



Universiteit
Leiden
The Netherlands

Statistical methods for mass spectrometry-based clinical proteomics

Kakourou, A.A.

Citation

Kakourou, A. A. (2018, March 8). *Statistical methods for mass spectrometry-based clinical proteomics*. Retrieved from <https://hdl.handle.net/1887/61138>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/61138>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/61138> holds various files of this Leiden University dissertation

Author: Kakourou, Alexia

Title: Statistical methods for mass spectrometry-based clinical proteomics

Date: 2018-03-08

5

Bayesian variable dimension logistic regression with paired proteomic measurements

Abstract

We explore the problem of variable selection in a case-control setting with high-resolution mass-spectrometry data consisting of paired measurements. Each pair corresponds to a distinct isotope cluster and each component within each pair represents a cluster summary of isotopic expression derived based on two different types of information: a) the overall intensity and b) the shape of the observed isotope cluster. Our objective is to identify a collection of isotope clusters associated with the disease outcome, on the one hand, and optimally integrate the two sources of information, on the other. We propose a Bayesian model formulation which exploits the paired structure of the proteomic data and allows us to assess the added-value of the shape source beyond the intensity source in predicting the class outcome of an individual while maintaining predictive performance. We evaluate the Bayesian selection model on proteomic data, structured into intensity-shape pairs, from a pancreatic cancer case-control study. We show results from an *a-posteriori* analysis of the

This chapter has been submitted for publication as : Alexia Kakourou and Bart Mertens (2017). Bayesian variable dimension logistic regression with paired proteomic measurements.

fitted model to give insight into which are the most interesting isotope clusters for further investigation as well as into the relative predictive power of shape when combined with intensity. Additionally, we present a simulation study to demonstrate how the method behaves under a controlled setting.

5.1 Introduction

Proteomics is the large-scale study of proteins which aim to provide a better understanding of the function of cellular and disease processes at the protein level. The most widely-used technology to assess proteomic expression is mass-spectrometry which has undergone remarkable evolution over the last twenty years. Particularly ultrahigh-resolution mass spectrometers (MS) such as Fourier-transform MS have become the most powerful and efficient tools for the quantitative analysis of complex protein mixtures in biological systems.

Irrespective of its type, which may characterize a mass-analyzer with respect to resolution, mass accuracy and sensitivity, a mass-spectrometer takes as input a molecular mixture and outputs a so-called mass spectrum. A mass spectrum (as shown in Figure 4.1 of chapter 4) is a sequence of intensity readings distributed over a mass range, generated from the detection of ionized molecules. In ultrahigh-resolution mass spectrometry, each species (such as peptide) is detected and expressed as a cluster of peaks, rather than a single peak, in the mass spectrum. These peaks represent ions of the same elemental composition but different isotopic composition due to the presence of additional neutrons in their nucleus. We refer to this set of isotope peaks as the isotopic cluster.

The late improvements in mass-spectrometry technologies, and thus the quality of the acquired data, turned the focus of recent research towards methods for optimally extracting and interpreting high-resolution mass spectral data at the individual level. Insufficient attention has been given however to statistical methods for optimally summarizing and analyzing the resulting data, especially in the context of clinical applications such as calibration of diagnostic rules for disease status allocation of patients. Moreover, while several approaches have been proposed in the literature for processing mass spectral proteomic data using knowledge on the properties of isotopes such as peak detection algorithms or deisotoping methods, the predictive potential of the isotope cluster information, inherent in high-resolution mass-spectrometry data had not been fully exploited until recently.

In a recent work Kakourou et al. (2016) proposed an approach for summarizing the proteomic expression in individual mass spectra by exploiting the isotope clustering information. In their paper they propose an approach which utilizes the known statistical properties of MALDI-FTICR MS data on isotope distributions in order to reduce the complete expression in the mass spectra to clusters of isotopic expression. They investigate various ways of translating the isotopic expression in each of the derived clusters into cluster summaries by using information on either the intensity or the shape of the observed isotope cluster pattern with the objective of assessing the impact of the different choices on prediction.

While summary measures based on intensity aim to capture, for each patient, the overall intensity level of the observed isotope cluster pattern, summary measures based on shape aim to estimate/describe the shape of the observed isotopic cluster pattern or, alternatively, the deviation of that observed pattern from the typical pattern (as defined in shape paper). The estimation of shape is invariant under altering the overall intensity level of an isotope cluster (for example under adding a constant to the log-transformed intensities or multiplying the absolute intensities of a cluster with a factor), hence we can assume that shape summary measures are, on a conceptual level at least, independent of intensity summary measures. Using the derived intensity and shape estimates as new input variables (separately) into a prediction model, the authors showed that both types of information are predictive of the health status of an individual though intensity has greater predictive capacity as compared to shape.

Having established the presence of an overall isotope cluster effect as well as the presence of a shape effect in addition to intensity effect, the authors further aimed to investigate the possibility of enhancing the predictive performance of the classification rule when integrating both types of information, or in other words, evaluate the additional predictive value of shape beyond intensity. The authors addressed this question by considering a “naive” combination based on stacking the intensity and shape summary measures. The conclusion from this analysis was that the additional value of shape information, if any, was insufficient as to allow for improved predictions when “naively” combined with intensity information.

Following up on the previously reported results asserting the predictive power of the pancreatic cancer data, and importantly the predictive ability of isotope clusters in high-resolution MS data, the aim of this work is twofold: a) to identify a collection of isotope clusters associated with the disease outcome and b) to optimally integrate the two sources of information. We wish to address these questions in a way which will allow us to assess the added-value of shape beyond intensity in predicting the class outcome of an individual while maintaining predictive performance. We propose an approach which makes use of the prior knowledge about the relative predictive power of each individual source as well as the distinctive structure of the proteomic data, resulting from the fact that intensity and shape measures are tied together in pairs of isotope expression. To explore the problem of isotope selection and at the same time address the problem of assessing the additional predictive value of shape, we use a Bayesian model formulation by which we can introduce multiple layers of selection. In doing so, we make the explicit assumption that the shape source is complementary. In terms of model fitting, this is translated by assuming that a shape measure can be included in the set of predictors on the condition that it is accompanied by/coupled with its corresponding intensity while the reverse does not need to hold. This assumption allows us to make simultaneous inference on which isotope clusters are the most informative with respect to the class outcome and for which isotopes - and to what extent - shape has a complementary value in separating the two groups.

The remainder of this chapter is organised as follows: We first introduce the pancreatic cancer data and their paired structure which is a key component of these particular data.

The problem of isotope selection, on the one hand, and assessment of the added-value of shape, on the other, through a Bayesian model formulation is explored next. We then present results from applying the Bayesian model to the pancreatic cancer data consisting of the paired intensity-shape measurements. Additionally, we show a simulation example to demonstrate how the proposed method behaves under a controlled setting. We finish with a discussion.

5.2 Data

In this chapter we re-analyse data from a case-control study which was carried out at the Leiden University Medical Centre. The study involved 273 individuals, consisting of 88 pancreatic cancer patients and 185 healthy volunteers. The samples collected from those individuals were distributed over three MALDI-target plates and thereafter mass-analyzed by a MALDI-FTICR MS system, giving rise to a single mass spectrum for each sample within the mass range of 1013 to 3700 Da (full details on the design and measurement protocol can be found in Nicolardi et al. (2014)).

In previous work (Kakourou et al., 2016), the authors applied to the pancreatic cancer data a peak detection algorithm in order to identify the isotopic clusters and their corresponding peaks. As a result, the complete proteomic expression in the individual spectra was reduced to clusters of isotopic expression on which summary measures could be defined. To derive the cluster summaries the authors proposed to use information on either the intensity or the shape of the observed isotope cluster pattern.

In this work, rather than considering single measurements of intensity or shape as our predictor variables, we recognize intensity and shape are tied together and regard them as paired measurements such that the components of each predictor pair represent cluster summaries of isotopic expression based on intensity information, in the case of the first component, and shape information, in the case of the second. The intensity component of a pair is denoted in the following by u and is defined as the sum of log-transformed peak intensities $l_j, j = 1, \dots, J$ within an isotope cluster (where J denotes the number of peaks in a cluster and is cluster-specific), given by

$$u := \sum_j l_j$$

The shape component, denoted by v , is defined as

$$v := \sum_j j p_j$$

the center of gravity of a distribution on the values $1, \dots, J$, where $p_j := \frac{x_j}{\sum_j x_j}$ and x_1, \dots, x_J denote the residuals which measure the deviation of the observed isotopic pattern from the typical pattern. For a more detailed description on how these summary measures were created/derived we refer the interested readers to (shape paper). We choose

these particular summary measures to represent the intensity-based and shape-based components due to their superiority, as compared to other proposed summary measures, with regard to their individual predictive ability.

Our final data set consists of 1289 pairs of intensity and shape summary measures. Our objective is to investigate whether we can develop methods to integrate both types of information in a way which will allow us to calibrate more interpretable rules and/or learn more about the interplay between intensity and shape in predicting the class outcome, while maintaining predictive performance.

5.3 Bayesian variable-selection model on intensity-shape pairs

5.3.1 The logistic regression model

Let the data be given by (\mathbf{y}, \mathbf{z}) , where $\mathbf{y} = (y_1, \dots, y_n)^\top$ is the binary case-control outcome with $y_i \in \{0, 1\}$ for $i = 1, \dots, n$ independent individuals while $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top$ represents the predictor source with each $\mathbf{z}_i = ((u_{i1}, v_{i1}), \dots, (u_{ip}, v_{ip}))$ consisting of a sequence of paired intensity and shape measurements for p isotope clusters. We consider the binary regression model

$$y_i \sim \text{Benoulli}(p_i)$$

with

$$\text{logit}(p_i) = \beta_0 + \mathbf{z}_i \boldsymbol{\beta}$$

where p_i is the case-probability for the i th observation and $\boldsymbol{\beta} = ((a_1, b_1), \dots, (a_p, b_p))$ represents the vector of paired regression parameters with the first and second elements of each pair corresponding to the effects of the intensity and shape measurements respectively.

5.3.2 Variable-dimension logistic regression model

We assume that only a subset of isotope clusters is relevant for predicting the class outcome and that \mathbf{u} is expected to carry more information on the class outcome than \mathbf{v} . Our main objective is to assess the added-value of the shape source \mathbf{v} on top of the intensity source \mathbf{u} in predicting the health status of future patients. Under the assumption that only a set of isotope clusters is predictive of the health status of an individual, the true model is given by

$$\begin{aligned} \text{logit}(p_i) &= \beta_0 + \tilde{\mathbf{z}}_i \tilde{\boldsymbol{\beta}} \\ &= \beta_0 + (\tilde{a}_1 \tilde{u}_{i1} + \tilde{b}_1 \tilde{v}_{i1}) + (\tilde{a}_2 \tilde{u}_{i2} + \tilde{b}_2 \tilde{v}_{i2}) + \dots + (\tilde{a}_k \tilde{u}_{ik} + \tilde{b}_k \tilde{v}_{ik}) \\ &= \beta_0 + \sum_{j=1}^k (\tilde{a}_j \tilde{u}_{ij} + \tilde{b}_j \tilde{v}_{ij}) \end{aligned} \quad (5.1)$$

where $\tilde{\mathbf{z}} = ((\tilde{u}_1, \tilde{v}_1), \dots, (\tilde{u}_k, \tilde{v}_k))$ represents the unknown set of paired measurements which are associated with the class outcome with regression coefficient vector $\tilde{\boldsymbol{\beta}} = ((\tilde{a}_1, \tilde{b}_1), \dots, (\tilde{a}_k, \tilde{b}_k))$ which is of unknown dimension $k \leq p$. The isotope dimensionality k is thus considered/treated as a parameter in the model and will be estimated from the data along with the unknown set of intensity-shape pairs and their corresponding regression coefficients.

In order to assess the added-value of the shape information over and above the intensity information we place a logical constraint on the inclusion of shape information which specifies that if $a = 0$ then $b = 0$ (if $b \neq 0$ then $a \neq 0$). The above formulation suggests that shape can be included in the model only on the condition that its corresponding intensity is included as well. Note that the reverse is not true. Hence, rather than forcing mutual selection of both components of the isotope pairs we let the data decide whether the first component (intensity) alone provides all the required information for separating the two groups. With this constraint, (5.1) reduces to

$$\text{logit}(p_i) = \beta_0 + \sum_{j=1}^{k_C} (\tilde{a}_j \tilde{u}_{ij} + \tilde{b}_j \tilde{v}_{ij}) + \sum_{l=1}^{k_I} \tilde{a}_l \tilde{u}_{il}$$

where k_C denotes the dimensionality of the “complete” isotope couples and indicates how many times a shape measure is included in the model in conjunction with its corresponding intensity and k_I denotes the dimensionality of intensity singletons so that $k = k_I + k_C$.

5.3.3 Prior specification

To complete the model formulation, we have to specify the prior structure for all the model parameters. For the intercept we assume a weakly informative normal prior $\beta_0 \sim N(0, 10^2)$. We specify independent normal priors on the regression parameters $\tilde{a}_j \sim N(0, \sigma_a^2)$ and $\tilde{b}_j \sim N(0, \sigma_b^2)$, for $j = 1, \dots, k$ and $b_j \neq 0$, where the variances $\sigma_a^2 = \tau_a^2 c_a$ and $\sigma_b^2 = \tau_b^2 c_b$ control the magnitude of included effects for intensity and shape respectively, τ_a and τ_b are known re-scaling factors while $c_a = 1/s_a$ and $c_b = 1/s_b$ are randomly distributed scale factors with gamma priors placed on s_a and s_b . Under the above prior assumption on the regression parameters, the covariance matrix Σ has a block-diagonal structure with block matrices along the diagonal of the form $\Sigma_j = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}$, if $b_j \neq 0$, and $\Sigma_j = \sigma_a^2$ otherwise. The prior specification is completed by assigning a prior to the isotope dimension parameter k . We use a discrete uniform prior on the set of integers $\{0, 1, 2, \dots, k_{max}\}$ with k_{max} a large positive integer corresponding to the maximum allowed isotope dimension.

5.3.4 MCMC model fitting

We present an adaptation of the reversible jump MCMC implementation described in Mertens (2016) for fitting the logistic model, in order to perform isotope selection, on the one hand, and assess the added-value of shape beyond intensity, on the other. Evaluation of the added-value of shape is achieved by introducing a second layer of shape selection on top of the isotope selection.

Our MCMC sampler is based on a random choice between 3 basic move steps: 1) BIRTH, 2) DEATH and 3) CHANGE. The first two of these steps propose moves between different isotope dimensions while the last one proposes moves within an isotope dimension and between different variable dimensions.

More specifically, in the BIRTH step, we propose with probability $b_k = 1/3$ to add a new randomly chosen isotope into the model set. In the DEATH step, we propose with probability $d_k = 1/3$ to remove a randomly chosen isotope from the current model set. Shape selection is facilitated by splitting each of these steps into two additional substeps such that, in the case of a BIRTH move proposal, we may choose between proposing to add either the entire pair (couple) into the model, with probability $b_k^C = b_k/2$, or solely the first component of the pair (intensity), with probability $b_k^I = b_k/2$. Note that the candidate set from which we may select a new isotope is the set containing all “complete” isotopes which do not currently have any component in the model set. Analogously, in the case of a DEATH move proposal, we may choose between proposing to remove either a complete isotope (couple) from the current set, with probability $d_k^C = d_k/2$, or an intensity singleton, with probability $d_k^I = d_k/2$.

Apart from the BIRTH and DEATH steps which propose moves between isotope dimensions by either adding or removing isotopes (singletons or pairs) to and from the current set, we may propose to change the composition of an isotope already included in the model, with probability $c_k = 1/3$. We do so by introducing a CHANGE step. Again here, within this step, we may choose between two substeps which either change an isotope couple into an intensity singleton i.e. remove shape from the included isotope pair, with probability $c_k^{C \rightarrow I} = c_k/2$, or change an intensity singleton into an isotope couple i.e. add shape to the included intensity singleton, with probability $c_k^{I \rightarrow C} = c_k/2$. In this way, we give a second chance to shape selection/deselection by allowing the data to judge whether a shape measure contributes to classification in addition to intensity and thus should join its corresponding intensity or does not provide any additional information over and above intensity and therefore could be omitted from the isotope pair in the model set.

We choose

$$b_{(k=0)}^I = b_{(k=0)}^C = d_{(k=k_{max}=k_I)}^I = d_{(k=k_{max}=k_C)}^C = c_{(k=k_{max}=k_I)}^{I \rightarrow C} = c_{(k=k_{max}=k_C)}^{C \rightarrow I} = 1/2,$$

$$d_{(0 < k_I, k_C < k_{max}=k)}^I = d_{(0 < k_I, k_C < k_{max}=k)}^C = c_{(0 < k_I, k_C < k_{max}=k)}^{C \rightarrow I} = c_{(0 < k_I, k_C < k_{max}=k)}^{I \rightarrow C} = 1/4,$$

$$b_{(k_I=0 < k_C < k_{max})}^I = b_{(k_I=0 < k_C < k_{max})}^C = d_{(k_I=0 < k_C < k_{max})}^I = c_{(k_I=0 < k_C < k_{max})}^{C \rightarrow I} = 1/4,$$

$$b_{(k_C=0 < k_I < k_{max})}^I = b_{(k_C=0 < k_I < k_{max})}^C = d_{(k_C=0 < k_I < k_{max})}^I = c_{(k_C=0 < k_I < k_{max})}^{I \rightarrow C} = 1/4,$$

$$b_{(k=k_{max})}^I = b_{(k=k_{max})}^C = d_{(k_I=0)}^I = d_{(k_C=0)}^C = c_{(k_C=0)}^{C \rightarrow I} = c_{(k_I=0)}^{I \rightarrow C} = 0$$

and

$$b_k^I = b_k^C = d_k^I = d_k^C = c_k^{C \rightarrow I} = c_k^{I \rightarrow C} = 1/6$$

in all other cases. Under this specification, the acceptance probability of a Metropolis-Hastings step for a proposal move from a model with parameters θ to a new model with parameters θ' is

$$A = \min \left\{ 1, \frac{P(D|\theta') p(\theta') q(\theta|\theta')}{P(D|\theta) p(\theta) q(\theta'|\theta)} \right\},$$

the ratio of marginal likelihoods of the newly proposed model to that of the old, prior and proposal distributions, with D the data and θ the vector containing the current values of the model parameters k , σ_a^2 and σ_b^2 , (see Appendix for detailed derivation of acceptance probabilities for all move types).

Note that with the above prior specification, normal inverse scaled chi-squared updating can be applied, conditional on the variance hyper-parameters. This implies that the “conditional” acceptance rates do not require actual calculation of the regression parameters on either the old or newly proposed models, as closed forms are available for the integrated likelihood functions conditional on the above variance terms. This leads to significant gains in computational speed. We use the auxiliary variable construction with Kolmogorov-Smirnow prior on the normal variance to obtain the logistic regression model form.

We present the basic structure of the algorithm in form of pseudo code below:

Algorithm 1 : RJ sampler

```

Set  $\theta^{(0)} = ()$ ;
Set  $t = 0$ ;
REPEAT
  Draw  $u_1$  from a  $U(0, 1)$  distribution
  If  $u_1 \leq b_k$  propose a birth step
    if  $u_1 \leq b_k/2$ 
       $\theta' = \text{birth of intensity singleton}$  – proposal;
    else
       $\theta' = \text{birth of intensity-shape pair}$  – proposal;
    end;
  elseif  $b_k \leq u_1 \leq b_k + d_k$  propose a death step
    if  $u_1 \leq b_k + d_k/2$ 
       $\theta' = \text{death of intensity singleton}$  – proposal;
    else
       $\theta' = \text{death of intensity-shape pair}$  – proposal;
    end;
  else propose a change step
    if  $u_1 \leq b_k + d_k + c_k/2$ 
       $\theta' = \text{change of singleton to pair}$  – proposal;
    else
       $\theta' = \text{change of pair to singleton}$  – proposal;
    end;
  End;
  Draw  $u_2$  from a  $U(0, 1)$  distribution
  If  $u_2 < \min \{1, A\}$ 
     $\theta^{(t+1)} = \theta'$ ;
  else
     $\theta^{(t+1)} = \theta^{(t)}$ ;
  End;
   $t = t + 1$ ;
  Store every  $m$ th value of  $\theta^{(t)}$  after initial burn-in;
END REPEAT;

```

5.4 Results

5.4.1 Application

We applied the isotope variable-dimension model to the pancreatic cancer data which contained 2578 intensity and shape measurements in total, coupled in 1289 isotope pairs. For the maximum allowed isotope dimensionality we used $k_{max} = 100$. We set the hyperparameters of the Gamma distribution to $\alpha = \beta = 1$ for the analysis presented in the paper, which results in a prior mean and variance of 1 for both s_a and s_b .

For the prior choice on the scaling factors τ_a and τ_b we make use of the knowledge that the variance-covariance matrix of the regression parameter vector in the logistic regression model is approximately $n * (\mathbf{X}^T \mathbf{X})^{-1}$, where n is the sample size and \mathbf{X} is the data matrix. We find that the inverse cross-products of $\mathbf{u}^T \mathbf{u}$, where $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ and $\mathbf{v}^T \mathbf{v}$, where $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_k)$, are of order 10^{-3} and 10^{-2} which, for a sample size of $n = 254$, points to a prior guess of 0.25 and 2.5 for τ_a and τ_b respectively.

5.4.2 Convergence

To perform the isotope selection, we sampled 2 sets of 500,000 simulations, after an initial burn-in of 200,000 samples which were discarded. Convergence was assessed by comparing the last set of simulations with the first through autocorrelation and kernel density plots on the model parameters such as the isotope dimensionality k , the intensity-shape pair and intensity singleton dimensionalities k_C and k_I , the regression parameters as well as the model variances for intensity and shape over the MCMC chains. For posterior inference on the model we combine the 2 sets of updates into a single set of simulations (1 million samples).

5.4.3 Post-hoc analysis

Previous results on the pancreatic cancer data showed evidence for the presence of predictive information in the isotope clusters which can be used for diagnostic purposes. More specifically, results from using isotope cluster summaries based on either intensity or shape information as input into a ridge logistic model independently, showed that both types of information are predictive of the class outcome, though intensity was found to be more informative than shape. Having established the presence of predictive capacity in the proteomic data, our focus turns towards identifying a collection of isotope clusters associated with the disease as well as investigating the predictive value of shape beyond intensity within the isotope clusters. We begin our exploration by investigating how often an isotope was selected in the model as well as the selection frequency of isotopes consisting of intensity alone and of isotopes consisting of both intensity and shape measures. Figure 5.1 shows the marginal probabilities of inclusion into the model for each of the 1289 isotopes plotted against the isotope cluster number. Crosses correspond to marginal posterior probabilities of inclusion of any specific isotope configuration, irrespective of

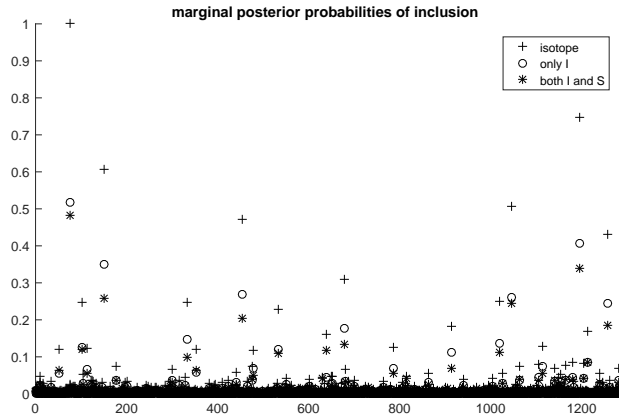


Figure 5.1: Posterior probabilities of inclusion into the model for any specific isotope configuration (crosses), intensity component alone (circles) and both intensity and shape components (asterisks) versus the isotope cluster number.

its composition, circles correspond to marginal posterior probabilities of inclusion of the intensity component alone while asterisks correspond to marginal posterior probabilities of inclusion of both the intensity and shape components such that the value of any crosses equals the sum of the values of the circles and asterisks. As can be seen, there is rather strong evidence in favor of specific isotopes. In particular, isotope 77 is selected into the

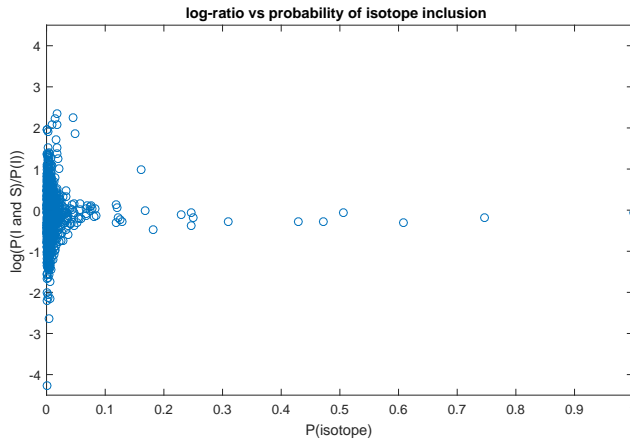


Figure 5.2: Estimates of ratio of inclusion probabilities for both intensity & shape and intensity only (on the log-scale) versus probability of isotope inclusion.

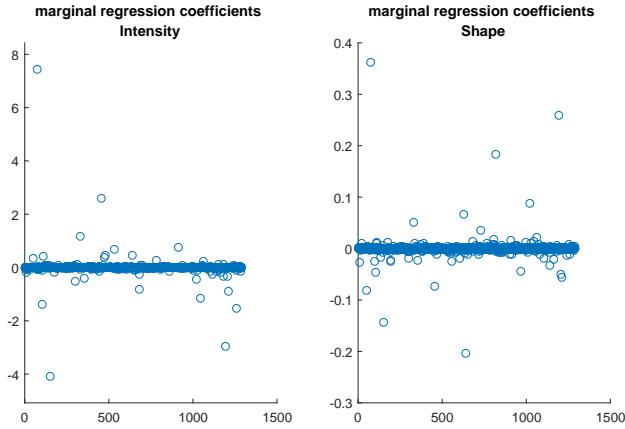


Figure 5.3: Marginal mean regression coefficients of the Bayesian model for intensity (left plot) and shape (right plot) versus the isotope cluster number.

model with a probability of almost one while half of the times the model selects both intensity and shape measures to be included in the set of predictors. We observe a general tendency of the model to select both intensity and shape measures almost as often as intensity alone. This is particularly true for the isotopes with high probability of inclusion.

This tendency of the model is more apparent in Figure 5.2 where we plot the ratio estimates of inclusion probabilities on the log scale, i.e. the ratio between the probability of the entire intensity-shape pair to be selected and the probability of intensity alone to be selected in the model, versus the probabilities of isotope inclusion. We notice that the isotopes for which mutual intensity and shape selection is more frequent than individual intensity selection (log-ratio estimates of inclusion probability above 0) are the ones with the lowest overall (isotope) probability of inclusion. On the other hand, the log-ratio estimates of the isotopes with the highest probabilities of inclusion are close to zero (ratio close to one) which means that for those isotopes the model selects both components with the same frequency that it selects the first component alone. This outcome suggests that in those cases, the model cannot distinguish between including or excluding shape to or from the model. In general, we observe that the log-ratio estimates get closer to zero for increasing values of the overall isotope inclusion probability.

The selection effect is reflected, apart from the selection itself, in the calibration of the regression coefficients. We further investigate the effect of selecting intensity or shape and their relative contribution to classification by calculating the marginal mean of regression coefficients as

$$\sum_{m \in M} \beta_m / M,$$

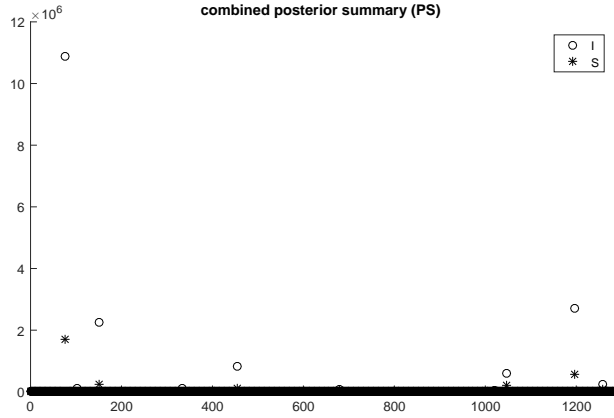


Figure 5.4: Combined posterior summary estimates for intensity (circles) and shape (asterisks) versus the isotope cluster number.

with m denoting the current model within the set of all simulated models and M denoting the total number of simulated models. Figure 5.3 plots these marginal mean regression coefficients separately for intensity (left plot) and shape (right plot) across all simulated models. We can see from these plots that shape is associated with effects of much lower magnitude, as compared to intensity, despite the relatively high frequency with which certain shape measures were selected in the model. This outcome may confirm our intuition that the intensity source carries more information on the class outcome than the shape source, however we must be careful when comparing the shape effects with the intensity effects since the observed differences in effect magnitude could be due to the systematic scale differences between the intensity and shape summary measures.

Essentially, the selection effect is shared between two types of parameters: 1) the inclusion probability and 2) the regression effect. To get additional insights into the relative contribution of intensity and shape we define a “combined” Posterior Summary (PS), which integrates the information from the effect of inclusion and the calibrated regression effect, as

$$PS = \sum_{m \in M} \left(|\beta_m| / \sum_{j=1}^{k'} |\beta_{m_j}| \right) \left(1/k' \right) / M$$

where $\sum_{j=1}^{k'} |\beta_{m_j}|$ is only across either all intensity or all shapes measures, with k' the total number of intensity measurements (equal to $k_I + k_C$) or the total number of shape measurements (equal to k_C) selected in model $m \in M$, depending on whether PS is calculated for an intensity or a shape measure respectively. We choose to compute PS

separately for intensity and shape measures as a way to account for the different scales of the two predictor sources. The combined posterior summary measures for intensity (circles) and shape (asterisks) are shown in Figure 5.4 from which it can be seen that there are in total 6 pairs which stand out as the most important predictor couples while the intensity components of these pairs have higher importance/contribution than their corresponding shape components. In the top plot of Figure 5.5 we plot the log-transformed combined posterior summaries for intensity (circles) and shape (asterisks) against the probability of isotope inclusion to get a more clear picture of the relative contribution of the two measures. Again here, we see 6 circles and 6 asterisks standing out which correspond to the most important intensity and shape measures according to the combined posterior summary while we see that effectively most asterisks representing the shape components lie continually below the circles representing the intensity components, especially for the pairs with large values of probability of isotope inclusion. Similar conclusions can be drawn when looking at the bottom plot of Figure 5.5, which shows the log-transformed combined posterior summaries of intensity versus the log-transformed combined posterior summaries of shape. For the pairs with the most important intensity components (according to their PS estimates), the PS values of the shape components are consistently smaller than those of the intensity components.

5.4.4 Assessment of predictive performance

To assess the predictive performance of the isotope dimension model we use 10-fold internal cross-validation. We do so by first partitioning the data into 10 mutually exclusive and exhaustive, equal sized, sub-samples. We then run 10 parallel chains in each of which

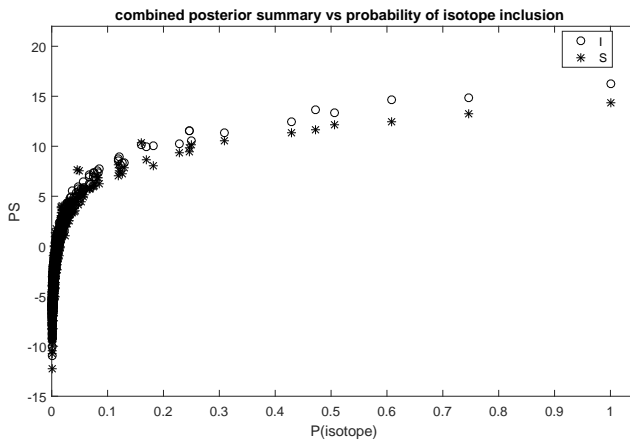


Figure 5.5: Combined posterior summary (on the log scale) of intensity (circles) and shape (asterisk) versus probability of isotope inclusion.

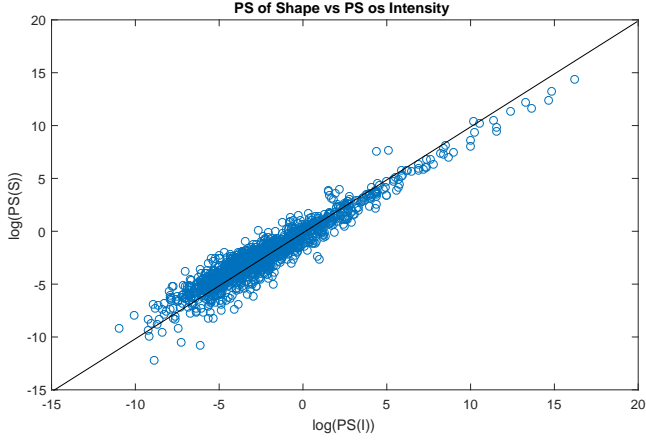


Figure 5.6: Combined posterior summary (on the log scale) of intensity versus combined posterior summary (on the log scale) of shape.

one sub-sample is used as validation data for evaluating the predictive performance of the model, while the remaining 9 sub-samples are used to calibrate the Bayesian model. Since the case-control ratio can greatly affect the selection, to get as consistent variable selection and estimates as possible across the different sets, we partition the original data such that the case-control ratio in the newly defined calibration and validation data is the same across all sub-samples. Evaluation of the model on each sub-sampled validation set is achieved by applying, within each MCMC step in each chain, the generated rule based on the calibration data to the profiles in the validation set and storing the sequence of validated predictions. To assign observations, we calculate for each observation the mean a-posteriori class probabilities of group-membership. That is, we compute the cross-validated probabilities

$$P(y_i = 1 | \tilde{\mathbf{u}}, \tilde{\mathbf{v}}) = \sum_{m \in \mathcal{M}} P_m(y_i = 1 | \tilde{\mathbf{u}}, \tilde{\mathbf{v}}) / M$$

for all $i = 1, \dots, n$, where P_m denotes the a-posteriori class probability calculated from the m th model simulated within the MCMC chain and the sum is across all models simulated. To estimate the error rate we use a cut-off value of 0.5. We assign an observation as a disease case if the mean a-posteriori class probability is greater than 0.5 and as a control otherwise. This assignment resulted in a misclassification error rate of 0.086. We also calculate the Brier score, defined as

$$B = \frac{1}{n} \sum_{i=1}^n (P(y_i = 1 | \tilde{\mathbf{u}}, \tilde{\mathbf{v}}) - c_i)^2$$

which equals 0.082, where c_i denotes the true class outcome of the i th individual. Finally,

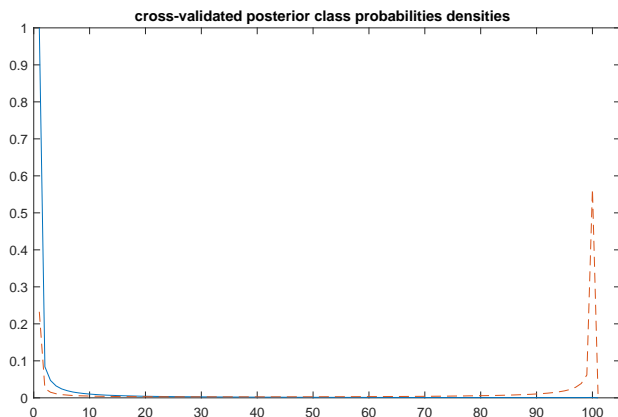


Figure 5.7: Cross-validated posterior class probability densities for controls (solid) and cases (dashed).

the Area Under the Curve (AUC) was found equal to 0.945. In Figure 5.6 we plot the posterior probability densities for the control (solid) and the case (dashed) groups. It is worth mentioning at this point that these results are in tune with previously reported results on the pancreatic cancer data. In particular, internally validated results reported in Chapter 3, obtained using intensity information across all observed isotope clusters into a ridge logistic model, gave an error-rate of 0.095 and an AUC of 0.948.

Finally, we investigate the consistency of the model selection across the partitioned data sets and the extent of agreement with the model selection based on the entire data. Table 5.1 shows the top 7 isotope clusters selected into the Bayesian model applied to the full data (second column) in decreasing order of posterior probability of isotope inclusion. The table also gives the rank of the top selected isotope clusters according to the average posterior probability of isotope inclusion calculated across the 10 sub-sampled data sets (third column) together with these average probability estimates (forth column) and their corresponding standard errors (last column). The top isotope cluster selected by the full model corresponds also to the top isotope cluster selected across sub-models with an average probability of 0.87. Nearly all top isotope clusters selected by the full model are identical to the ones selected across sub-models, also with identical ranking, except for clusters 1047 and 455 which interchange their ranks as well as cluster 639 which drops from the 7th rank to the 12th, with an average (isotope) inclusion probability of just 0.16.

5.4.5 A simulation example

We use a simplified simulation example to demonstrate how the method behaves in a controlled setting. We generate data to have the same number of patients as in the pancreatic cancer data set but smaller dimensionality under the independence assumption. We gen-

Rank (full)	Isotope nr	Rank (average)	P(isotope)	SE
1	77	1	0.87	0.10
2	1196	2	0.50	0.20
3	151	3	0.43	0.13
4	1047	5	0.36	0.15
5	455	4	0.39	0.18
6	1258	6	0.35	0.20
7	639	12	0.16	0.25

Table 5.1: Ranks (left column) of top 7 selected isotope clusters (second column) based on Bayesian model applied to the full data, rank of top selected isotope clusters according to average probability of isotope inclusion across partitions (third column), average probability estimates of isotope inclusion (forth column) and standard errors of average probability estimates (last column).

erate 200 variables in total such that half of them represent the first and the other half the second components of 100 paired measurements. For both components of each pair we simulate data from a normal distribution where we specify the independent normal random variables u and v as $N(0, 2.5^2)$. We simulate the binary outcome data according to the logistic model

$$\begin{aligned} \text{logit}(p) &= \beta_0 + \beta \mathbf{z} \\ &= \beta_0 + \mathbf{a}u + \mathbf{b}v \end{aligned}$$

where \mathbf{a} and \mathbf{b} are 100-dimensional vectors containing the first and second component effects on the class outcome. To induce associations between the predictor components and the outcome, we draw binary response variables from a Bernoulli distribution with

$$p = \frac{1}{1 + e^{-(\beta_0 + \mathbf{a}u + \mathbf{b}v)}}$$

We consider three different scenarios in which we vary the number of true, non-zero effects, as well as their magnitude. In all 3 scenarios we use $k_{max} = 50$, $\alpha = \beta = 1$ for the hyperparameters of the Gama distribution and $\tau_a = \tau_b = 1$ while we vary the values of the non-zero elements of β from 1 to 3.5.

In the first scenario we select solely the first component of the last pair to have a non-zero effect and we set its value to be equal to 3.5. We wish to test whether the model can identify the single predictive pair - and more specifically the single predictive component within the pair, on the one hand, and estimate its true effect, on the other. The top 3 plots of Figure 5.7 show estimates of the inclusion probabilities, marginal effects and combined summaries for the first (circles) and second (asterisks) components of the 100 pairs across 200,000 simulations. The key point from these plots is that the model can identify the single discriminating feature corresponding to the first component of the last pair and it can accurately estimate its true effect.

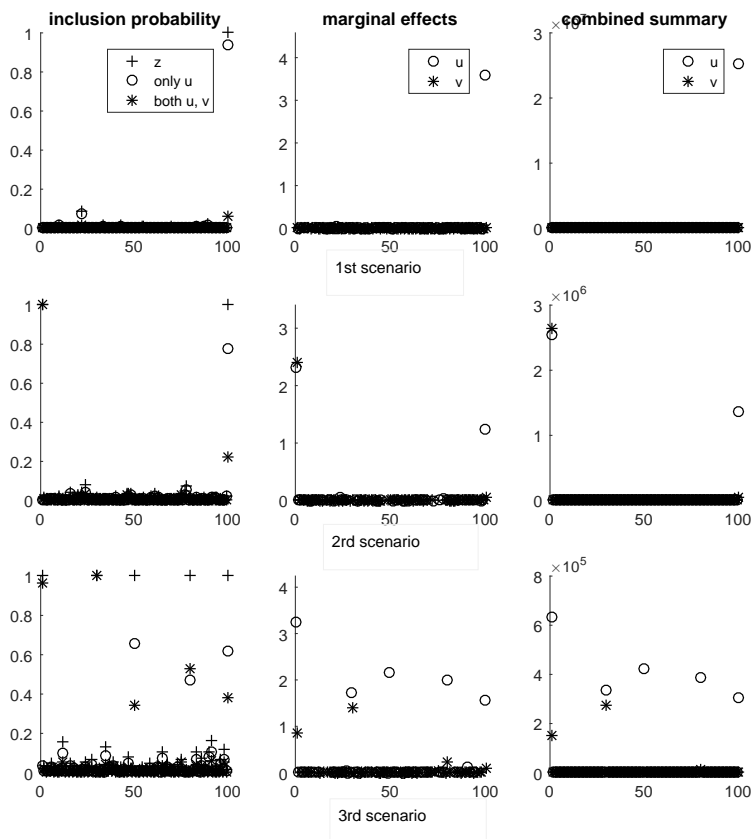


Figure 5.8: Inclusion probabilities (left plots), marginal regression effects (middle plots) and combined posterior summaries (right plots) of the first (circles) and second (asterisks) components of the pairs for the first (top plots), second (middle plots) and third (bottom plots) scenarios.

In the second scenario, apart from the first component of the last pair, we assign non-zero effects to both components of the first pair. We set the effect for the two components of the first pair equal to 2.5 and the effect of the first component of the last pair equal to 1.5. Looking at the first plot in the middle row of Figure 5.7 which shows the inclusion probability estimates, we observe that the second component of the last pair is selected, alongside with the first, with a probability of about 0.2, although this component does not carry any information on the class outcome. The relative importance of this non-predictive component is better captured/reflected by the marginal effect estimates and the estimates of combined summary. We see from the two last plots in the middle row of Figure 5.7

which show these estimates that, although the model selects the non-informative component with a probability 0.2, it assigns it a zero effect, such that the only prominent points in the two plots are the ones which correspond to the 3 components which are predictive of the outcome.

In the third scenario we increase the number of pairs with both components discriminative to two and the number of pairs with only the first component discriminative to three such that $a_k \neq 0$ for $k = 1, 30, 50, 70, 100$ and $b_k \neq 0$ for $k = 1, 30$. We also assume smaller effects for the second components than the first, i.e. we assume that the first components are more informative than the second. In particular, we set $a_1 = 3.5$, $a_{30} = 2$, $a_{50} = a_{70} = 2.5$, $a_{100} = 1.5$ and $b_1 = 1.5$, $b_{30} = 1$. From the first plot at the bottom of Figure 5.7, we see that when both components are predictive, the entire pair is included into the model with a probability almost 1, even though the second components are attributed smaller effects than the first. For the pairs (u_{50}, v_{50}) and (u_{100}, v_{100}) , the probability that the entire pair is selected is overestimated, yet smaller than the probability than only the first component is selected. The same is not true for pair (u_{70}, v_{70}) of which the probability of selecting the entire pair exceeds, even though marginally, the probability of selecting the single component. This overestimation of the inclusion probabilities for the non-informative components is counterbalanced by assigning them a zero effect, as can be seen from the marginal effects plot at the bottom of Figure 5.7. We see from this plot that, in general, the model manages to estimate the true magnitude of the various effects. A more clear picture of the relative importance/contribution of each predictor component is given in the last plot of Figure 5.7 which shows the estimates of combined summaries. The most prominent points in this plot correspond to the informative components of the 5 predictive pairs, which suggests that PS provides an adequate summary for assessment of the true impact of the component variables in the model.

5.5 Discussion

In this work, we addressed the problem of isotope cluster selection through a Bayesian model formulation. Results from applying the Bayesian selection model to the pancreatic cancer data showed rather strong evidence in favor of specific isotope clusters being associated with the class outcome while most often a limited number of about 15 isotope clusters was selected by the model as a sufficient subset for separating the two groups. In addition to the isotope selection, the model formulation allows for assessment of the added-value of the shape source over and above the intensity source through the employment/application of an additional layer of shape selection within the isotope cluster selection. To perform the shape selection we make the explicit assumption that this type of information is complementary to the intensity information which, in the model fitting process, can be formulated as shape being included in the model on the condition that its corresponding intensity is already selected or will be selected to the model alongside with shape. This assumption, and hence this “asymmetrical” manner in which we select/deselect shape, is based on our prior knowledge from previous work that the overall

intensity level is a superior source of predictive information as compared to shape. The main outcome with regards to assessing the added-value of shape was that most shape measures were associated with effects of much smaller magnitude as compared to intensity (conditional on the assumed model), regardless of their inclusion or not in the model, suggesting that intensity alone provides all the required information for separating the two groups.

Variable selection methods have seen many applications within the Bayesian statistics field. Among these, a popular method for performing variable selection is via spike-and-slab priors (George and McCulloch, 1993). Sparsity is achieved in this case by placing a mixture prior on the regression effect of each predictor consisting of a “spike” either exactly at or around zero, corresponding to exclusion of a specific variable from the model, and a flat “slab”, corresponding to inclusion of the variable to the model. Variable selection using spike-and-slab priors is performed by introducing latent binary indicator variables for each predictor to denote whether the predictor belongs to the slab or spike part of the prior. Priors are placed on the binary indicator variables to encourage sparsity. MCMC sampling is often used to fit this type of models in which fixed prior parameters are frequency specified in order to reduce the computational burden. A variant of the slab-and-spike priors was considered by Dellaportas et al. (2000, 2002) who proposed including the binary indicator variables γ_j in the likelihood so that the variables which do

not contribute to the linear predictor, which is now of the form $\beta_0 + \sum_{j=1}^k \gamma_j \beta_j X_j$, drop off.

In our Bayesian selection approach we used a reversible-jump implementation in which the level of sparsity is controlled through a prior on the model (isotope) dimension. This allows us, on the one hand, to estimate the optimal dimensionality of the isotope clusters predictor set, and on the other, to assess the additional value of shape at the isotope cluster level, through the Metropolis-Hastings acceptance ratios, which account for inclusion or exclusion of shape to or from the model.

In a recent paper, Rodríguez-Girondo et al. (2016) proposed a frequentist two-step approach for assessing the augmented predictive value of a secondary source on top of a primary source based on a sequential double cross-validation procedure. Apart from the selective nature of our Bayesian model, a key difference between this sequential approach and our Bayesian selection method is that the latter takes explicitly into account the pairing structure of our data and could be extended to deal also with more complex grouping data structures such as triplets, quadruples etc. Moreover, in contrast to the sequential approach, our Bayesian approach can be used not solely to evaluate whether there is additional predictive information in the secondary source (shape) after correcting for the primary (intensity) source but also to address the question of where this extra information, if any, comes from. This is essentially achieved through the “asymmetrical” layer of shape selection within the isotope cluster selection which evaluates - and estimates - for each unique isotope cluster the relative contribution of each shape measure to classification.

An alternative way of implementing the double-layered selection of isotope clusters,

on the one hand, and shape measures within the isotope cluster configuration, on the other, would be to restrict the parameter space by modifying the prior specification on the regression coefficient parameters. More specifically, rather than assuming Normal priors for the intensity and shape effects we may constrain the regression coefficients to be within the set of values $\{-1, 0, 1\}$. In this way, we replace calibration of the regression coefficients with selection by restricting the effect to be absent, present-negative or present-positive. Moreover, by assuming pre-specified values for the magnitude of the regression effects, no further assumptions, for instance on the distribution of the variances for the intensity and shape effects, are needed. A limitation of this particular prior specification is that simple conjugate updating rules are no longer applicable. This suggests that calculation of the acceptance probabilities would require actual estimation of the integrated likelihood functions of the old and newly proposed models which could increase the computational time considerably.

We restricted our discussion to intensity-shape combinations in which we considered simple linear effects for both intensity and shape summary measures, mainly due to the fact that such linear rules can be calibrated and interpreted with relative ease in practical application. In reality however the relationship between the class outcome and the two predictor sources - as well as the relationship between the intensity source and the shape source - may be more complex. In fact, it would be interesting to consider more complex effect structures, for example by including also quadratic terms for intensity and shape as well as interactions between intensity and shape into the model in addition to the linear effects. Allowing for this flexibility in the model could potentially improve the predictive capacity of intensity and shape integration and give us the chance to learn more about the interplay between these two different types of information. The idea of considering more complex structures in order to capture the true relationships in the data, in the hope that this will result in improved predictions and/or more thorough inference, is a promising topic for future research. Preliminary results from an exploratory analysis showed that including quadratic as well as interaction terms in a univariate model (i.e. a model with only one intensity predictor and one shape predictor fitted separately for each isotope cluster), could lead to improvement of the model fit, as compared to univariate models with only linear terms, for a considerable number of clusters.

An undesired feature of the proposed Bayesian selection approach is that it tends to overselect shape components even in the cases where there is no true effect. This tendency of the model is more pronounced for isotope clusters which are selected with high overall isotope probability of inclusion. The overestimation of the inclusion probabilities for the non-informative shape components is partly counterbalanced by the calibration of the marginal regression coefficients for shape which were effectively zero for almost all shape measures that were included to the model with high frequency. Overestimation of inclusion probabilities is actually a known deficiency of reversible-jump estimation procedures for (linear) models, yet not discussed/identified in the related literature.

For posterior inference, we proposed a posterior summary which combines the information that can be extracted from the inclusion probability estimates and the marginal

regression effect estimates as a way to deal with the “inflated” shape selection. Results from both the real data analysis and a simulation study showed that the “combined” posterior summary gives a reasonable assessment of the relative contribution of shape in separating the two groups when combined with intensity. A more formal solution to this problem would be to modify the assumptions on the distribution of the variance for the shape effects (or the previously mentioned -1,0,1 based model proposal which decouples effect estimation from variable selection). For example, instead of an Inverse Gamma distribution for the shape variance we could assume a Uniform distribution in order to shift the values of the variance distribution to larger values. This would result in more conservative shape selection as, for higher values of σ_b , the acceptance probability for adding a shape measure becomes smaller while the acceptance probability for removing a shape measure becomes larger. The drawback however with this solution is that conjugate updating is not applicable for this family of distributions. An other alternative could be to use a more general and flexible prior specification for the variance of the intensity and shape effects so that the model can adapt to the data at hand. A solution towards this direction could be to assume independent - across isotopes - Inverse Gamma priors with different hyperparameters for the intensity and shape variances with additional priors placed on these hyperparameters. This prior specification could give us more flexibility and control in borrowing strength across the shape measures as compared to setting the hyperparameters of the Inverse Gamma priors to fixed values.

Appendix

A. Calculation of acceptance probabilities

Let T be the total number of isotopes in our data set and $k = 0, 1, \dots, k_{max}$, the model (isotope) dimension. Within each model dimension k , isotopes may be included as either singletons (intensity) or couples (intensity & shape) such that the isotope dimension k equals the number of intensity singletons k_I plus the number of intensity-shape couples k_C . We use the following 6 move types:

- 1) birth - intensity (B_I),
- 2) birth - couple (B_C),
- 3) death - intensity (D_I),
- 4) death - couple (D_C),
- 5) change - add shape (C_{AS}),
- 6) change - remove shape (C_{RS}).

We assume that within each isotope dimension, all models are equally likely. We choose the proposal probabilities as

$$\begin{aligned}
 b_{(k=0)}^I &= b_{(k=0)}^C = d_{(k=k_{max}=k_I)}^I = d_{(k=k_{max}=k_C)}^C = c_{(k=k_{max}=k_I)}^{I \rightarrow C} = c_{(k=k_{max}=k_C)}^{C \rightarrow I} = 1/2, \\
 d_{(0 < k_I, k_C < k_{max}=k)}^I &= d_{(0 < k_I, k_C < k_{max}=k)}^C = c_{(0 < k_I, k_C < k_{max}=k)}^{C \rightarrow I} = c_{(0 < k_I, k_C < k_{max}=k)}^{I \rightarrow C} = 1/4, \\
 b_{(k_I=0 < k_C < k_{max})}^I &= b_{(k_I=0 < k_C < k_{max})}^C = d_{(k_I=0 < k_C < k_{max})}^C = c_{(k_I=0 < k_C < k_{max})}^{C \rightarrow I} = 1/4, \\
 b_{(k_C=0 < k_I < k_{max})}^I &= b_{(k_C=0 < k_I < k_{max})}^C = d_{(k_C=0 < k_I < k_{max})}^I = c_{(k_C=0 < k_I < k_{max})}^{I \rightarrow C} = 1/4, \\
 b_{(k=k_{max})}^I &= b_{(k=k_{max})}^C = d_{(k_I=0)}^I = d_{(k_C=0)}^C = c_{(k_C=0)}^{C \rightarrow I} = c_{(k_I=0)}^{I \rightarrow C} = 0
 \end{aligned}$$

and

$$b_k^I = b_k^C = d_k^I = d_k^C = c_k^{C \rightarrow I} = c_k^{I \rightarrow C} = 1/6$$

in all other cases.

The acceptance probability for a proposal move from a model with parameters θ to a new model with parameters θ' is given by

$$a = \min \left\{ 1, \frac{P(D|\theta') p(\theta') q(\theta|\theta')}{P(D|\theta) p(\theta) q(\theta'|\theta)} \right\},$$

the ratio of 1) marginal likelihoods, 2) priors and 3) proposal distributions.

For the prior distribution of θ we use a discrete uniform specification of the form

$$p(\theta) = \binom{T}{k}^{-1} \binom{k}{k'}^{-1} \frac{1}{k_{max} + 1}$$

where k' is either one of k_I or k_C . The first two terms in the equation ensure that each model with isotope dimension k and intensity singleton dimension k_I or intensity-shape couple dimension k_C is equally likely. In this way we assume that, given k , any set of potential isotope predictors is found by sampling k isotope clusters from the candidate set \mathcal{T} without replacement. The last term assumes that each possible isotope dimension $k \in \{0, 1, \dots, k_{max}\}$ is equally likely.

Before we show how to calculate the prior and proposal ratios for all 6 move types based on the description in Denison et al. (2002), we explain the notion of the proposal ratio. This involves understanding both the BIRTH and the reverse DEATH moves. Suppose we add a new intensity singleton component, then we use the proposal density $q(\theta'|\theta) = b_{I(k)}/(T - k)$. This consists of the probability of attempting this particular birth move and the probability of choosing this particular new component which can be done in $T - k$ ways. The probability of proposing the reverse move is $q(\theta'|\theta) = d_{I(k+1)}/(k_I + 1)$ which is the probability of proposing the death of an intensity singleton and then of choosing the proposed component as the one to remove. The ratio of marginal likelihoods BF can be calculated according to the equations given in Denison et al. (2002).

BIRTH MOVE

Birth of an I singleton

$$k \rightarrow k + 1, \quad k_I \rightarrow k_I + 1, \quad k_C \rightarrow k_C$$

$$\text{prior ratio} = \frac{p(\theta')}{p(\theta)} = \frac{\frac{1}{\binom{T}{k+1}}}{\frac{1}{\binom{T}{k}}} \frac{\frac{1}{\binom{k+1}{k_I+1}}}{\frac{1}{\binom{k}{k_I}}} \frac{1}{k_{max} + 1} = \frac{k_I + 1}{T - k}$$

$$\text{proposal ratio} = \frac{q(\theta|\theta')}{q(\theta'|\theta)} = \frac{\frac{d_{I(k+1)}}{k_I + 1}}{\frac{b_{I(k+1)}}{T - k}} = \frac{T - k}{k_I + 1} \frac{d_{I(k+1)}}{b_{I(k)}}$$

$$R = \frac{p(\theta')}{p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)} = \frac{d_{I(k+1)}}{b_{I(k)}}$$

$$\text{If } k = 0 \rightarrow \frac{d_{I(k+1)}}{b_{I(k)}} = \frac{1}{6}, \text{ else } \frac{d_{I(k+1)}}{b_{I(k)}} = 1$$

Birth of an IS pair

$$k \rightarrow k + 1, \quad k_I \rightarrow k_I, \quad k_C \rightarrow k_C + 1$$

$$\text{prior ratio} = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} = \frac{\frac{1}{\binom{T}{k+1}}}{\frac{1}{\binom{T}{k}}} \frac{\frac{1}{\binom{k+1}{k_C+1}}}{\frac{1}{\binom{k}{k_C}}} \frac{1}{k_{max}+1} = \frac{k_C+1}{T-k}$$

$$\text{proposal ratio} = \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{\frac{d_{C(k+1)}}{k_C+1}}{\frac{b_{C(k+1)}}{T-k}} = \frac{T-k}{k_C+1} \frac{d_{C(k+1)}}{b_{C(k)}}$$

$$R = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{d_{C(k+1)}}{b_{C(k)}}$$

$$\text{If } k = 0 \rightarrow \frac{d_{C(k+1)}}{b_{C(k)}} = \frac{1}{6}, \text{ else } \frac{d_{C(k+1)}}{b_{C(k)}} = 1$$

DEATH MOVE**Death of an I singleton**

$$k \rightarrow k - 1, \quad k_I \rightarrow k_I - 1, \quad k_C \rightarrow k_C$$

$$\text{prior ratio} = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} = \frac{\frac{1}{\binom{T}{k-1}}}{\frac{1}{\binom{T}{k}}} \frac{\frac{1}{\binom{k-1}{k_I-1}}}{\frac{1}{\binom{k}{k_I}}} \frac{1}{k_{max}+1} = \frac{T-k+1}{k_I}$$

$$\text{proposal ratio} = \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{\frac{b_{I(k-1)}}{T-k+1}}{\frac{d_{I(k)}}{k_I}} = \frac{k_I}{T-k+1} \frac{b_{I(k-1)}}{d_{I(k)}}$$

$$R = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{b_{I(k-1)}}{d_{I(k)}}$$

$$\text{If } k = k_{max} \rightarrow \frac{b_{I(k-1)}}{d_{I(k)}} = \frac{1}{6}, \text{ else } \frac{b_{I(k-1)}}{d_{I(k)}} = 1$$

Death of an IS pair

$$k \rightarrow k - 1, \quad k_I \rightarrow k_I, \quad k_C \rightarrow k_C - 1$$

$$\text{prior ratio} = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} = \frac{\frac{1}{\binom{T}{k-1}} \frac{1}{\binom{k_C-1}{k_C-1}} \frac{1}{k_{max}+1}}{\frac{1}{\binom{T}{k}} \frac{1}{\binom{k}{k_C}} \frac{1}{k_{max}+1}} = \frac{T-k+1}{k_C}$$

$$\text{proposal ratio} = \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{\frac{b_{C(k-1)}}{T-k+1}}{\frac{d_{C(k)}}{k_C}} = \frac{k_C}{T-k+1} \frac{b_{C(k-1)}}{d_{C(k)}}$$

$$R = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{b_{C(k-1)}}{d_{C(k)}}$$

$$\text{If } k = k_{max} \rightarrow \frac{b_{C(k-1)}}{d_{C(k)}} = \frac{1}{6}; \quad \text{else } \frac{b_{C(k-1)}}{d_{C(k)}} = 1$$

CHANGE MOVE**Change of an I singleton to an IS pair**

$$k \rightarrow k, \quad k_I \rightarrow k_I - 1, \quad k_C \rightarrow k_C + 1$$

$$\text{prior ratio} = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} = \frac{\frac{1}{\binom{T}{k}} \frac{1}{k_{max}+1}}{\frac{1}{\binom{T}{k}} \frac{1}{k_{max}+1}} = 1$$

$$\text{proposal ratio} = \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{\frac{c_{RS(k)}}{k_C+1}}{\frac{c_{AS(k)}}{k_I}} = \frac{k_I}{k_C+1} \frac{c_{RS(k)}}{c_{AS(k)}}$$

$$R = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{k_I}{k_C+1} \frac{c_{RS(k)}}{c_{AS(k)}}$$

$$\text{If } k = k_I \rightarrow \frac{c_{RS(k)}}{c_{AS(k)}} = \frac{1}{6}, \text{ else } \frac{c_{RS(k)}}{c_{AS(k)}} = 1$$

Change of an IS pair to an I singleton

$$k \rightarrow k, k_I \rightarrow k_I + 1, k_C \rightarrow k_C - 1$$

$$\text{prior ratio} = \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} = \frac{\frac{1}{\binom{T}{k}}}{\frac{1}{\binom{T}{k}} \frac{1}{k_{max} + 1}} = 1$$

$$\text{proposal ratio} = \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{\frac{c_{AS(k)}}{k_I + 1}}{\frac{c_{RS(k)}}{k_C}} = \frac{k_C}{k_I + 1} \frac{c_{AS(k)}}{c_{RS(k)}}$$

$$R = \frac{p(\boldsymbol{\theta}') q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{p(\boldsymbol{\theta}) q(\boldsymbol{\theta}'|\boldsymbol{\theta})} = \frac{k_C}{k_I + 1} \frac{c_{AS(k)}}{c_{RS(k)}}$$

$$\text{If } k = k_C \rightarrow \frac{c_{AS(k)}}{c_{RS(k)}} = \frac{1}{6}, \text{ else } \frac{c_{AS(k)}}{c_{RS(k)}} = 1$$

