Model-based Noise PSD Estimation from Speech in Non-stationary Noise

Nielsen, Jesper Kjær; Kavalekalam, Mathew Shaji; Christensen, Mads Græsbøll; Boldt, Jesper Bünsow

Published in:
IEEE International Conference on Acoustics, Speech, and Signal Processing

DOI (link to publication from Publisher):
10.1109/ICASSP.2018.8461683

Publication date:
2018

Document Version
Other version

Citation for published version (APA):
Nielsen, J. K., Kavalekalam, M. S., Christensen, M. G., & Boldt, J. B. (2018). Model-based Noise PSD Estimation from Speech in Non-stationary Noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 5424-5428). Article 8461683 IEEE (Institute of Electrical and Electronics Engineers). https://doi.org/10.1109/ICASSP.2018.8461683

Variational Bayesian Inference for Model-based Noise PSD Estimation

Jesper Kjær Nielsen[1], Mathew Shaji Kavalekam[1], Mads Græsbøll Christensen[1], and Jesper Boldt[2]
[1] Audio Analysis Lab, CREATE, Aalborg University
[2] GN ReSound

Last compiled: April 17, 2018

1 Introduction

This document gives the details of the variational Bayesian (VB) noise power spectral density (PSD) estimator proposed in our ICASSP-paper [2]. In the first section of the document, a general description of the VB approach is given to joint parameter estimation and model comparison. If the reader is familiar with the VB approach, this section can be skipped. In the second section, the VB algorithm is applied to the model described in [2]. We strongly encourage the reader to read the ICASSP paper before continuing reading this document.

2 A VB tutorial

Suppose we wish to compute the posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{k=1}^{K} p(\boldsymbol{\theta}|\boldsymbol{x}, \mathcal{M}_k) p(\mathcal{M}_k|\boldsymbol{x}) \ , \tag{1}$$

for the model parameters $\boldsymbol{\theta}$, but cannot do this analytically. Sometimes, we can simplify the problem considerably if we introduce latent variables $\boldsymbol{s}$. An example of this is for Gaussian mixture models where the introduction of latent variables allows us to use the EM-algorithm to produce a solution. If we let $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{s}^T & \boldsymbol{\theta}^T \end{bmatrix}^T$, the joint posterior over the latent parameters and the model parameters is

$$p(\boldsymbol{z}|\boldsymbol{x}) = \sum_{k=1}^{K} p(\boldsymbol{z}|\boldsymbol{x}, \mathcal{M}_k) p(\mathcal{M}_k|\boldsymbol{x}) \ . \tag{2}$$

Often, this posterior cannot be computed analytically, so we assume that it factorises for each model $\mathcal{M}_k$ as

$$p(\boldsymbol{z}|\boldsymbol{x}, \mathcal{M}_k) \approx q_k(\boldsymbol{z}) = \prod_{i=1}^{M} q_{ik}(\boldsymbol{z}_i) \tag{3}$$

where we have divided $\boldsymbol{z}$ into M non-overlapping subgroups. In the sequel, we will use the more compact notation $q_{ik} = q_{ik}(\boldsymbol{z}_i)$. Also note that even though we do not assume any factorisation for $p(\mathcal{M}_k|\boldsymbol{x})$, we will only obtain an approximation $q(\mathcal{M}_k)$ of it since we use $q_k(\boldsymbol{z})$ instead of $p(\boldsymbol{z}|\boldsymbol{x}, \mathcal{M}_k)$. Thus, we have that

$$p(\boldsymbol{z}, \mathcal{M}_k|\boldsymbol{x}) \approx q(\boldsymbol{z}, \mathcal{M}_k) = q(\mathcal{M}_k)q_k = q(\mathcal{M}_k)\prod_{i=1}^{M} q_{ik} \ . \tag{4}$$

Thus, we have now reformulated the problem from finding $p(\boldsymbol{z}, \mathcal{M}_k|\boldsymbol{x})$ to that of finding the different factors q_{ik} and $q(\mathcal{M}_k)$. If we select the factorisation in a clever way, it is much easier to find these factors (using VB) than it is to find the exact posterior $p(\boldsymbol{z}, \mathcal{M}_k|\boldsymbol{x})$. The art of applying the VB framework is, therefore, to elicit a signal model (i.e., an observation model $p(\boldsymbol{x}|\boldsymbol{z}, \mathcal{M}_k)$ and prior distributions $p(\boldsymbol{z}|\mathcal{M}_k)$ and $p(\mathcal{M}_k)$) and a factorisation so that the various factors can easily be computed while still being a good approximation to the exact posterior. In the sequel, we assume that we have elicited these and focus on describing how the various factors of the elicited factorisation are computed from the signal model.

For any pdf q, we have that the log of the model evidence can be written as [1, p. 473]

$$\ln p(\boldsymbol{x}) = \mathcal{L}(q) - \sum_{k=1}^{K} q(\mathcal{M}_k) \int q_k \ln \frac{p(\boldsymbol{z}, \mathcal{M}_k|\boldsymbol{x})}{q_k q(\mathcal{M}_k)} d\boldsymbol{z} \tag{5}$$

where

$$\mathcal{L}(q) = \sum_{k=1}^{K} q(\mathcal{M}_k) \int q_k \ln \frac{p(\boldsymbol{z}, \boldsymbol{x}, \mathcal{M}_k)}{q_k q(\mathcal{M}_k)} d\boldsymbol{z} \ . \tag{6}$$

It turns out that maximising the lower bound $\mathcal{L}(q)$ w.r.t. q produces the best approximation to the posterior $p(\boldsymbol{z}, \mathcal{M}_k|\boldsymbol{x})$. To perform this maximisation, the lower bound is first rewritten as

$$\mathcal{L}(q) = \sum_{k=1}^{K} q(\mathcal{M}_k)\mathcal{L}_k(q_k) + \sum_{k=1}^{K} q(\mathcal{M}_k) \ln \frac{p(\mathcal{M}_k)}{q(\mathcal{M}_k)} \tag{7}$$

where

$$\mathcal{L}_k(q_k) = \int q_k \ln \frac{p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)}{q_k} d\boldsymbol{z} \tag{8}$$

$$= \int q_k \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) d\boldsymbol{z} - \int q_k \ln q_k d\boldsymbol{z} \tag{9}$$

$$= \int q_{jk} \left[\int \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) \left(\prod_{i \neq j} q_{ik} d\boldsymbol{z}_i \right) \right] d\boldsymbol{z}_j - \int \sum_{i=1}^{M} \ln q_{ik} \left(\prod_{i=1}^{K} q_{ik} d\boldsymbol{z}_i \right) \tag{10}$$

$$= \int q_{jk} \left[\int \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) \left(\prod_{i \neq j} q_{ik} d\boldsymbol{z}_i \right) \right] d\boldsymbol{z}_j - \sum_{i=1}^{M} \int q_{ik} \ln q_{ik} d\boldsymbol{z}_i \tag{11}$$

$$= \int q_{jk} \left[\int \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) \left(\prod_{i \neq j} q_{ik} d\boldsymbol{z}_i \right) - \ln q_{jk} \right] d\boldsymbol{z}_j - \sum_{i \neq j} \int q_{ik} \ln q_{ik} d\boldsymbol{z}_i \ . \tag{12}$$

Given the data $\boldsymbol{x}$ and the pdfs q_{ik} for $i \neq j$, the integral in the bracket is a function of $\boldsymbol{z}_j$. However, this function does not necessarily integrate to one, but such a function can easily be defined as

$$\ln \tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k) = \mathbb{E}_{i \neq j}[\ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)] - \ln Z_{jk} \tag{13}$$

where $\mathbb{E}_{i \neq j}[\ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)]$ and Z_{jk} are the expectation operator w.r.t. to $\prod_{i \neq j} q_{ik}$ and the normalisation constant so that $\tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k)$ integrates to one, respectively. These are given by

$$\mathbb{E}_{i \neq j}[\ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)] = \int \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) \left(\prod_{i \neq j} q_{ik} d\boldsymbol{z}_i \right) \tag{14}$$

$$Z_{jk} = \int \exp \left\{ \mathbb{E}_{i \neq j}[\ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)] \right\} d\boldsymbol{z}_j \ . \tag{15}$$

Consequently, we can now write $\mathcal{L}_k(q_k)$ as

$$\mathcal{L}_k(q_k) = \int q_{jk} \left[\ln \tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k) + \ln Z_{jk} - \ln q_{jk} \right] d\boldsymbol{z}_j - \sum_{i \neq j} \int q_{ik} \ln q_{ik} d\boldsymbol{z}_i \tag{16}$$

$$= -\mathrm{KL}(q_{jk}||\tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k)) + \ln Z_{jk} + \sum_{i \neq j} H[q_{ik}] \tag{17}$$

where $\mathrm{KL}(q_{jk}||\tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k))$ and $H[q_{ik}]$ are the Kullback-Leibler (KL) divergence and the entropy, respectively.

We have now written the lower bound $\mathcal{L}(q)$ in a form which allows us to maximise it w.r.t. to q_{jk}. Since the KL divergence is the only term which depends on the functional q_{jk}, we maximise the lower bound by minimising the KL divergence. Fortunately, this is easy since the minimum of the KL divergence is zero, and this minimum value is obtained if and only if

$$q_{jk} = \tilde{p}(\boldsymbol{z}_j|\boldsymbol{x}, \boldsymbol{M}_k) \tag{18}$$

and the solution to the optimisation problem, therefore, is that

$$\ln q_{jk} = \mathbb{E}_{i\neq j}[\ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k)] + \text{const.} \tag{19}$$

The above solution is not a closed-form solution since the expectation operator depends on $\{q_{ik}\}_{i\neq j}$ which are also unknown. However, if we optimise for the individual q_{jk}'s iteratively, we are guaranteed to converge to a solution [1, p. 466].

2.1 The VB Lower Bound

For every model order, we can keep an eye on the convergence by monitoring the lower bound $\mathcal{L}_k(q_k)$. For the optimal form of q_{jk}, the KL divergence vanishes, and the lower bound is given by

$$\mathcal{L}_k(q_k) = \ln Z_{jk} + \sum_{i\neq j} H[q_{ik}]. \tag{20}$$

Since this should hold for all j's, we can stop the algorithm when $\mathcal{L}_k(q_k)$ is nearly the same for all j's. An alternative formulation for the lower bound can also be constructed directly from the definition of $\mathcal{L}_k(q_k)$ as

$$\begin{aligned} \mathcal{L}_k(q_k) &= \int q_k \ln p(\boldsymbol{z}, \boldsymbol{x}|\mathcal{M}_k) d\boldsymbol{z} - \int q_k \ln q_k d\boldsymbol{z} & (21)\\ &= \int q_k \ln p(\boldsymbol{x}|\boldsymbol{z}, \mathcal{M}_k) d\boldsymbol{z} + \int q_k \ln p(\boldsymbol{z}|\mathcal{M}_k) d\boldsymbol{z} + \sum_{i=1}^{M} H[q_{ik}] & (22)\\ &= \mathbb{E}_{q_k}[\ln p(\boldsymbol{x}|\boldsymbol{z}, \mathcal{M}_k)] + \mathbb{E}_{q_k}[\ln p(\boldsymbol{z}|\mathcal{M}_k)] + \sum_{i=1}^{M} H[q_{ik}] \ . & (23) \end{aligned}$$

2.2 The VB Model Comparison

We can also do model comparison in the VB framework. To do this, we rewrite $\mathcal{L}(q)$ as

$$\begin{aligned} \mathcal{L}(q) &= \sum_{k=1}^{K} q(\mathcal{M}_k)\mathcal{L}_k(q_k) + \sum_{k=1}^{K} q(\mathcal{M}_k) \ln \frac{p(\mathcal{M}_k)}{q(\mathcal{M}_k)} & (24)\\ &= \sum_{k=1}^{K} q(\mathcal{M}_k) \left[\ln\left(p(\mathcal{M}_k) \exp(\mathcal{L}_k(q_k))\right) - \ln q(\mathcal{M}_k)\right] \ . & (25) \end{aligned}$$

Given the data $\boldsymbol{x}$ and the joint pdf q_k, the first term in the bracket is a function of the model $\mathcal{M}_k$. This function, however, does not necessarily integrate to one, but such a function can easily be defined as

$$\ln \tilde{p}(\mathcal{M}_k|\boldsymbol{x}) = \mathcal{L}_k(q_k) + \ln p(\mathcal{M}_k) - \ln Z_k \tag{26}$$

where Z_k is a normalisation constant given by

$$Z_k = \sum_{k=1}^{K} p(\mathcal{M}_k) \exp\left[\mathcal{L}_k(q_k)\right] \ . \tag{27}$$

Using these definitions, we can write $\mathcal{L}(q)$ as

$$\mathcal{L}(q) = -\text{KL}(q(\mathcal{M}_k)||\tilde{p}(\mathcal{M}_k|\boldsymbol{x})) + \ln Z_k \tag{28}$$

which is clearly maximised for

$$q(\mathcal{M}_k) = \tilde{p}(\mathcal{M}_k|\boldsymbol{x}) \tag{29}$$

or, equivalently,

$$\ln q(\mathcal{M}_k) = \mathcal{L}_k(q_k) + \ln p(\mathcal{M}_k) + \text{const.} \tag{30}$$

$$= \ln Z_{jk} + \sum_{i \neq j} H[q_{ik}] + \ln p(\mathcal{M}_k) + \text{const.} \tag{31}$$

for any j.

3 Noise PSD Estimation Using VB

We will now apply the VB framework to solve the problem of computing the posterior distribution on the excitaiton noise variances from a mixture of two periodic AR processes. That is, we observe

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{e} \tag{32}$$

where

$$p(\boldsymbol{e}|\sigma_\text{e}^2, \mathcal{M}_k) = \mathcal{N}(\boldsymbol{0}, \sigma_\text{e}^2 \boldsymbol{R}_\text{e}(\boldsymbol{b}_k)) \tag{33}$$

$$p(\boldsymbol{s}|\sigma_\text{s}^2, \mathcal{M}_k) = \mathcal{N}(\boldsymbol{0}, \sigma_\text{s}^2 \boldsymbol{R}_\text{s}(\boldsymbol{a}_k)) \tag{34}$$

$$\boldsymbol{R}_\text{s}(\boldsymbol{a}_k) = N^{-1}\boldsymbol{F}\boldsymbol{D}_\text{s}(\boldsymbol{a}_k)\boldsymbol{F}^H \tag{35}$$

$$\boldsymbol{R}_\text{e}(\boldsymbol{b}_k) = N^{-1}\boldsymbol{F}\boldsymbol{D}_\text{e}(\boldsymbol{b}_k)\boldsymbol{F}^H \tag{36}$$

$$[\boldsymbol{F}]_{nl} = \exp(j2\pi(n-1)(l-1)/N) \,, \qquad n, l = 1, \ldots, N \tag{37}$$

$$\boldsymbol{\Lambda}_\text{s}(\boldsymbol{a}_k) = \text{diag}(\boldsymbol{F}^H \begin{bmatrix} \boldsymbol{a}_k^T & \boldsymbol{0} \end{bmatrix}^T) \tag{38}$$

$$\boldsymbol{\Lambda}_\text{e}(\boldsymbol{b}_k) = \text{diag}(\boldsymbol{F}^H \begin{bmatrix} \boldsymbol{b}_k^T & \boldsymbol{0} \end{bmatrix}^T) \tag{39}$$

$$\boldsymbol{D}_\text{s}(\boldsymbol{a}_k) = \left(\boldsymbol{\Lambda}_\text{s}^H(\boldsymbol{a}_k)\boldsymbol{\Lambda}_\text{s}(\boldsymbol{a}_k)\right)^{-1} \tag{40}$$

$$\boldsymbol{D}_\text{e}(\boldsymbol{b}_k) = \left(\boldsymbol{\Lambda}_\text{e}^H(\boldsymbol{b}_k)\boldsymbol{\Lambda}_\text{e}(\boldsymbol{b}_k)\right)^{-1} \,. \tag{41}$$

Moreover, we assume that the prior for the excitation noise variances are given by

$$p(\sigma_\text{e}^2|\mathcal{M}_k) = \text{Inv-}\mathcal{G}(\alpha_{\text{e},k}, \beta_{\text{e},k}) \tag{42}$$

$$p(\sigma_\text{s}^2|\mathcal{M}_k) = \text{Inv-}\mathcal{G}(\alpha_{\text{s},k}, \beta_{\text{s},k}) \tag{43}$$

whose hyperparameters we initially assume known. The latens variable of our model is $\boldsymbol{s}$ whereas the model parameters are $\sigma_{\rm s}^2$ and $\sigma_{\rm e}^2$. Note that the AR parameters are given by the model. From the above model, it also follows that

$$p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k) = \mathcal{N}(\boldsymbol{s},\sigma_{\rm e}^2\boldsymbol{R}_{\rm e}(\boldsymbol{b}_k)) \tag{44}$$

so that joint distribution over the observations, the latent variables, and the model parameters factorise as

$$p(\boldsymbol{y},\boldsymbol{s},\sigma_{\rm s}^2,\sigma_{\rm e}^2|\mathcal{M}_k) = p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k)p(\boldsymbol{s}|\sigma_{\rm s}^2,\mathcal{M}_k)p(\sigma_{\rm s}^2|\mathcal{M}_k)p(\sigma_{\rm e}^2|\mathcal{M}_k)\ . \tag{45}$$

In our VB algorithm, we seek to compute and approximation of the joint posterior $p(\boldsymbol{s},\sigma_{\rm s}^2,\sigma_{\rm e}^2|\mathcal{M}_k)$ using the factorisation

$$p(\boldsymbol{s},\sigma_{\rm s}^2,\sigma_{\rm e}^2|\mathcal{M}_k) \approx q_k(\boldsymbol{s},\sigma_{\rm s}^2,\sigma_{\rm e}^2) = q_{1k}(\boldsymbol{s})q_{2k}(\sigma_{\rm s}^2)q_{3k}(\sigma_{\rm e}^2) \tag{46}$$

where the factors can be viewed as approximations to the marginal posteriors with

$$q_{1k}(\boldsymbol{s}) \approx p(\boldsymbol{s}|\boldsymbol{y},\mathcal{M}_k) \tag{47}$$
$$q_{2k}(\sigma_{\rm s}^2) \approx p(\sigma_{\rm s}^2|\boldsymbol{y},\mathcal{M}_k) \tag{48}$$
$$q_{3k}(\sigma_{\rm e}^2) \approx p(\sigma_{\rm e}^2|\boldsymbol{y},\mathcal{M}_k)\ . \tag{49}$$

We also wish to compute an approximation $q(\mathcal{M}_k)$ of the posterior pmf for the models. Before we can do that, however, we first have to find the functional expression of the factors q_{ik}, their entropies $H[q_{ik}]$, and the normalisation constants Z_{ik}.

3.1 Functional expressions for the factor $q_{1k}(\boldsymbol{s})$

According to the above tutorial, we have that

$$\ln q_{1k}(\boldsymbol{s}) = \mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{z},\boldsymbol{y}|\mathcal{M}_k)] + \text{const.} \tag{50}$$

Inserting the expression for $p(\boldsymbol{z},\boldsymbol{y}|\mathcal{M}_k)$ in the expectation operator, we obtain

$$\begin{aligned}\mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{z},\boldsymbol{y}|\mathcal{M}_k)] = \mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k)] + \mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{s}|\sigma_{\rm s}^2,\mathcal{M}_k)] \\ + \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\rm s}^2|\mathcal{M}_k)] + \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\rm e}^2|\mathcal{M}_k)]\ .\end{aligned} \tag{51}$$

Of the four terms, the first two terms depend on $\boldsymbol{s}$ whereas the last two do not. However, we still have to compute the last two terms in order to compute the normalisation constant Z_{1k}.

Since

$$\begin{aligned}\ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k) =& -(N/2)\ln(2\pi\sigma_{\rm e}^2) - (1/2)\ln|\det(\boldsymbol{R}_{\rm e}(\boldsymbol{b}_k))| \\ & - \frac{1}{2\sigma_{\rm e}^2}(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\rm e}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s}) \end{aligned} \tag{52}$$

$$\begin{aligned}\ln p(\boldsymbol{s}|\sigma_{\rm s}^2,\mathcal{M}_k) =& -(N/2)\ln(2\pi\sigma_{\rm s}^2) - (1/2)\ln|\det(\boldsymbol{R}_{\rm s}(\boldsymbol{a}_k))| \\ & - \frac{1}{2\sigma_{\rm s}^2}\boldsymbol{s}^T\boldsymbol{R}_{\rm s}^{-1}(\boldsymbol{a}_k)\boldsymbol{s} \end{aligned} \tag{53}$$

$$\ln p(\sigma_{\rm s}^2|\mathcal{M}_k) = \alpha_{{\rm s},k}\ln\beta_{{\rm s},k} - \ln\Gamma(\alpha_{{\rm s},k}) - (\alpha_{{\rm s},k}+1)\ln\sigma_{\rm s}^2 - \frac{\beta_{{\rm s},k}}{\sigma_{\rm s}^2} \tag{54}$$

$$\ln p(\sigma_{\rm e}^2|\mathcal{M}_k) = \alpha_{{\rm e},k}\ln\beta_{{\rm e},k} - \ln\Gamma(\alpha_{{\rm e},k}) - (\alpha_{{\rm e},k}+1)\ln\sigma_{\rm e}^2 - \frac{\beta_{{\rm e},k}}{\sigma_{\rm e}^2}\ , \tag{55}$$

we have for the four terms that

$$\mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k)] = \int q_{3k}(\sigma_{\rm e}^2)\ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\rm e}^2,\mathcal{M}_k)d\sigma_{\rm e}^2 \tag{56}$$

$$\begin{aligned} =& -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{3k}}[\ln\sigma_{\rm e}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\rm e}(\boldsymbol{b}_k))| \\ & - (1/2)(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\rm e}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\,\mathbb{E}_{q_{3k}}[\sigma_{\rm e}^{-2}] \end{aligned} \tag{57}$$

$$\begin{aligned} =& \ln\mathcal{N}(\boldsymbol{s},\mathbb{E}_{q_{3k}}^{-1}[\sigma_{\rm e}^{-2}]\boldsymbol{R}_{\rm e}(\boldsymbol{b}_k)) \\ & - (N/2)\left\{\ln\mathbb{E}_{q_{3k}}[\sigma_{\rm e}^{-2}] + \mathbb{E}_{q_{3k}}[\ln\sigma_{\rm e}^2]\right\} \end{aligned} \tag{58}$$

$$\mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{s}|\sigma_{\rm s}^2,\mathcal{M}_k)] = \int q_{2k}(\sigma_{\rm s}^2)\ln p(\boldsymbol{s}|\sigma_{\rm s}^2,\mathcal{M}_k)d\sigma_{\rm s}^2 \tag{59}$$

$$\begin{aligned} =& -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{2k}}[\ln\sigma_{\rm s}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\rm s}(\boldsymbol{a}_k))| \\ & - (1/2)\boldsymbol{s}^T\boldsymbol{R}_{\rm s}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\,\mathbb{E}_{q_{2k}}[\sigma_{\rm s}^{-2}] \end{aligned} \tag{60}$$

$$\begin{aligned} =& \ln\mathcal{N}(\boldsymbol{0},\mathbb{E}_{q_{2k}}^{-1}[\sigma_{\rm s}^{-2}]\boldsymbol{R}_{\rm s}(\boldsymbol{a}_k)) \\ & - (N/2)\left\{\ln\mathbb{E}_{q_{2k}}[\sigma_{\rm s}^{-2}] + \mathbb{E}_{q_{2k}}[\ln\sigma_{\rm s}^2]\right\} \end{aligned} \tag{61}$$

$$\begin{aligned}\mathbb{E}_{i\neq 1}[\ln p(\sigma_{\rm s}^2|\mathcal{M}_k)] =& \alpha_{{\rm s},k}\ln\beta_{{\rm s},k} - \ln\Gamma(\alpha_{{\rm s},k}) - (\alpha_{{\rm s},k}+1)\,\mathbb{E}_{q_{2k}}[\ln\sigma_{\rm s}^2] \\ & - \beta_{{\rm s},k}\,\mathbb{E}_{q_{2k}}[\sigma_{\rm s}^{-2}] \end{aligned} \tag{62}$$

$$\begin{aligned}\mathbb{E}_{i\neq 1}[\ln p(\sigma_{\rm e}^2|\mathcal{M}_k)] =& \alpha_{{\rm e},k}\ln\beta_{{\rm e},k} - \ln\Gamma(\alpha_{{\rm e},k}) - (\alpha_{{\rm e},k}+1)\,\mathbb{E}_{q_{3k}}[\ln\sigma_{\rm e}^2] \\ & - \beta_{{\rm e},k}\,\mathbb{E}_{q_{3k}}[\sigma_{\rm e}^{-2}] \end{aligned} \tag{63}$$

By only retaining the terms from these four expressions which depend on $\boldsymbol{s}$, we obtain that

$$q_{1k}(\boldsymbol{s}) \propto \mathcal{N}(\boldsymbol{s},\mathbb{E}_{q_{3k}}^{-1}[\sigma_{\rm e}^{-2}]\boldsymbol{R}_{\rm e}(\boldsymbol{b}_k))\mathcal{N}(\boldsymbol{0},\mathbb{E}_{q_{2k}}^{-1}[\sigma_{\rm s}^{-2}]\boldsymbol{R}_{\rm s}(\boldsymbol{a}_k)) \tag{64}$$

from which we can derive that

$$q_{1k}(\boldsymbol{s}) = \mathcal{N}(\hat{\boldsymbol{s}},\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}) \tag{65}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}} = \left[\mathbb{E}_{q_{2k}}[\sigma_{\rm s}^{-2}]\boldsymbol{R}_{\rm s}^{-1}(\boldsymbol{a}_k) + \mathbb{E}_{q_{3k}}[\sigma_{\rm e}^{-2}]\boldsymbol{R}_{\rm e}^{-1}(\boldsymbol{b}_k)\right]^{-1} \tag{66}$$

$$\hat{\boldsymbol{s}} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}\,\mathbb{E}_{q_{3k}}[\sigma_{\rm e}^{-2}]\boldsymbol{R}_{\rm e}^{-1}(\boldsymbol{b}_k)\boldsymbol{y}\ . \tag{67}$$

and that

$$\int \mathcal{N}(\boldsymbol{s}, \mathbb{E}^{-1}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}]\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k))\mathcal{N}(\boldsymbol{0}, \mathbb{E}^{-1}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}]\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))d\boldsymbol{s} = \mathcal{N}(\boldsymbol{0}, \mathbb{E}^{-1}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}]\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k) + \mathbb{E}^{-1}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}]\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k)) \quad (68)$$

by using standard Bayesian inference. Thus, the log-normalisation factor $\ln Z_{1k}$ is given by

$$\ln Z_{1K} = \ln\left[\int \exp\left\{\mathbb{E}_{i\neq 1}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)]\right\} d\boldsymbol{s}\right] \quad (69)$$

$$= \ln \mathcal{N}(\boldsymbol{0}, \mathbb{E}^{-1}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}]\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k) + \mathbb{E}^{-1}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}]\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k)) - (N/2)\left\{\ln \mathbb{E}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}] + \mathbb{E}_{q_{2k}}[\ln \sigma_{\mathrm{s}}^2] + \ln \mathbb{E}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}] + \mathbb{E}_{q_{3k}}[\ln \sigma_{\mathrm{e}}^2]\right\} \quad (70)$$

$$\mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] + \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)]\ . \quad (71)$$

Finally, since $q_{1k}(\boldsymbol{s})$ is a multivariate Gaussian distribution, its entropy is given by

$$H[q_{1k}] = (N/2)\ln(2\pi\exp(1)) + (1/2)\ln|\det(\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}})|\ . \quad (72)$$

3.2 Functional expressions for the factor $q_{2k}(\sigma_{\mathbf{s}}^2)$

According to the above tutorial, we have that

$$\ln q_{2k}(\sigma_{\mathrm{s}}^2) = \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)] + \text{const.} \quad (73)$$

Inserting the expression for $p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)$ in the expectation operator, we obtain

$$\mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)] = \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] + \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] + \mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] + \mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)]\ . \quad (74)$$

Of the four terms, the second and third terms depend on σ_{s}^2 whereas the first and fourth do not. However, we still have to compute all the terms to obtain the normalisation constant Z_{2k}. For the four terms, we have that

$$\mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] = \int q_{1k}(\boldsymbol{s})q_{3k}(\sigma_{\mathrm{e}}^2)\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)d\boldsymbol{s}d\sigma_{\mathrm{e}}^2 \quad (75)$$

$$= -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{3k}}[\ln \sigma_{\mathrm{e}}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k))| - (1/2)\,\mathbb{E}_{q_{1k}}\left[(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\mathrm{e}}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\right]\mathbb{E}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}] \quad (76)$$

$$\mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] = \int q_{1k}(\boldsymbol{s})\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)d\boldsymbol{s} \quad (77)$$

$$= -(N/2)\ln(2\pi) - (N/2)\ln \sigma_{\mathrm{s}}^2 - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))| - \frac{1}{2\sigma_{\mathrm{s}}^2}\,\mathbb{E}_{q_{1k}}\left[\boldsymbol{s}^T\boldsymbol{R}_{\mathrm{s}}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\right] \quad (78)$$

$$\mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] = \alpha_{\mathrm{s},k}\ln\beta_{\mathrm{s},k} - \ln\Gamma(\alpha_{\mathrm{s},k}) - (\alpha_{\mathrm{s},k}+1)\ln\sigma_{\mathrm{s}}^2 - \frac{\beta_{\mathrm{s},k}}{\sigma_{\mathrm{s}}^2} \quad (79)$$

$$\mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] = \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] \quad (80)$$

From this, we have that

$$\ln q_{2k}(\sigma_{\mathrm{s}}^2) = -(N/2 + \alpha_{\mathrm{s},k} + 1)\ln \sigma_{\mathrm{s}}^2 - \frac{(1/2)\,\mathbb{E}_{q_{1k}}\left[\boldsymbol{s}^T \boldsymbol{R}_{\mathrm{s}}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\right] + \beta_{\mathrm{s},k}}{\sigma_{\mathrm{s}}^2} + \text{const.} \tag{81}$$

where all the terms independent of σ_{s}^2 are absorbed in the constant. Thus,

$$q_{2k}(\sigma_{\mathrm{s}}^2) = \text{Inv-}\mathcal{G}(a_{\mathrm{s},k}, b_{\mathrm{s},k}) \tag{82}$$

where

$$a_{\mathrm{s},k} = N/2 + \alpha_{\mathrm{s},k} \tag{83}$$

$$b_{\mathrm{s},k} = (1/2)\,\mathbb{E}_{q_{1k}}\left[\boldsymbol{s}^T \boldsymbol{R}_{\mathrm{s}}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\right] + \beta_{\mathrm{s},k}\;. \tag{84}$$

The log-normalisation factor $\ln Z_{2K}$ can now also be found to

$$\ln Z_{2K} = \ln\left[\int \exp\left\{\mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)]\right\} d\sigma_{\mathrm{s}}^2\right] \tag{85}$$

$$\begin{aligned} &= \ln\left[\int (\sigma_{\mathrm{s}}^2)^{-(a_{\mathrm{s},k}+1)} \exp\left(-\frac{b_{\mathrm{s},k}}{\sigma_{\mathrm{s}}^2}\right) d\sigma_{\mathrm{s}}^2\right] + \alpha_{\mathrm{s},k}\ln\beta_{\mathrm{s},k} - \ln\Gamma(\alpha_{\mathrm{s},k}) \\ &\quad - (N/2)\ln(2\pi) - (N/2)\ln\sigma_{\mathrm{s}}^2 - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))| \\ &\quad + \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] + \mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] \end{aligned} \tag{86}$$

$$\begin{aligned} &= -a_{\mathrm{s},k}\ln b_{\mathrm{s},k} + \ln\Gamma(a_{\mathrm{s},k}) + \alpha_{\mathrm{s},k}\ln\beta_{\mathrm{s},k} - \ln\Gamma(\alpha_{\mathrm{s},k}) \\ &\quad - (N/2)\ln(2\pi) - (N/2)\ln\sigma_{\mathrm{s}}^2 - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))| \\ &\quad + \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] + \mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] \end{aligned} \tag{87}$$

Finally, since $q_{2k}(\sigma_{\mathrm{s}}^2)$ is an inverse Gamma distribution, its entropy is given by

$$H[q_{2k}] = a_{\mathrm{s},k} + \ln(b_{\mathrm{s},k}\Gamma(a_{\mathrm{s},k})) - (1 + a_{\mathrm{s},k})\Psi(a_{\mathrm{s},k}) \tag{88}$$

where $\Psi(\cdot)$ is the digamma function.

3.3 Functional expressions for the factor $q_{3k}(\sigma_{\mathbf{e}}^2)$

According to the above tutorial, we have that

$$\ln q_{3k}(\sigma_{\mathrm{e}}^2) = \mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)] + \text{const.} \tag{89}$$

Inserting the expression for $p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)$ in the expectation operator, we obtain

$$\begin{aligned} \mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{z}, \boldsymbol{y}|\mathcal{M}_k)] &= \mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] + \mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] \\ &\quad + \mathbb{E}_{i\neq 3}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] + \mathbb{E}_{i\neq 3}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)]\;. \end{aligned} \tag{90}$$

Of the four terms, the first and the fourth terms depend on σ_{e}^2 whereas the second and third do not. However, we still have to compute all the terms to obtain the normalisation constant Z_{3k}. For the four terms, we have that

$$\begin{aligned}
\mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\mathrm{e}}^2,\mathcal{M}_k)] &= \int q_{1k}(\boldsymbol{s}) \ln p(\boldsymbol{y}|\boldsymbol{s},\sigma_{\mathrm{e}}^2,\mathcal{M}_k) d\boldsymbol{s} \qquad (91)\\
&= -(N/2)\ln(2\pi) - (N/2)\ln\sigma_{\mathrm{e}}^2 - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k))| \\
&\quad - \frac{1}{2\sigma_{\mathrm{e}}^2}\mathbb{E}_{q_{1k}}\left[(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\mathrm{e}}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\right] \qquad (92)\\
\mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2,\mathcal{M}_k)] &= \int q_{1k}(\boldsymbol{s}) q_{2k}(\sigma_{\mathrm{s}}^2)\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2,\mathcal{M}_k) d\boldsymbol{s} d\sigma_{\mathrm{s}}^2 \qquad (93)\\
&= -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{2k}}[\ln\sigma_{\mathrm{s}}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))| \\
&\quad - (1/2)\,\mathbb{E}_{q_{1k}}\left[\boldsymbol{s}^T\boldsymbol{R}_{\mathrm{s}}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\right]\mathbb{E}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}] \qquad (94)\\
\mathbb{E}_{i\neq 3}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] &= \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] \qquad (95)\\
\mathbb{E}_{i\neq 3}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] &= \alpha_{\mathrm{e},k}\ln\beta_{\mathrm{e},k} - \ln\Gamma(\alpha_{\mathrm{e},k}) - (\alpha_{\mathrm{e},k}+1)\ln\sigma_{\mathrm{e}}^2 - \frac{\beta_{\mathrm{e},k}}{\sigma_{\mathrm{e}}^2} \qquad (96)\\
& \qquad (97)
\end{aligned}$$

From this, we have that

$$\ln q_{3k}(\sigma_{\mathrm{e}}^2) = -(N/2+\alpha_{\mathrm{e},k}+1)\ln\sigma_{\mathrm{e}}^2 - \frac{(1/2)\,\mathbb{E}_{q_{1k}}\left[(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\mathrm{e}}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\right]+\beta_{\mathrm{e},k}}{\sigma_{\mathrm{e}}^2} + \text{const.} \qquad (98)$$

where all the terms independent of σ_{e}^2 are absorbed in the constant. Thus,

$$q_{3k}(\sigma_{\mathrm{s}}^2) = \text{Inv-}\mathcal{G}(a_{\mathrm{e},k}, b_{\mathrm{e},k}) \qquad (99)$$

where

$$\begin{aligned}
a_{\mathrm{e},k} &= N/2 + \alpha_{\mathrm{e},k} \qquad (100)\\
b_{\mathrm{e},k} &= (1/2)\,\mathbb{E}_{q_{1k}}\left[(\boldsymbol{y}-\boldsymbol{s})^T\boldsymbol{R}_{\mathrm{e}}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\right] + \beta_{\mathrm{e},k}\,. \qquad (101)
\end{aligned}$$

The log-normalisation factor $\ln Z_{3K}$ can now also be found to

$$\ln Z_{3K} = \ln \left[\int \exp \left\{ \mathbb{E}_{i \neq 3}[\ln p(\boldsymbol{z}, \boldsymbol{y} | \mathcal{M}_k)] \right\} d\sigma_3^2 \right] \tag{102}$$

$$\begin{aligned} &= \ln \left[\int (\sigma_e^2)^{-(a_{e,k}+1)} \exp \left(-\frac{b_{e,k}}{\sigma_e^2} \right) d\sigma_e^2 \right] + \alpha_{e,k} \ln \beta_{e,k} - \ln \Gamma(\alpha_{e,k}) \\ &\quad - (N/2) \ln(2\pi) - (N/2) \ln \sigma_e^2 - (1/2) \ln |\det(\boldsymbol{R}_e(\boldsymbol{b}_k))| \\ &\quad + \mathbb{E}_{i \neq 3}[\ln p(\boldsymbol{s} | \sigma_e^2, \mathcal{M}_k)] + \mathbb{E}_{i \neq 3}[\ln p(\sigma_s^2 | \mathcal{M}_k)] \end{aligned} \tag{103}$$

$$\begin{aligned} &= -a_{e,k} \ln b_{e,k} + \ln \Gamma(a_{e,k}) + \alpha_{e,k} \ln \beta_{e,k} - \ln \Gamma(\alpha_{e,k}) \\ &\quad - (N/2) \ln(2\pi) - (N/2) \ln \sigma_e^2 - (1/2) \ln |\det(\boldsymbol{R}_e(\boldsymbol{b}_k))| \\ &\quad + \mathbb{E}_{i \neq 3}[\ln p(\boldsymbol{s} | \sigma_e^2, \mathcal{M}_k)] + \mathbb{E}_{i \neq 3}[\ln p(\sigma_s^2 | \mathcal{M}_k)] \end{aligned} \tag{104}$$

Finally, since $q_{3k}(\sigma_e^2)$ is an inverse Gamma distribution, its entropy is given by

$$H[q_{3k}] = a_{e,k} + \ln(b_{e,k} \Gamma(a_{e,k})) - (1 + a_{e,k}) \Psi(a_{e,k}) \; . \tag{105}$$

3.4 Evaluating the expectations

To summarize the main result above, we have that

$$q_{1k}(\boldsymbol{s}) = \mathcal{N}(\hat{\boldsymbol{s}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}}) \tag{106}$$

$$q_{2k}(\sigma_s^2) = \text{Inv-}\mathcal{G}(a_{s,k}, b_{s,k}) \tag{107}$$

$$q_{3k}(\sigma_e^2) = \text{Inv-}\mathcal{G}(a_{e,k}, b_{e,k}) \tag{108}$$

where

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}} = \left[\mathbb{E}_{q_{2k}}[\sigma_s^{-2}] \boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) + \mathbb{E}_{q_{3k}}[\sigma_e^{-2}] \boldsymbol{R}_e^{-1}(\boldsymbol{b}_k) \right]^{-1} \tag{109}$$

$$\hat{\boldsymbol{s}} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}} \, \mathbb{E}_{q_{3k}}[\sigma_e^{-2}] \boldsymbol{R}_e^{-1}(\boldsymbol{b}_k) \boldsymbol{y} \tag{110}$$

$$a_{s,k} = N/2 + \alpha_{s,k} \tag{111}$$

$$b_{s,k} = (1/2) \, \mathbb{E}_{q_{1k}} \left[\boldsymbol{s}^T \boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) \boldsymbol{s} \right] + \beta_{s,k} \tag{112}$$

$$a_{e,k} = N/2 + \alpha_{e,k} \tag{113}$$

$$b_{e,k} = (1/2) \, \mathbb{E}_{q_{1k}} \left[(\boldsymbol{y} - \boldsymbol{s})^T \boldsymbol{R}_e^{-1}(\boldsymbol{b}_k) (\boldsymbol{y} - \boldsymbol{s}) \right] + \beta_{e,k} \; . \tag{114}$$

Since we now have derived the form of the distributions for the factors, we can evaluate the expectations. These are

$$\mathbb{E}_{q_{2k}}[\sigma_s^{-2}] = \frac{a_{s,k}}{b_{s,k}} \tag{115}$$

$$\mathbb{E}_{q_{3k}}[\sigma_e^{-2}] = \frac{a_{e,k}}{b_{e,k}} \tag{116}$$

$$\mathbb{E}_{q_{1k}} \left[\boldsymbol{s}^T \boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) \boldsymbol{s} \right] = \hat{\boldsymbol{s}}^T \boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) \hat{\boldsymbol{s}} + \text{tr} \left(\boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) \boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}} \right) \tag{117}$$

$$\mathbb{E}_{q_{1k}} \left[(\boldsymbol{y} - \boldsymbol{s})^T \boldsymbol{R}_s^{-1}(\boldsymbol{a}_k) (\boldsymbol{y} - \boldsymbol{s}) \right] = \hat{\boldsymbol{e}}^T \boldsymbol{R}_e^{-1}(\boldsymbol{b}_k) \hat{\boldsymbol{e}} + \text{tr} \left(\boldsymbol{R}_e^{-1}(\boldsymbol{b}_k) \boldsymbol{\Sigma}_{\hat{\boldsymbol{s}}} \right) \tag{118}$$

where $\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{s}}$. For the evaluation of the normalisation constants, we also have that

$$\mathbb{E}_{q_{2k}}[\ln \sigma_{\mathrm{s}}^2] = \ln b_{\mathrm{s},k} - \Psi(a_{\mathrm{s},k}) \tag{119}$$

$$\mathbb{E}_{q_{3k}}[\ln \sigma_{\mathrm{e}}^2] = \ln b_{\mathrm{e},k} - \Psi(a_{\mathrm{e},k}) \ . \tag{120}$$

3.5 Computing the model factor $q(\mathcal{M}_k)$

To evaluate the model factor $q(\mathcal{M}_k)$, we have to compute the lower bound $\mathcal{L}_k(q_k)$ for all models. This lower bound can be computed from the normalisation factors $\{Z_{ik}\}_{i=1}^3$ and the entropies for $q_{1k}(\boldsymbol{s})$, $q_{2k}(\sigma_{\mathrm{s}}^2)$, and $q_{3k}(\sigma_{\mathrm{e}}^2)$. Alternatively, we also have from (23) that $\mathcal{L}_k(q_k)$ can be written as

$$\mathcal{L}_k(q_k) = \mathbb{E}_{q_k}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] + \mathbb{E}_{q_k}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] + \mathbb{E}_{q_k}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] + \mathbb{E}_{q_k}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] + \sum_{i=1}^{M} H[q_{ik}] \ . \tag{121}$$

Since none of the expectations depends on all three sets of unknowns, we have that

$$\mathbb{E}_{q_k}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] = \mathbb{E}_{i\neq 2}[\ln p(\boldsymbol{y}|\boldsymbol{s}, \sigma_{\mathrm{e}}^2, \mathcal{M}_k)] \tag{122}$$

$$= -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{3k}}[\ln \sigma_{\mathrm{e}}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{e}}(\boldsymbol{b}_k))| - (1/2)\,\mathbb{E}_{q_{1k}}\left[(\boldsymbol{y}-\boldsymbol{s})^T \boldsymbol{R}_{\mathrm{e}}^{-1}(\boldsymbol{b}_k)(\boldsymbol{y}-\boldsymbol{s})\right]\mathbb{E}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}] \tag{123}$$

$$\mathbb{E}_{q_k}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] = \mathbb{E}_{i\neq 3}[\ln p(\boldsymbol{s}|\sigma_{\mathrm{s}}^2, \mathcal{M}_k)] \tag{124}$$

$$= -(N/2)\ln(2\pi) - (N/2)\,\mathbb{E}_{q_{2k}}[\ln \sigma_{\mathrm{s}}^2] - (1/2)\ln|\det(\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{a}_k))| - (1/2)\,\mathbb{E}_{q_{1k}}\left[\boldsymbol{s}^T \boldsymbol{R}_{\mathrm{s}}^{-1}(\boldsymbol{a}_k)\boldsymbol{s}\right]\mathbb{E}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}] \tag{125}$$

$$\mathbb{E}_{q_k}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] = \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] = \mathbb{E}_{i\neq 3}[\ln p(\sigma_{\mathrm{s}}^2|\mathcal{M}_k)] \tag{126}$$

$$= \alpha_{\mathrm{s},k}\ln\beta_{\mathrm{s},k} - \ln\Gamma(\alpha_{\mathrm{s},k}) - (\alpha_{\mathrm{s},k}+1)\,\mathbb{E}_{q_{2k}}[\ln \sigma_{\mathrm{s}}^2] - \beta_{\mathrm{s},k}\,\mathbb{E}_{q_{2k}}[\sigma_{\mathrm{s}}^{-2}] \tag{127}$$

$$\mathbb{E}_{q_k}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] = \mathbb{E}_{i\neq 1}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] = \mathbb{E}_{i\neq 2}[\ln p(\sigma_{\mathrm{e}}^2|\mathcal{M}_k)] \tag{128}$$

$$= \alpha_{\mathrm{e},k}\ln\beta_{\mathrm{e},k} - \ln\Gamma(\alpha_{\mathrm{e},k}) - (\alpha_{\mathrm{e},k}+1)\,\mathbb{E}_{q_{3k}}[\ln \sigma_{\mathrm{e}}^2] - \beta_{\mathrm{e},k}\,\mathbb{E}_{q_{3k}}[\sigma_{\mathrm{e}}^{-2}] \ . \tag{129}$$

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006. ISBN: 0387310738.

[2] J. K. Nielsen et al. "Model-based Noise PSD Estimation from Speech in Non-stationary Noise". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* 2018.