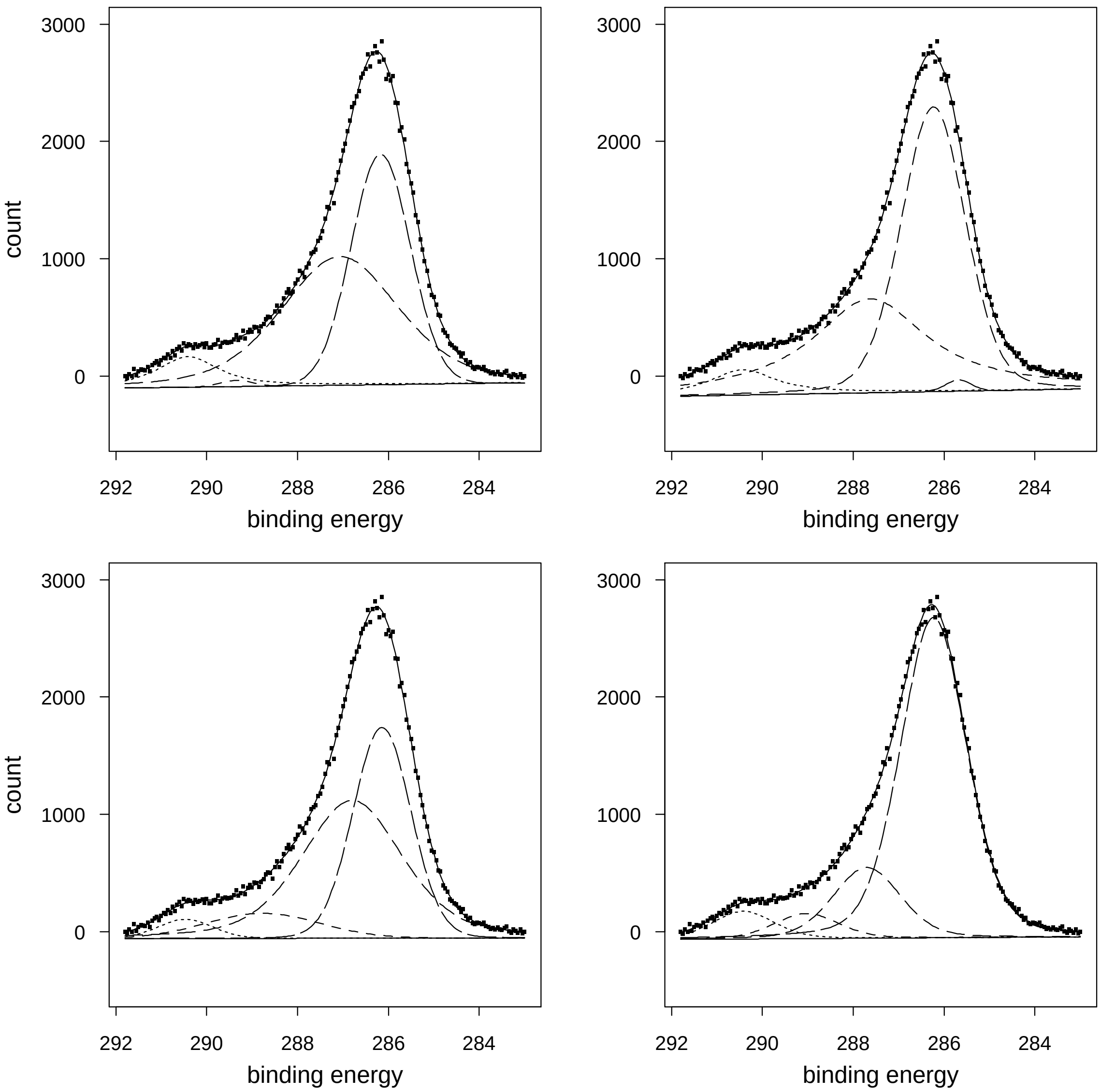

Figure 2: Four different decompositions. Broken lines: sub-peaks; solid lines: superpositon and background; dots: data points

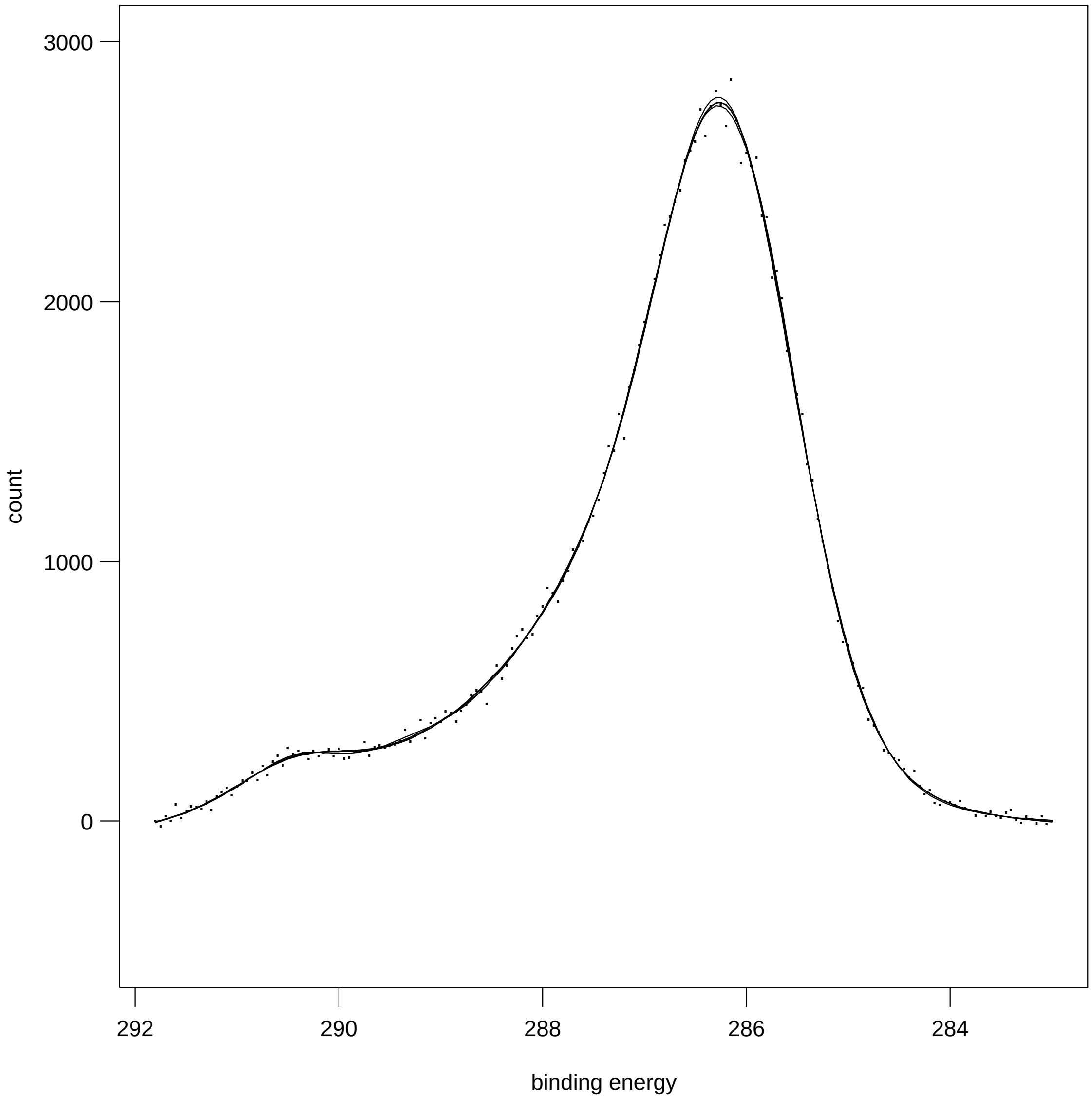

Figure 3: Four fits plotted on top of each other together with the data

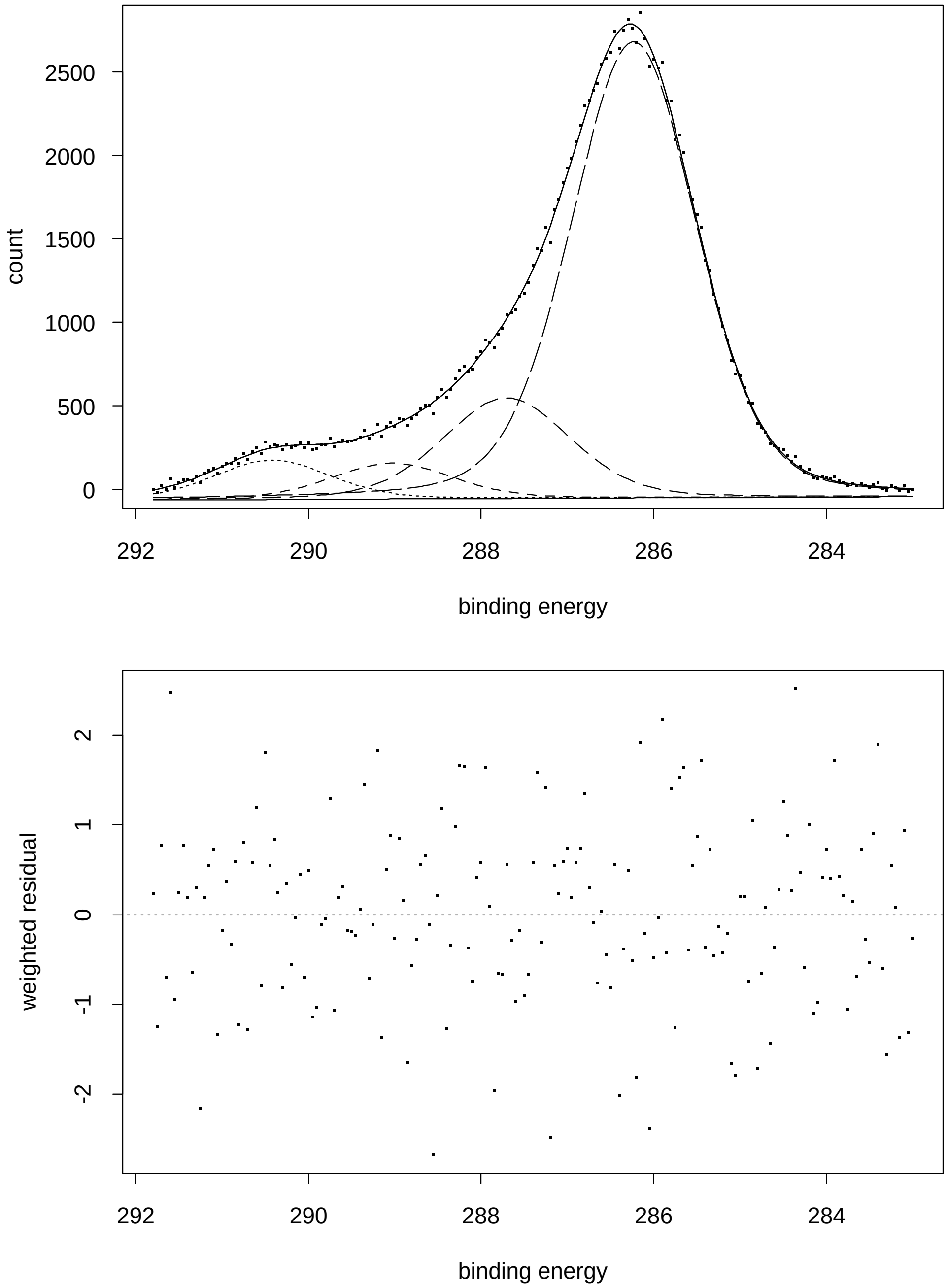

Figure 4: Fit of penalized model and weighted residuals.

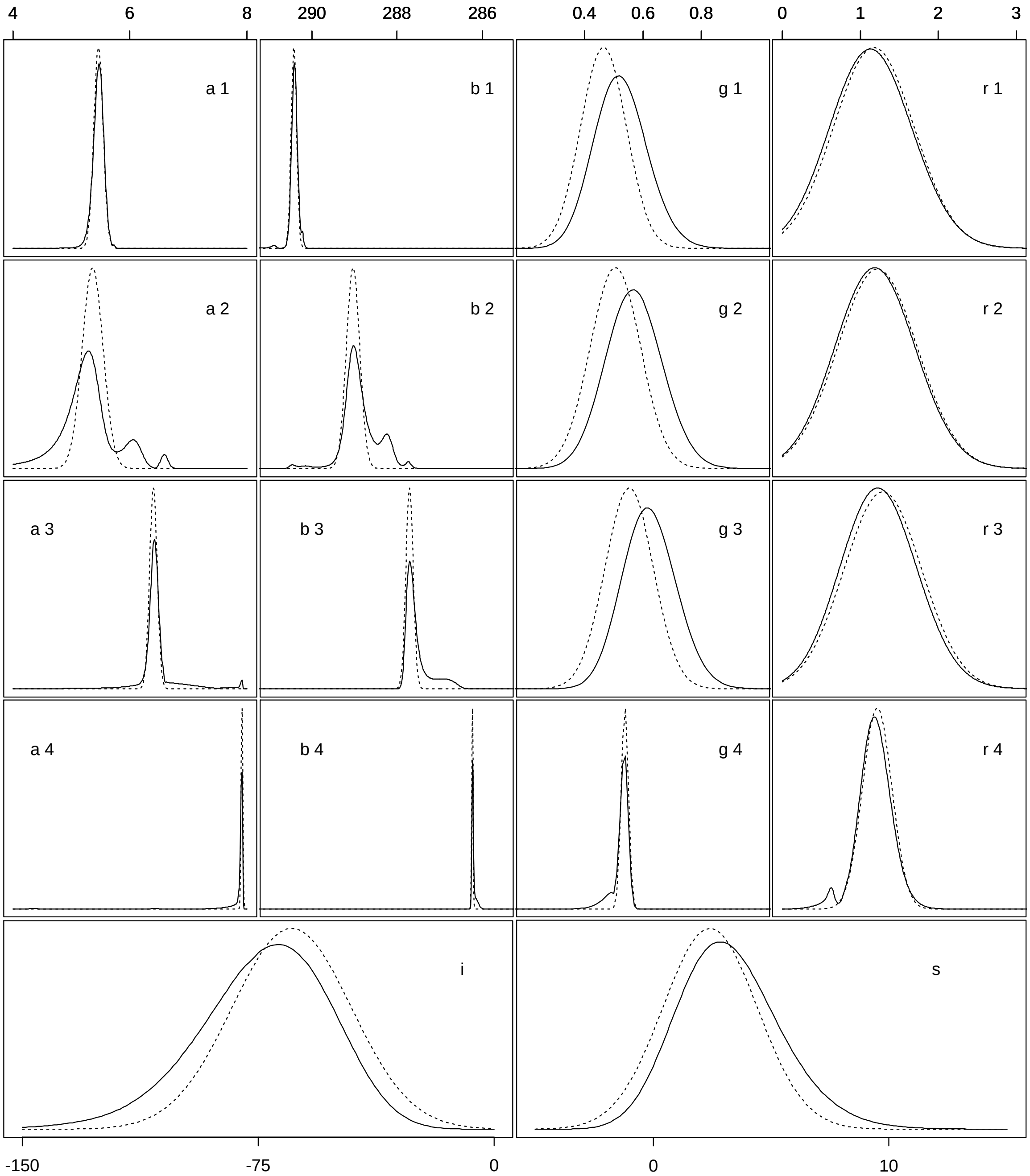

Figure 5: Modified profiles, i.e. Bayesian marginals, (solid) and normal approximations (dotted). All curves have been normalized to integrate to one over the displayed intervals.

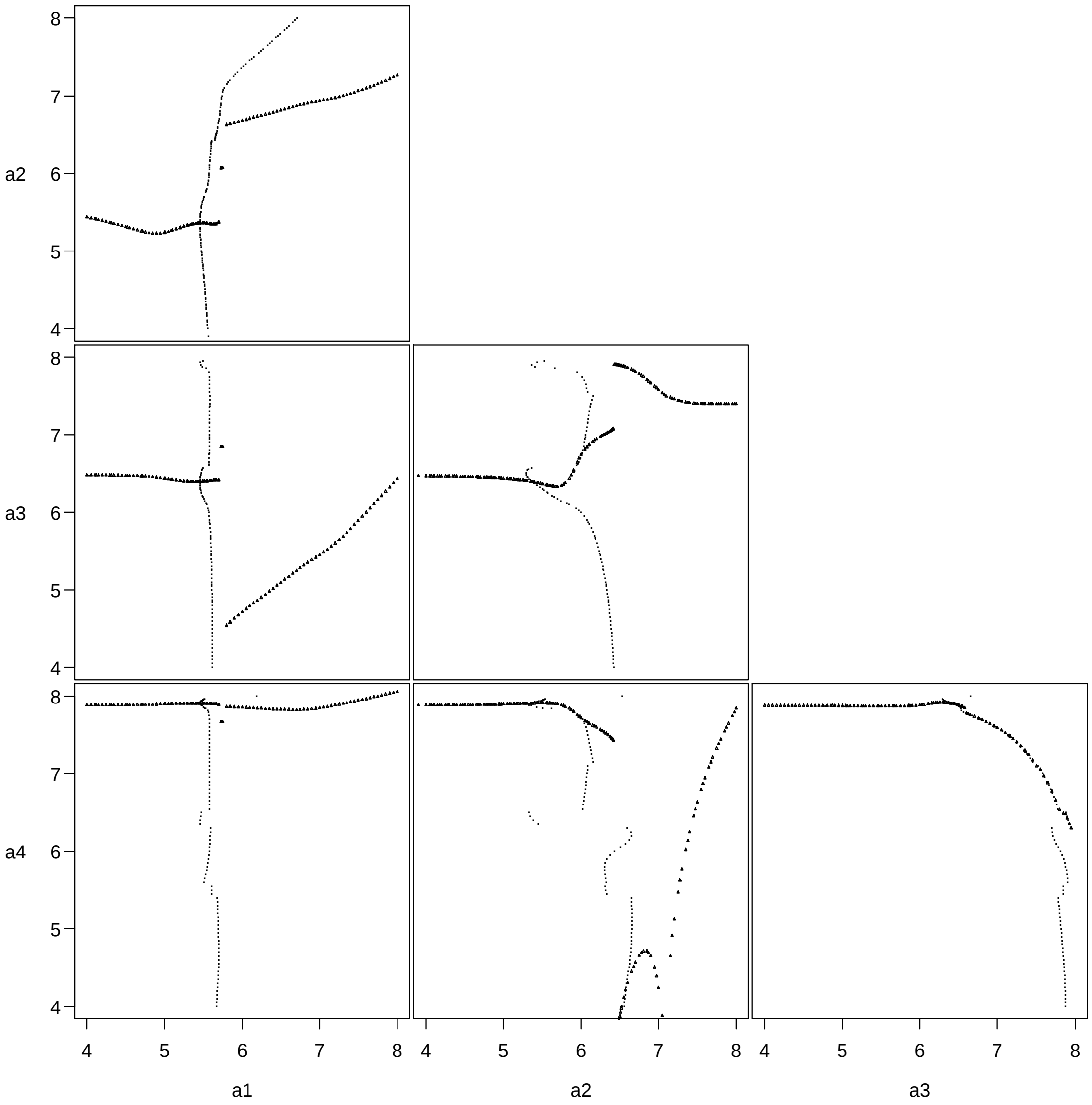

Figure 6: Pairwise profile diagnostic plots for the height parameters.

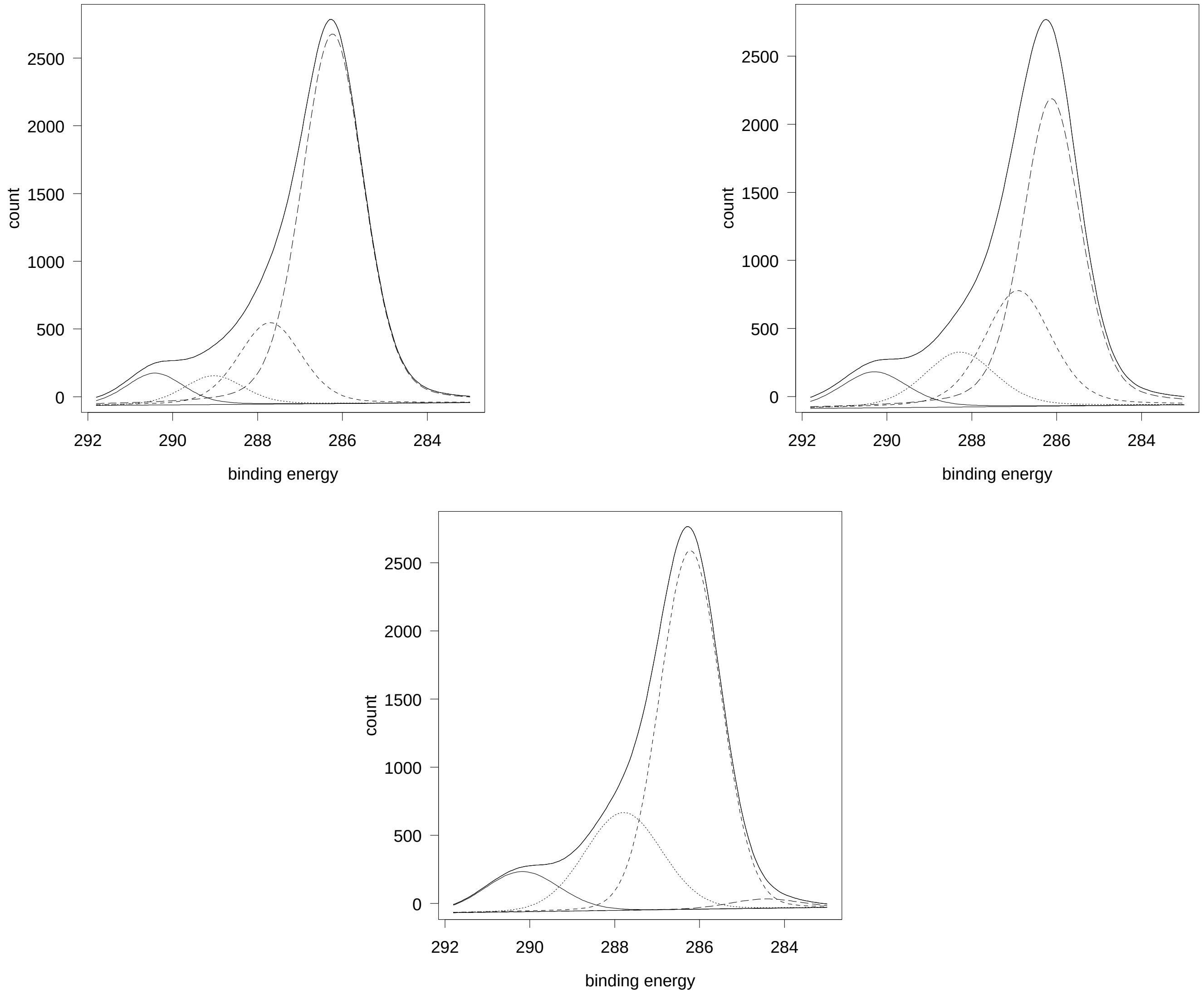

Figure 7: Three fits corresponding to the three modes in the marginal distribution of a_2.

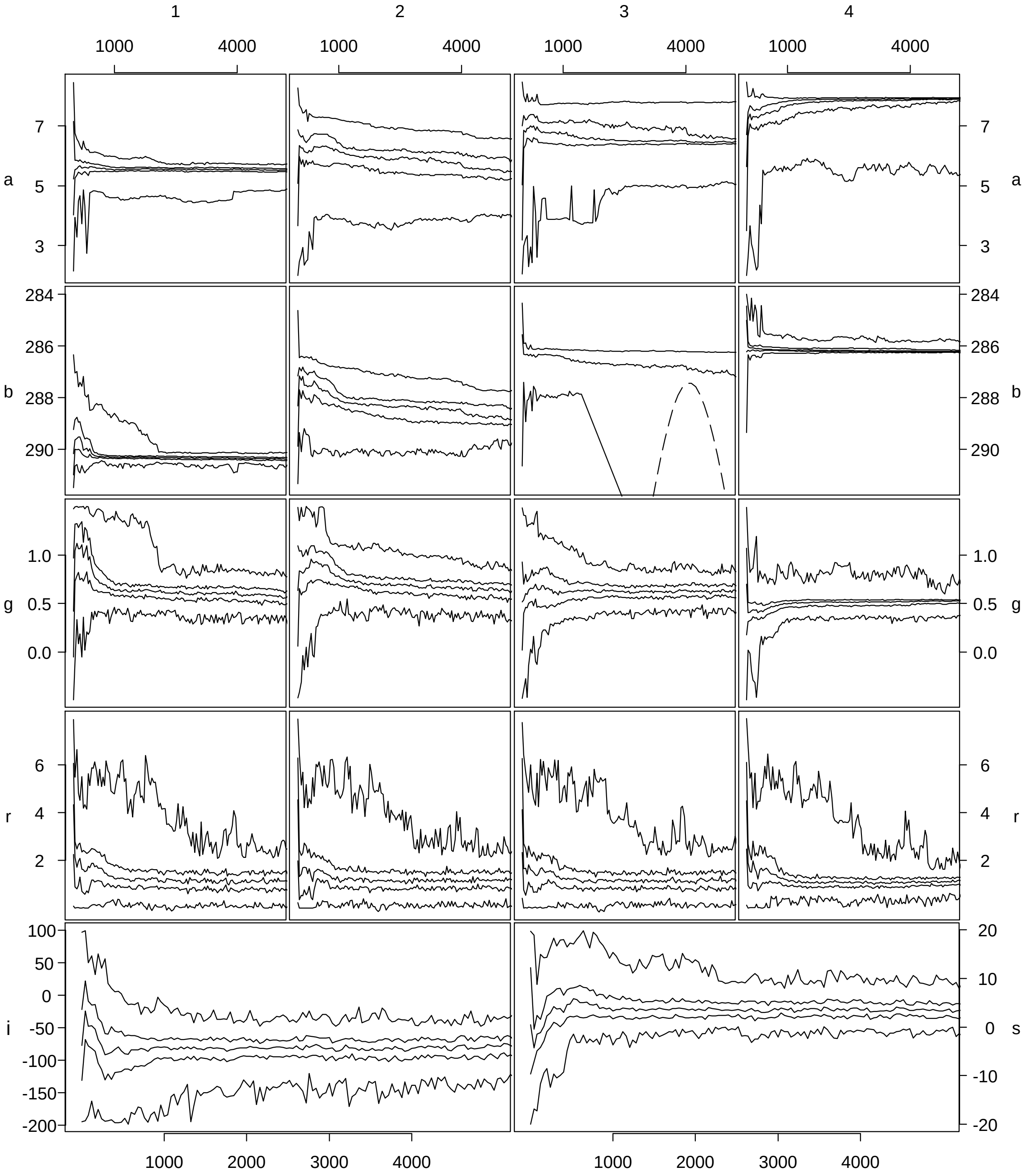

Figure 8: Quartiles (0th, 25th, 50th, 75th, 100th percentiles) by iteration for all parameters. First 200 iterations: Griddy Gibbs sampler, remaining iterations: Griddy Hastings algorithm.

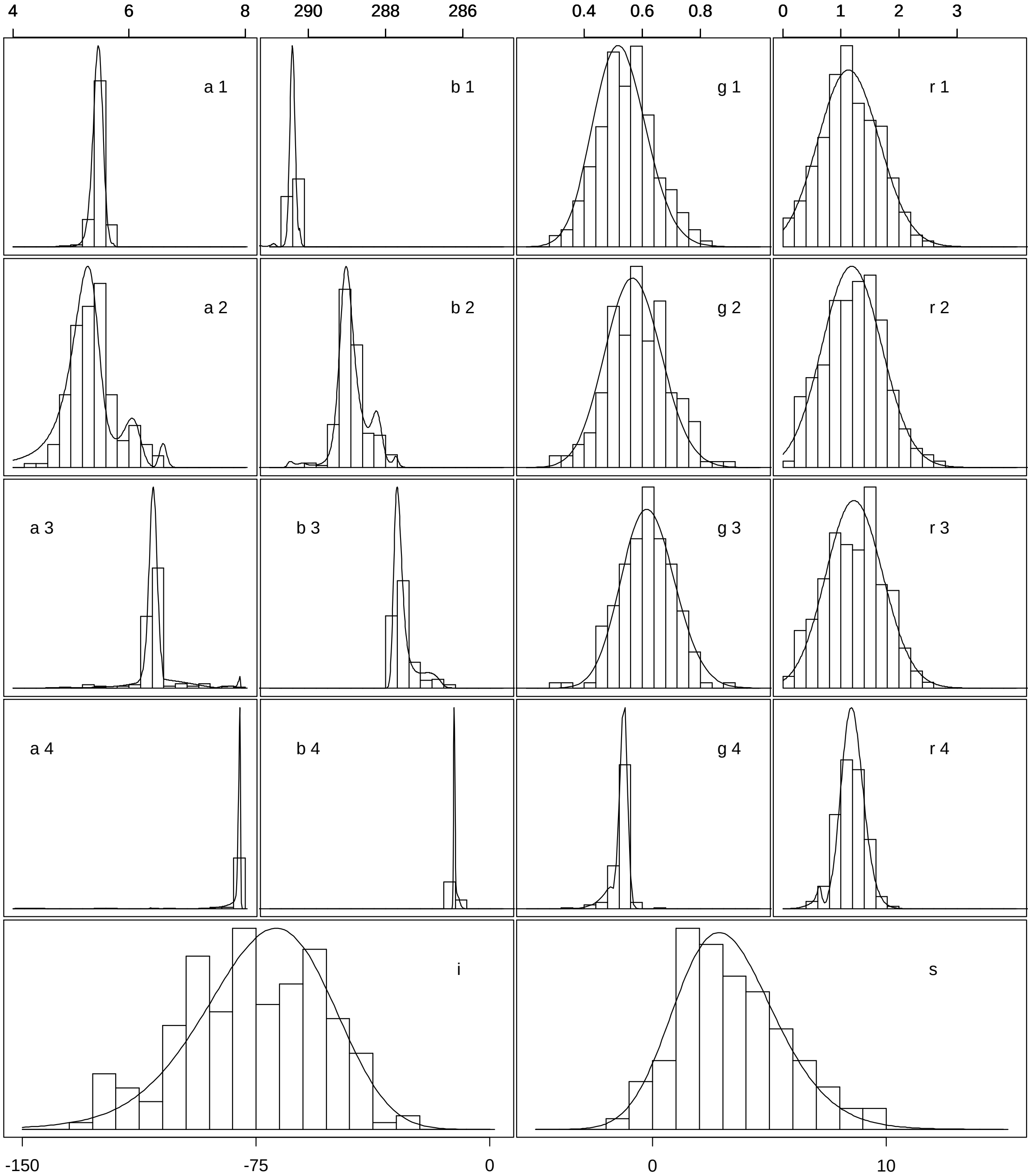

Figure 9: Approximate marginals for all parameters. Smooth: Laplace's approximation, histogram: Markov chain sample. Smooth representations of the marginals of the Markov chain sample can be found in Figure 11

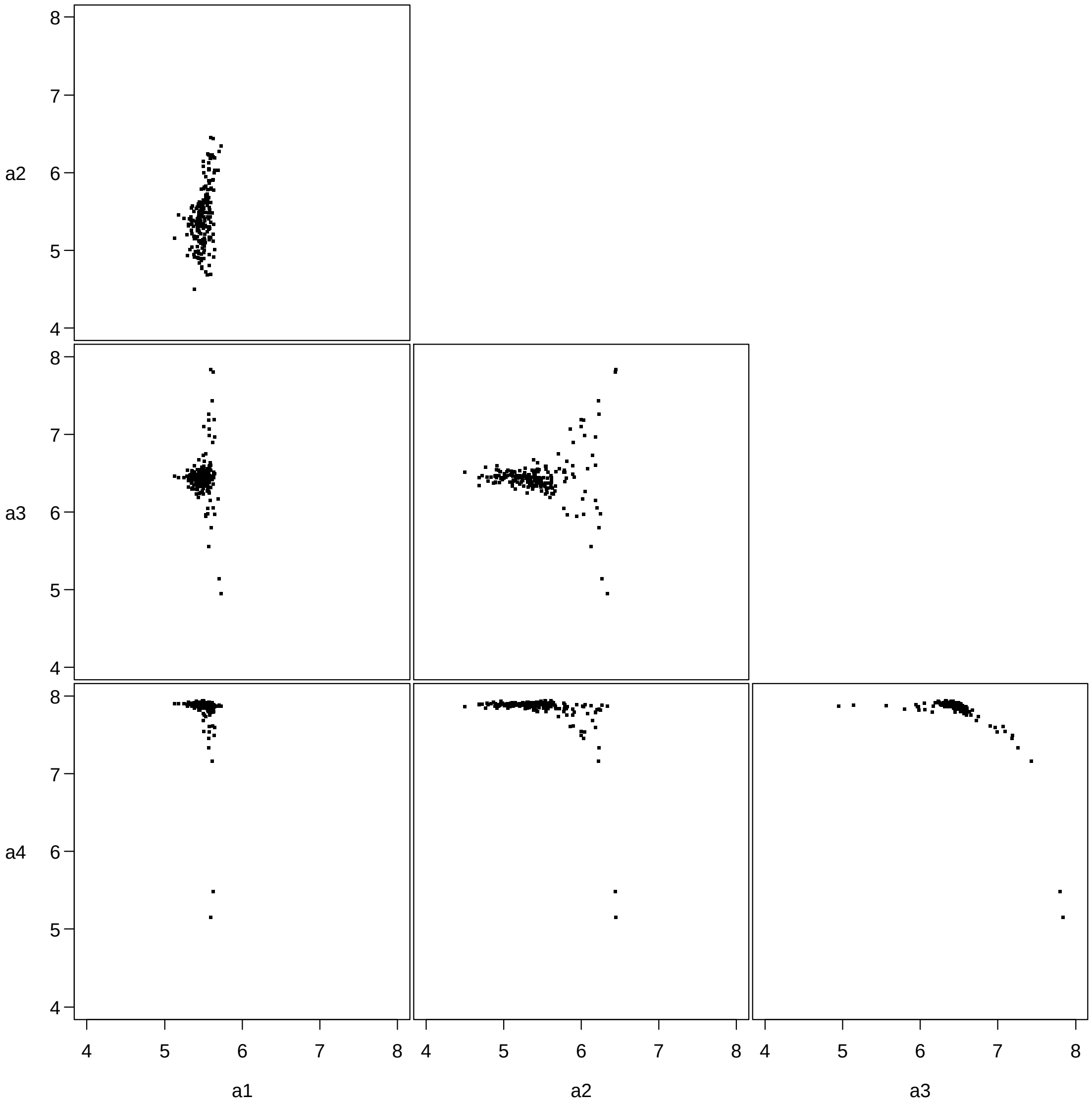

Figure 10: Pairwise scatter-plots of the height parameters of the Markov chain sample after 5000 iterations

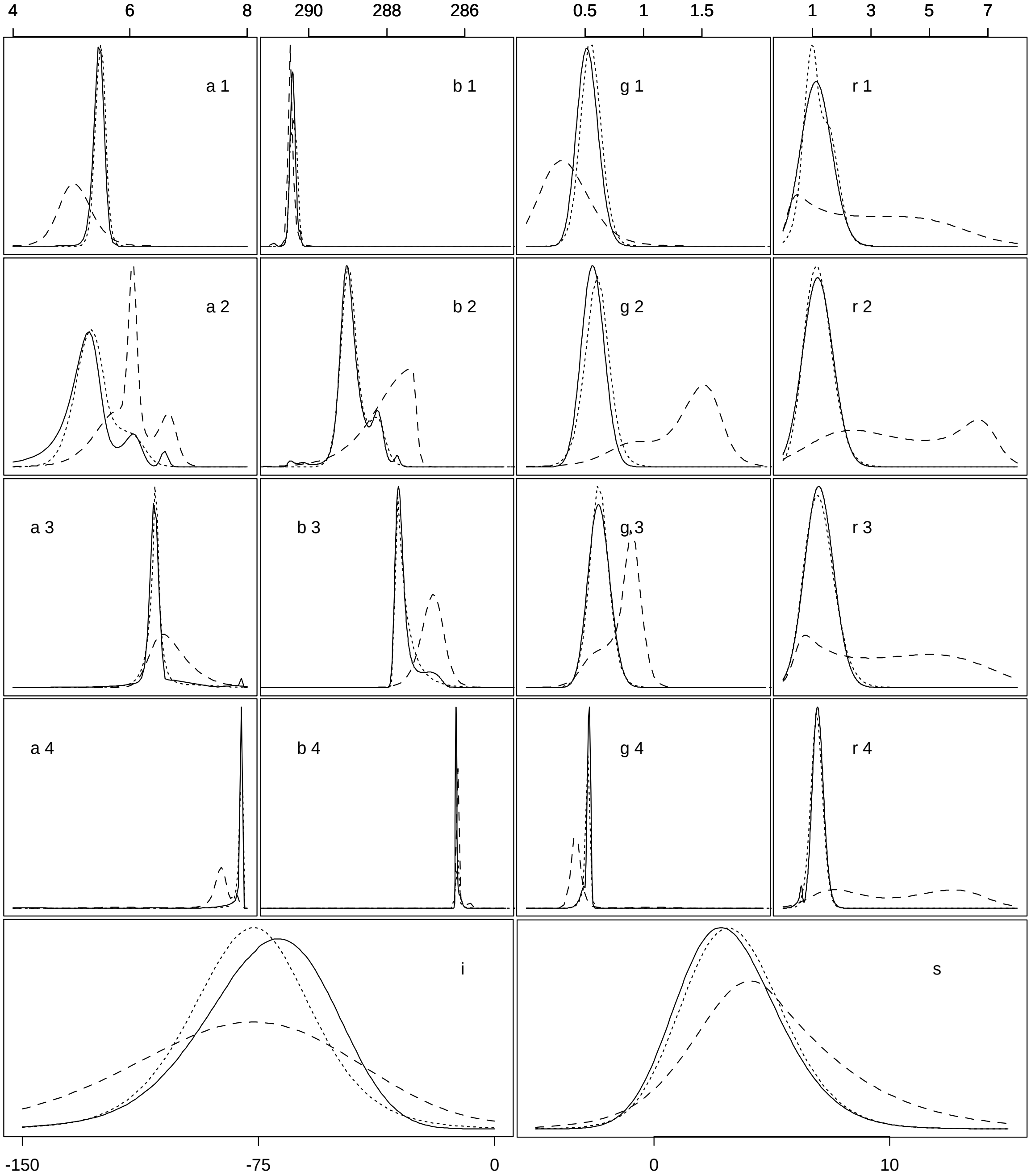

Figure 11: Comparison of penalized and unpenalized problem. Modified profiles of the penalized problem (solid), smooth density estimates of a Markov chain sample of the penalized problem (dotted), smooth density estimates of a Markov chain sample of the unpenalized problem (dashed).

	Value	Std. Error
a_1	5.46	0.08
a_2	5.37	0.18
a_3	6.40	0.07
a_4	7.91	0.01

	Value	Std. Error
b_1	-290.42	0.06
b_2	-289.03	0.16
b_3	-287.70	0.09
b_4	-286.23	0.02

	Value	Std. Error
g_1	0.46	0.08
g_2	0.50	0.09
g_3	0.55	0.08
g_4	0.54	0.01

	Value	Std. Error
r_1	1.24	0.52
r_2	1.29	0.51
r_3	1.35	0.50
r_4	1.21	0.19

	Value	Std. Error
i	-62.90	18.77
s	2.27	1.99

Table 1: Parameter estimates and approximate standard error