

Identifying influential model choices in Bayesian hierarchical models

Ida Scheel^{*} Peter J. Green[†] Jonathan C. Rougier[†]

May, 2008

Abstract

Real-world phenomena are frequently modelled by Bayesian hierarchical models. The building-blocks in such models are the distribution of each variable conditional on parent and/or neighbour variables in the graph. The specifications of centre and spread of these conditional distributions may be well-motivated, while the tail specifications are often left to convenience. However, the posterior distribution of a parameter may depend strongly on such arbitrary tail specifications. This is not easily detected in complex models. In this paper we propose a graphical diagnostic which identifies such influential statistical modelling choices at the node level in any chain graph model. Our diagnostic, *the local critique plot*, examines local conflict between the information coming from the parents and neighbours (local prior) and from the children and co-parents (lifted likelihood). It identifies properties of the local prior and the lifted likelihood that are influential on the posterior density. We illustrate the use of the local critique plot with applications involving models of different levels of complexity. The local critique plot can be derived for all parameters in a chain graph model, and is easy to implement using the output of posterior sampling.

1 Introduction

Bayesian hierarchical models are now widely used to model complex, structured data. Such models are built from a large number of individual factors, representing the conditional distributions of each variable given those higher in the hierarchy, or, in the case of undirected models, potential functions for cliques of variables. Responsible, disciplined model-building requires that specification of all these factors should properly take into account prior information, whether this codifies scientific laws, earlier experiments, or degrees of subjective belief. However, this specification is a very challenging task, and there will often be a concern that it has been done imperfectly. In particular, while it may be relatively easy to specify the location and spread of a marginal or conditional distribution, the *shape* of the distribution, especially in the tail, is a more taxing question. Yet the posterior distribution of all unknowns given data may depend on the trading off of tails of individual model factors. It is important that this phenomenon be detected so that the modeller's attention can be drawn to particular statistical choices that are influential in the analysis, in order to confirm them or to reconsider.

In a simple Bayesian model, conflict between prior and data is easily detected, and this provides a diagnostic for criticising statistical modelling choices. Suppose we have independent data (y_1, \dots, y_{10}) with $\bar{y} = 12$ which are modelled as Normal with mean μ and precision 0.1. The prior distribution for μ is Normal with mean 8 and precision 0.5. For this simple example it is easy to see that much of the posterior density for

^{*}Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway
Email: idasch@math.uio.no (corresponding author)

[†]Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK

μ is dependent on the right tail of the prior and the left tail of the likelihood. It is easy to illustrate (as in Figure 1) because both the prior and the likelihood for μ are fixed (i.e. they have no random parameters). In a general hierarchical model, identifying conflict between the sources of information contributing to the posterior distribution of a single node is a more subtle matter. This paper introduces a graphical diagnostic for this purpose.

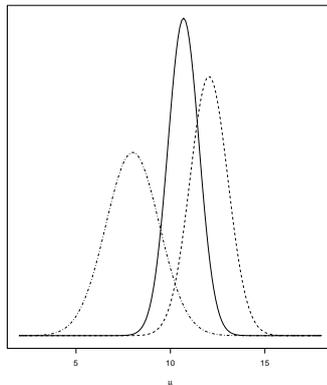


Figure 1: Plot of prior (dotdashed), normalised likelihood (dashed) and posterior (solid) with $N(8, 0.5)$ prior for μ

Bayesian model criticism is often performed by considering a Bayesian p -value describing the compatibility of the observed data and the model. Such a p -value is typically obtained from some test-statistic or discrepancy measure (possibly depending on parameters as well as data) reflecting important aspects of the model, and a predictive distribution for this discrepancy measure. The type of predictive distribution used varies, e.g. the prior predictive distribution (Box, 1980), the posterior predictive distribution (Guttman, 1967; Rubin, 1984; Gelman et al., 1996; Meng, 1994), and the partial posterior predictive distribution (Bayarri and Berger, 1999, 2000; Bayarri and Castellanos, 2007). The latter approach avoids the need for informative prior distributions, as in the prior predictive approach, as well as the conservatism caused by the double use of data, as in the posterior predictive approach. This conservatism may also be handled by calibration (Hjort et al., 2006). These p -values are usually directed at one specific aspect of a model, not considering model fit at the individual nodes of a hierarchical model. A method for checking all stages of a hierarchical model was proposed by Dey et al. (1998), though it is highly computationally intensive. Our idea of looking for conflict between the prior and likelihood information at the node level is not new. O’Hagan (2003) extends the node level residual analysis of Chaloner (1994) to other measures of conflict, to look for conflict between the different sources of information provided for the node in question. In practice, this is done by looking at how much the densities representing two different sources of information overlap, measured by the height of the densities (normalised to have unit maximum height) at the point where the two cross. A modification of this, avoiding double use of data, is proposed in Dahl et al. (2007). Marshall and Spiegelhalter (2007) propose a similar p -value for measuring conflict at the node level in hierarchical models, which also avoids specifying a discrepancy measure and acts as an approximation to their cross-validators, mixed p -value (Marshall and Spiegelhalter, 2003, 2007), when it exists.

However, none of the above-mentioned conflict measures really address the nature of the conflict and the impact certain aspects of the prior and the likelihood have on the posterior analysis. The diagnostic we propose examines conflict at the node level by identifying where the posterior samples of a variable are located in what we call the local prior (the information coming from the parents and/or neighbours) and what we call the lifted likelihood (the information coming from children and co-parents). It has the

ability to identify properties of the local prior and lifted likelihood that are influential on the posterior density. The tool we propose may be used in any chain graph model, thus is applicable to virtually all hierarchical models in routine statistical practice. It is easy to implement using the output of a sampler, such as the Gibbs sampler. In our examples we have run posterior simulations using WinBUGS (Lunn et al., 2000) and the R (<http://cran.r-project.org>) package BRugs which contains OpenBUGS (<http://www.mathstat.helsinki.fi/openbugs>).

The outline of the paper is as follows. In Section 2 we start by providing some theory on chain graphs, then we introduce the local prior, the lifted likelihood and our graphical diagnostic for chain graph models, *the local critique plot*, and explain how to interpret this new diagnostic. In Section 3 we illustrate the use of the local critique plot by applying it to three different types of chain graph models; a 2-level Directed acyclic graph (DAG), a 3-level DAG and a DAG combined with a Markov Random Field (MRF). In Section 4 we conclude on the abilities of the local critique plot.

2 Methodology

2.1 Chain graphs

Consider a graph $\mathcal{G} = (V, E)$ where V is the set of vertices and E the set of edges. Vertex i represents the random variable x_i , which can be observed or unobserved. Constants are not represented in the graph. For a subset $A \subseteq V$ let $x_A = \{x_i\}_{i \in A}$ and $x = x_V$. Let O be the subset of vertices representing the observed variables (the data), and U be subset of vertices representing the unobserved variables (parameters or missing data), such that $O \cup U = V$. This paper has a Bayesian perspective and we consider parameters as random variables. We only consider continuous parameters defined on the real line, but similar theory could be developed for discrete parameters. The edges in the graph can be directed or undirected. A subset of a graph where all the vertices are connected by directed or undirected edges is said to form a complete subgraph. If there is a directed edge from vertex j to vertex i , j is a parent of i and i is a child of j , and we write $\text{pa}(i)$ for the set of parents and $\text{ch}(i)$ for the set of children of vertex i ($\text{pa}(i)$ or $\text{ch}(i)$ may be empty). Furthermore, $\text{pa}(A)$ is the set of parents of the vertices in subset A (may be empty). If there is an undirected edge between vertex j and vertex i , j is in the neighbourhood of i (and vice versa) and we write $i \sim j$. The set $\text{ne}(i) = \{j : i \sim j\}$ is then called the neighbourhood of vertex i (may be empty). The directed and undirected edges of the graph encode conditional independence properties among the variables, in a sense shortly to be made precise.

In this paper we focus on a class of graphs called chain graphs. Suppose that V can be partitioned into numbered subsets $V(c)$, $c \in \mathcal{T} = \{1, \dots, T\}$, such that all edges between subsets are directed, with direction from the subset with the lower number to the one with the higher number, and edges within subsets are undirected. Denote the edges within $V(c)$ as $E(c)$. If and only if $(V(c), E(c))$ form undirected subgraphs (which do not have to be complete) for all c then \mathcal{G} is a chain graph and $V(c)$, $c \in \mathcal{T}$ are called chain components (see Lauritzen, 1996, chapter 2). \mathcal{G} has no directed cycles. If a vertex i has no undirected edges connected to it, i forms a chain component with itself as the only member. In the special case that all chain components are such single vertices, \mathcal{G} is a Directed acyclic graph (DAG).

Now suppose the vector of all random variables x of the chain graph takes values in \mathcal{X} , and that probabilistic statements on the graph are described by some probability distribution \mathcal{P} on \mathcal{X} with joint density $p(x)$. Markov properties are implied by two factorisation assumptions on p , the first of which is

$$(1) \quad p(x) = \prod_{c \in \mathcal{T}} p(x_{V(c)} | x_{\text{pa}(V(c))}),$$

where $p(x_A|x_B)$ denotes a density of the variables x_A for any subset $A \in V$, given the variables x_B . The second assumption is a further factorisation of the factors in (1). Let \mathcal{G}_c be the undirected subgraph with nodes $V(c) \cup \text{pa}(V(c))$ and undirected edges between two nodes if either they are both in $\text{pa}(V(c))$ or there is a directed or undirected edge between them in \mathcal{G} . Let $W(c)$ be the collection of all subsets of $V(c) \cup \text{pa}(V(c))$ that form complete subgraphs in \mathcal{G}_c . The second factorisation assumption is

$$(2) \quad p(x_{V(c)}|x_{\text{pa}(V(c))}) = \prod_{a \in W(c)} \phi_a(x), \quad c \in \mathcal{T},$$

where $\phi_a(x)$ is some function that depends on x only through x_a . The probability distribution \mathcal{P} is said to factorise according to \mathcal{G} if p satisfies both (1) and (2) (see Lauritzen, 1996, chapter 3.2.3).

2.2 The local critique plot

The assumptions (1) and (2) lead to the following full conditional distribution for the variable x_i

$$(3) \quad p(x_i|x_{-i}) \propto p(x_i|x_{\text{pa}(i)}, x_{\text{ne}(i)}) \prod_{c:i \in \text{pa}(V(c))} p(x_{V(c)}|x_{\text{pa}(V(c))}), \quad i \in V,$$

where x_{-i} are all variables except variable x_i . In this paper we call

$$(4) \quad p_i(x) = p(x_i|x_{\text{pa}(i)}, x_{\text{ne}(i)}), \quad i \in U$$

the local prior for x_i and

$$(5) \quad l_i(x) = \prod_{c:i \in \text{pa}(V(c))} p(x_{V(c)}|x_{\text{pa}(V(c))}), \quad i \in U$$

the lifted likelihood for x_i . The justification for the names "local prior" and "lifted likelihood" lies in that in the case of a simple two-level model in which x_i is alone at the higher level (i.e. the only parameter), (4) would in Bayesian statistics be called the prior and (5) the likelihood for x_i . One should be aware that if \mathcal{G} is a non-DAG, $p(x_{V(c)}|x_{\text{pa}(V(c))})$ in (5) may contain functions of x_i that may be unexpected from a glance at the graph. This follows from the factorisation requirement (2). In the special case of only two chain components, one of which is a Markov Random Field (MRF) with vertices $i \in Q$ and the other is a single vertex called p , which is a parent to all the vertices in the MRF, (5) for x_p reduces to

$$(6) \quad l_p(x) = p(x_Q|x_p).$$

In the special case that \mathcal{G} is a DAG, (5) reduces to

$$(7) \quad l_i(x) = \prod_{j:i \in \text{pa}(j)} p(x_j|x_{\text{pa}(j)}), \quad i \in U.$$

To illustrate the concepts "local priors" and "lifted likelihoods", we consider a DAG example called Pumps. It concerns the numbers of failures of $n = 10$ power plant pumps. Pump i has y_i failures, operation time t_i (in 1000s of hours) and failure rate θ_i . The model implemented in Spiegelhalter et al. (2004) is

$$(8) \quad \begin{aligned} y_i|\theta_i &\sim \text{Poisson}(\theta_i t_i), \quad i = 1, \dots, n \\ \theta_i|\alpha, \beta &\sim \text{Gamma}(\alpha, \beta), \quad i = 1, \dots, n \\ \alpha &\sim \text{Gamma}(1, 1) \\ \beta &\sim \text{Gamma}(0.1, 1). \end{aligned}$$

This example originates from Gaver and O’Muirheartaigh (1987), with α and β fixed at empirical Bayes estimates. Gelfand and Smith (1990) also use an empirical Bayes estimate for α , but assume a Gamma distribution for β , while George et al. (1993) assume Gamma distributions for both α and β .

Here $x = (y_1, \dots, y_n, \theta_1, \dots, \theta_n, \alpha, \beta)$. The DAG can be seen in Figure 2. The local priors for the parameters in this model are

$$(9) \quad \begin{aligned} p_\alpha(x) &= \exp(-\alpha) \\ p_\beta(x) &= \frac{\beta^{-0.9}}{\Gamma(0.1)} \exp(-\beta) \\ p_{\theta_i}(x) &= \frac{\beta^\alpha \theta_i^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta_i), \quad i = 1, \dots, n, \end{aligned}$$

and the lifted likelihoods

$$(10) \quad \begin{aligned} l_\alpha(x) &= l_\beta(x) = \prod_{i=1}^n \frac{\beta^\alpha \theta_i^{\alpha-1} \exp(-\beta\theta_i)}{\Gamma(\alpha)} \\ l_{\theta_i}(x) &\propto \theta_i^{y_i} \exp(-t_i\theta_i), \quad i = 1, \dots, n. \end{aligned}$$

” $l_i(x) \propto f(x)$ ” implies throughout this paper that $l_i(x)$ is proportional to $f(x)$ with respect to x_i . Furthermore, in applications the parameter name is frequently used as subscript on functions instead of i (e.g. $l_\alpha(x)$ instead of $l_{2n+1}(x)$ in Pumps).

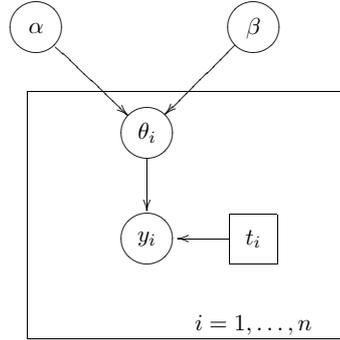


Figure 2: DAG for the Pumps model (squares represent constants that are not true vertices in the graph).

Based on (3-5) we propose two functions to identify influential statistical modelling choices. Assuming that the local prior $p_i(x)$ is proper, $\pi_i(x)$ is the distribution function of the local prior for x_i

$$(11) \quad \pi_i(x) = \int_{-\infty}^{x_i} p_i(x^{i \rightarrow u}) du, \quad i \in U.$$

where $x^{i \rightarrow u}$ denotes x with the i 'th element replaced by u . Notice that it is a function of x only through x_i , $x_{\text{pa}(i)}$ and $x_{\text{ne}(i)}$. It expresses where x_i is located in the support of the local prior density given the states of $\text{pa}(i)$ and $\text{ne}(i)$ in x . The cumulative distribution function is used in order to map x_i onto a standard probability scale given $x_{\text{pa}(i)}$ and $x_{\text{ne}(i)}$.

The second function we propose is the cumulative normalised lifted likelihood of x_i

$$(12) \quad \psi_i(x) = \frac{\int_{-\infty}^{x_i} l_i(x^{i \rightarrow u}) du}{\int_{-\infty}^{\infty} l_i(x^{i \rightarrow u}) du}, \quad i \in U,$$

We assume for the moment that the lifted likelihood is integrable with respect to x_i , so that the denominator is finite and (12) well-defined; further we assume that the resulting $\psi_i(x)$ is available in closed form or equals a known cumulative distribution function. We discuss other cases in Section 2.4. $\psi_i(x)$ identifies where x_i is located in the support of its lifted likelihood function given the state of $\{x_{V(c)}, x_{\text{pa}(V(c))}\}_{c:i \in \text{pa}(V(c))}$ in x . A cumulative distribution is used in order to map x_i onto a standard $[0, 1]$ scale given $\{x_{V(c)}, x_{\text{pa}(V(c))}\}_{c:i \in \text{pa}(V(c))}$. Because the lifted likelihood (5) is not a distribution with respect to the parameter x_i , normalisation is needed. Notice that $\psi_i(x)$ is a function of x only through the set $\{x_{V(c)}, x_{\text{pa}(V(c))}\}_{c:i \in \text{pa}(V(c))}$, which contains x_i . The intersection of the random variables involved in $\psi_i(x)$ and the ones involved in $\pi_i(x)$ may contain more than just x_i .

A third way to score x_i is in terms of its posterior distribution. Let $p_i(x^{i \rightarrow u} | x_O)$ be the marginal posterior density of x_i and

$$(13) \quad \xi_i(x) = \int_{-\infty}^{x_i} p_i(x^{i \rightarrow u} | x_O) du, \quad i \in U,$$

which is the cumulative posterior distribution function for x_i . For consistency in notation, we treat it as a function of the whole vector x . $\xi_i(x)$ is intended to be supplementary to $\pi_i(x)$ and $\psi_i(x)$, which are of main interest. Our diagnostic, which we call *the local critique plot*, is a graphical visualisation of the joint posterior density of $\pi_i(x)$, $\psi_i(x)$ and $\xi_i(x)$ for each $i \in U$, given the data x_O . Generally, $\pi_i(x)$ and $\psi_i(x)$ are not independent conditional on $\xi_i(x)$ and x_O . Except for very simple models, the joint posterior distribution of $\pi_i(x)$, $\psi_i(x)$ and $\xi_i(x)$ can be very complex and intractable analytically. Because this is the case also for the posterior density of x_U , an obvious solution is Markov Chain Monte Carlo (MCMC) simulation. The posterior sample of x_U provided by MCMC simulation can be directly plugged into the $\pi_i(x)$ and $\psi_i(x)$ functions. Assume a MCMC simulation with a total of M iterations (after burn in) is performed. The set $x_A^{(t)}$ are the values of x_A at the t 'th iteration. We plug the samples $\{x^{(t)}\}_{t=1}^M$ into explicit formulas for $\pi_i(x)$ and $\psi_i(x)$ to obtain the samples $\pi_i(x^{(t)})$ and $\psi_i(x^{(t)})$. An estimate of $\xi_i(x)$ can be derived directly from the MCMC simulation

$$(14) \quad \hat{\xi}_i(x^{(t)}) = \frac{1}{M} \sum_{s=1}^M I_{x_i^{(s)} \leq x_i^{(t)}}, \quad i \in U.$$

For each x_i , $i \in U$, we record the value of our diagnostic by plotting the sampled values of $\pi_i(x)$, $\psi_i(x)$ and $\xi_i(x)$, that is $\{(\pi_i(x^{(t)}), \psi_i(x^{(t)})), \hat{\xi}_i(x^{(t)})\}_{t=1}^M$, on $[0, 1] \times [0, 1] \times [0, 1]$, with π along the vertical axis, ψ along the horizontal axis and the ξ dimension expressed by marking of the points. The marking used in this paper is done by substituting each point by a line segment with an angle $\in [0, \frac{\pi}{2}]$, where the angle relative to the π axis represents the value of $\xi_i(x)$, transformed from $[0, 1]$ to $[0, \frac{\pi}{2}]$. The value of $(\pi_i(x), \psi_i(x))$ is the centre of the plotting line segment. A vertical line segment indicates $\xi_i(x) = 0$ and a horizontal line segment indicates $\xi_i(x) = 1$. ξ can be rendered in various ways. Rainbow colouring of the points according to $\xi_i(x)$ seems to be the easiest to interpret, but because colour printers still are rare, we have chosen the alternative described above. Grey scale is a possibility, but it may encourage the interpretation that the light coloured points are of less importance, which is not the case.

2.3 Interpretation of the local critique plots

If the mass is concentrated in the upper left corner or the lower right corner of the plotting region, it means that $x_i | x_O$ is located mainly in the tails of both the local prior and the lifted likelihood (opposite tails). In other words, the marginal posterior density of x_i is a result of a trading off these tails. If x_i is considered important, there is reason to rethink the model carefully to ensure that the model is well specified. In particular, one should reconsider the tails of the local prior and the lifted likelihood. Less extreme, but still a reason to re-examine the model, is the case when the mass is well spread out in the π dimension, but concentrated close to one of the vertical edges in the plot, which means that the posterior density of x_i is using only the tail of its lifted likelihood, but most of its local prior. Conversely, if the posterior density of x_i is using most of its lifted likelihood, but mainly the tail of its local prior, the mass will be well spread out in the ψ dimension, but concentrated close to one of the horizontal edges in the plot.

If the plot shows that the mass is gathered along a thin line parallel to the ψ axis, it means that the posterior density of x_i is using only a small part of the local prior. This is not surprising if the prior specification for x_i is non-informative, or if the lifted likelihood is highly informative because of large amounts of data. On the other hand, if the plot shows that the mass is gathered along a thin line parallel to the π axis, it means that the posterior density of x_i is using only a small part of the lifted likelihood, which may be more disturbing. Of course, there may be good reasons why the lifted likelihood is non-informative on x_i .

The mass will never be concentrated in the lower left corner or the upper right corner of the plotting region, unless there is something wrong in the simulation. The reason is simply that x_i values located in the same side tails of the local prior and the lifted likelihood will not have the highest posterior density values.

Generally, $\pi_i(x)$ and $\psi_i(x)$ do not follow a $\text{Unif}(0, 1)$ distribution. How close the distributions of $\pi_i(x)$ and $\psi_i(x)$ are to $\text{Unif}(0, 1)$ depends on the similarity between the local prior and the full conditional and between the lifted likelihood and the full conditional, respectively. In the special case that the lifted likelihood is completely flat, i.e. $l_i(x) \propto 1$, then $p_i(x) = p(x_i|x_{-i})$ and $\pi_i(x) \sim \text{Unif}(0, 1)$. Similarly, in the special case that the local prior is completely flat, i.e. $p_i(x) \propto 1$, then $l_i(x) = p(x_i|x_{-i})$ and $\psi_i(x) \sim \text{Unif}(0, 1)$. Suppose neither the local prior or the lifted likelihood are flat, but that they agree well on the location and spread of the distribution of x_i . Generally, the degree of such similarity between $p_i(x)$ and $l_i(x)$ will not be constant because they can both depend on random parameters. The distributions of $\pi_i(x)$ and $\psi_i(x)$ will in such a case not be $\text{Unif}(0, 1)$, but their supports will still cover most of $[0, 1]$ and their modes will be approximately 0.5.

To illustrate the local critique plots, we return to the simple Normal example from the Introduction. First we introduce some general notation on distributions. If z is Normal with mean μ and precision (inverse variance) τ , we write $z \sim N(\mu, \tau)$ and $\Phi(u; \mu, \tau) = P(z \leq u)$. If z is Gamma with shape α and rate (inverse scale) β , we write $z \sim \text{Ga}(\alpha, \beta)$ and $\Gamma(v; \alpha, \beta) = P(z \leq v)$. Lastly, $z \sim \log N(\mu, \tau)$ means that z is log-Normal where the mean and precision of the logarithm of z are μ and τ , respectively, and $\Omega(v; \mu, \tau) = P(z \leq v)$. The model for the simple Normal example is

$$(15) \quad \begin{aligned} y_i | \mu, \tau &\sim N(\mu, \tau), \quad i = 1, \dots, n \\ \mu &\sim N(\mu_0, \tau_0), \end{aligned}$$

with either fixed precision τ , in which case $x = (y_1, \dots, y_n, \mu)$, or random precision τ , in which case $x = (y_1, \dots, y_n, \mu, \tau)$. We focus our attention on the parameter μ . This is a DAG where the local prior and

lifted likelihood for μ are identical to the conventional prior and likelihood

$$(16) \quad \begin{aligned} \pi_\mu(x) &= \Phi(\mu; \mu_0, \tau_0) \\ l_\mu(x) &= \prod_{i=1}^n p(y_i | \mu, \tau) \propto \exp\left(-\frac{\tau n}{2} (\mu - \bar{y})^2\right) \Rightarrow \psi_\mu(x) = \Phi(\mu; \bar{y}, \tau n). \end{aligned}$$

The data used in this example have $n = 10$ and $\bar{y} = 12$. Figure 3 (a) shows the local critique plot for μ when $\tau = 0.1$, $\mu_0 = 8$ and $\tau_0 = 0.5$ (Model1) based on 10000 MCMC posterior simulations. It is a further demonstration of what can be seen in Figure 1; the posterior samples of μ are located in the right tail of its local prior and the left tail of its lifted likelihood. The local critique plot for μ when $\tau \sim \text{Ga}(1e-6, 1e-6)$, $\mu_0 = 11$ and $\tau_0 = 0.5$ (Model2) can be seen in Figure 3 (b) (also based on 10000 MCMC posterior simulations). The posterior samples of μ for Model2 are more spread out across both the local prior and the lifted likelihood than when using Model1, where the local prior mean for μ is farther away from \bar{y} . Also, because τ is now made random, there is randomness in ψ direction, which is explained below. The information from the local critique plots in Figure 3 (a) and (b) may be obvious just from looking at the model, but for models with more complicated structures conflicts of this kind are not at all obvious, and our graphical diagnostic serves a useful purpose.

For some variables the relationship between $\pi_i(x)$ and $\xi_i(x)$ or between $\psi_i(x)$ and $\xi_i(x)$ is one-to-one, given the observed data x_O . If i has no random parents or neighbours, i.e. if

$$(17) \quad \text{pa}(i) \cup \text{ne}(i) \subseteq O,$$

then the local prior (4) is a function only of x_i , i.e. $p_i(x) = p(x_i)$, and therefore $\pi_i(x) = g(x_i)$, where g is strictly increasing. Hence the posterior distribution of $\pi_i(x)$ can be completely described in terms of the posterior distribution of x_i , i.e. $\xi_i(x)$, and g . If i is the parent only of observed data and has no random co-parents, i.e. if

$$(18) \quad \text{ch}(i) \subseteq O \quad \text{and} \quad \cup_{c:i \in \text{pa}(V(c))} \text{pa}(V(c)) = \{i\},$$

then x_i is the only random variable in the lifted likelihood (5), i.e. $l_i(x) = p(x_O | x_i)$, and consequently $\psi_i(x) = h(x_i)$, where h is strictly increasing. Hence the posterior distribution of $\psi_i(x)$ can be completely described in terms of the posterior distribution of x_i , i.e. $\xi_i(x)$, and h . In Model2, μ has no random parents or neighbours, but is a co-parent with the random parameter τ , thus (17) is true but not (18). In Figure 3 (b) we see that the angle of the plotting line segment relative to the π axis (which corresponds to the value of $\xi_\mu(x)$) is strictly increasing as a function of $\pi_\mu(x)$. In Model1, both (17) and (18) are true, i.e. μ has no random parents, neighbours or co-parents and is the parent only of observed data. Then given a value of $\xi_\mu(x)$, both $\pi_\mu(x)$ and $\psi_\mu(x)$ are fixed and the angle of the plotting line segment relative to the π axis is strictly increasing both as a function of $\pi_\mu(x)$ and of $\psi_\mu(x)$, as can be seen in Figure 3 (a). In the limiting case that both (17) and (18) are true and the lifted likelihood and the local prior provide exactly the same information about x_i , i.e. $l_i(x) \propto p_i(x)$, then the local critique plot for x_i consists of the straight line $\pi_i(x) = \psi_i(x)$.

2.4 Approximate integration of the lifted likelihood

Computationally it is convenient if the lifted likelihood of the variable x_i has a conditional conjugate family, because then $\psi_i(x)$ equals a known cumulative distribution function. But sometimes it is not easy to integrate the lifted likelihood $l_i(x)$ (5), or $l_i(x)$ may not even be integrable. In the latter case, $\psi_i(x)$ is not well-defined by (12). As a solution to both these issues, we suggest first to determine an interval $[a, b]$

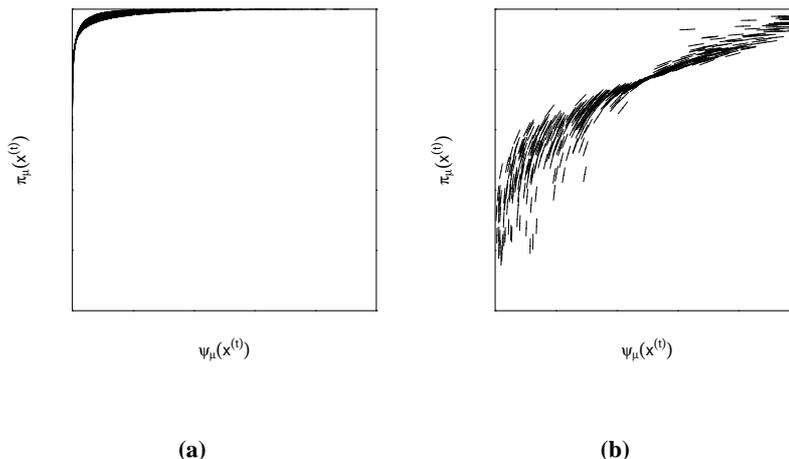


Figure 3: The local critique plots for μ in (a) Model1 ($\tau = 0.1$, $\mu_0 = 8$ and $\tau_0 = 0.5$) and (b) Model2 ($\tau \sim \text{Ga}(1e-6, 1e-6)$, $\mu_0 = 11$ and $\tau_0 = 0.5$) ($M = 10000$, results are shown for a random subsample of size 500).

that covers the bulk of the $l_i(x)$ mass, and then perform numerical integration on this interval to obtain $\hat{\psi}_i(x)$. When x_i has random, unobserved children and/or random co-parents, new integration boundaries $[a, b]$ must be determined at each simulation iteration to be plotted. Hence the algorithms for determining $[a, b]$ and performing the integrations should be as efficient as possible. Suppose x_i lies between a_{\min} and b_{\max} . The algorithm we propose for determining suitable bounds $[a, b]$ for the numerical integration starts with an initial interval $[a_0, b_0]$ (e.g. the 10th and 90th percentiles in the posterior sample of x_i), inside which $\log l_i(x)$ is evaluated at m equally spaced abscissas. The interval is then expanded iteratively, until the logarithm of the lifted likelihood at either end is at least $\frac{q}{2}$ less than the maximum in the interval, where q is a high quantile of the χ_1^2 distribution (e.g. 99%). Pseudocode for this program can be seen in the supplemental material. Once an appropriate interval has been determined, $\psi_i(x)$ can be approximated using the Trapezium Rule. Now, $\hat{\psi}_i(x) = 0$ for $x_i \leq a$ and $\hat{\psi}_i(x) = 1$ for $x_i \geq b$. Because $[a, b]$ generally does not cover all of the support of $l_i(x)$, the values 0 and 1 do not have the same interpretations for $\hat{\psi}_i(x)$ as for $\psi_i(x)$. To visualise this fact we suggest suppressing the plotting of the boundaries for ψ at 0 and 1.

To illustrate the use of this approach we return to the Pumps example. The lifted likelihood for α is

$$(19) \quad l_\alpha(x) \propto \left(\frac{\left(\beta \prod_{i=1}^n \theta_i^{\frac{1}{n}} \right)^\alpha}{\Gamma(\alpha)} \right)^n = \left(\frac{\kappa^\alpha}{\Gamma(\alpha)} \right)^n, \quad \kappa = \beta \prod_{i=1}^n \theta_i^{\frac{1}{n}}.$$

It is not possible to integrate this lifted likelihood analytically. The local critique plot for α with $\psi_\alpha(x)$ estimated by $\hat{\psi}_\alpha(x)$, using $m = 6$ and the Trapezoid Rule on 200 subintervals, can be seen in Figure 4 (a).

To assess how well this approximation is doing we performed another approximation. $l_\alpha(x)$ depends on the parameters β and θ_i , $i = 1, \dots, n$ only through κ . An examination of $l_\alpha(x)$ for different values of κ suggests that the normalised $l_\alpha(x)$ can be approximated well by a Normal distribution with mean equal to the mode of $l_\alpha(x)$ (α^* , found by optimisation) and precision equal to the estimated Fisher Information (estimated by setting $\alpha = \alpha^*$). The Fisher Information for α from $l_\alpha(x)$ is $n\phi_1(\alpha)$, where $\phi_1(\alpha)$ is the

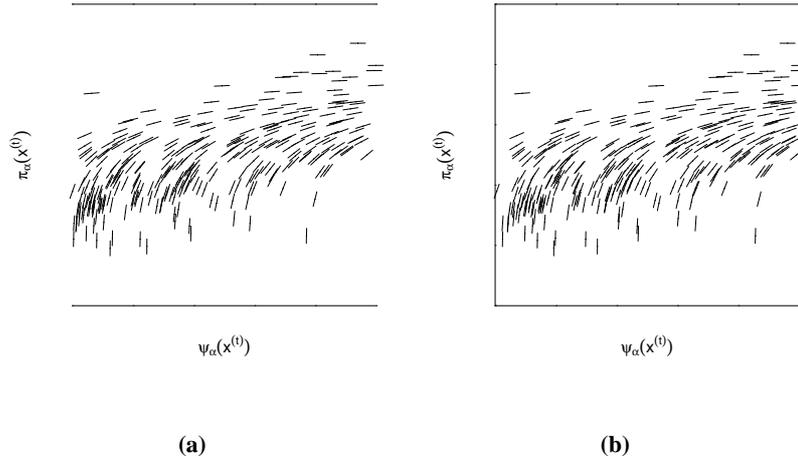


Figure 4: The local critique plots for α in the Pumps example where the $\psi_\alpha(x^{(t)})$ is (a) estimated using the numerical approach described in section 2.4 and (b) approximated by a Normal distribution ($M = 10000$, results are shown for a random subsample of size 300).

trigamma function. The local critique plot for α with $\psi_\alpha(x)$ estimated by this Normal approximation can be seen in Figure 4 (b). We see that the two local critique plots in (a) and (b) are very similar.

We also want to mention the possibility of multiplying $l_i(x)$ with a non-informative, proper prior (e.g. the Jeffreys prior). If this results in a (unnormalised) posterior density $\hat{l}_i(x)$ which can be integrated analytically, it can provide a reasonable approximation to $\psi_i(x)$. In addition to the advantage of $\hat{l}_i(x)$ being integrable, the use of the Jeffreys prior makes $\hat{\psi}_i(x)$ invariant to transformation, which $\psi_i(x)$ generally is not, as we show in the next section. We have not explored this idea; indeed, we have found it difficult to find non-trivial examples where the Jeffreys prior does result in an analytically-integrable $\hat{l}_i(x)$. Another possibility is to exploit the lack of invariance of $\psi_i(x)$ by transforming x_i in order to achieve an analytically tractable, integrable lifted likelihood.

2.5 Invariance

Our graphical tool is invariant in the π -direction to one-to-one transformations of x_i , but not always in the ψ -direction. Consider a one-to-one transformation $\tilde{x}_i = g(x_i)$ of $x_i, i \in U$. The other variables are not transformed, so $\tilde{x} = x^{i \rightarrow \tilde{x}_i}$. The full conditional distribution for the variable \tilde{x}_i is then

$$(20) \quad p(\tilde{x}_i | x_{-i}) \propto p(\tilde{x}_i | x_{\text{pa}(i)}, x_{\text{ne}(i)}) \prod_{c:i \in \text{pa}(V(c))} p(x_{V(c)} | \tilde{x}_{\text{pa}(V(c))}), \quad i \in V.$$

The model is the same as before the transformation, including the local prior on $x_i = g^{-1}(\tilde{x}_i)$, in the sense that the local prior on \tilde{x}_i is

$$(21) \quad \tilde{p}_i(\tilde{x}) = p(\tilde{x}_i | x_{\text{pa}(i)}, x_{\text{ne}(i)}) = p_i(x) \left| \frac{dg^{-1}(\tilde{x}_i)}{d\tilde{x}_i} \right|.$$

Because x_i is the only transformed variable, i.e. $\tilde{x}_{-i} = x_{-i}$, and because the model is the same as before the transformation, we have for the lifted likelihood

$$(22) \quad \tilde{l}_i(\tilde{x}) = \prod_{c:i \in \text{pa}(V(c))} p(x_{V(c)} | \tilde{x}_{\text{pa}(V(c))}) = \prod_{c:i \in \text{pa}(V(c))} p(x_{V(c)} | x_{\text{pa}(V(c))}) = l_i(x).$$

We suppose in the following, for simplicity, that g is increasing. Now, because $\pi_i(x)$ is equivalent to the cumulative distribution of the local prior of x_i , and because the local prior distribution is unchanged (see (21)), we have

$$(23) \quad \tilde{\pi}_i(\tilde{x}) = \int_{-\infty}^{\tilde{x}_i} \tilde{p}_i(x^{i \rightarrow \tilde{u}}) d\tilde{u} = \int_{-\infty}^{x_i} p_i(x^{i \rightarrow u}) \frac{dg^{-1}(\tilde{u})}{d\tilde{u}} \frac{dg(u)}{du} du = \int_{-\infty}^{x_i} p_i(x^{i \rightarrow u}) du = \pi_i(x).$$

i.e. $\tilde{\pi}_i(\tilde{x})$ is identical to $\pi_i(x)$ for g increasing, or the complement $1 - \pi_i(x)$ for g decreasing. Thus $\pi_i(x)$ is invariant to this type of transformation. When creating $\tilde{\psi}_i(\tilde{x})$ the integrand is the same as for $\psi_i(x)$ (see (22)) even though the variable to integrate over is transformed. We have

$$(24) \quad \tilde{\psi}_i(\tilde{x}) = \frac{\int_{-\infty}^{\tilde{x}_i} \tilde{l}_i(x^{i \rightarrow \tilde{u}}) d\tilde{u}}{\int_{-\infty}^{\infty} \tilde{l}_i(x^{i \rightarrow \tilde{u}}) d\tilde{u}} = \frac{\int_{-\infty}^{x_i} l_i(x^{i \rightarrow u}) \frac{dg(u)}{du} du}{\int_{-\infty}^{\infty} l_i(x^{i \rightarrow u}) \frac{dg(u)}{du} du}.$$

Hence, for linear transformations, $\tilde{\psi}_i(\tilde{x})$ is identical to $\psi_i(x)$ (see (12)), but for non-linear transformations it is generally not. Thus $\psi_i(x)$ is not invariant to non-linear one-to-one transformations. For $j \neq i$, the functions $\pi_j(x)$ and $\psi_j(x)$ are obviously unaltered by this kind of transformation of x_i .

The non-invariance of $\psi_i(x)$ is obviously unappealing because there is no "true" parametrisation, and thus no "true" $\psi_i(x)$. There is an analogous freedom of choice of focus parameters when defining p_D , the effective number of parameters in a model, and the resulting deviance information criterion (DIC), proposed by Spiegelhalter et al. (2002). Because of this lack of invariance, the local critique plot can potentially give a false warning about a lifted likelihood $l_i(x)$, or conceal a warning that should have been given about $l_i(x)$. But in the following example we see that relatively drastic transformations do not cause a great discrepancy in ψ . We return again to the Pumps example. Alternatively to the parametrisation used in (8), one could reparametrise using for example the two power transformations $\eta_i = \log(\theta_i)$ or $\nu_i = \frac{1}{\theta_i}$ instead of θ_i . All three parametrisations are natural choices: θ_i is the mean value parameter, η_i is the canonical parameter and ν_i is the mean time to event parameter. The two alternative models are

$$(25) \quad \begin{aligned} y_i | \eta_i &\sim \text{Poisson}(\exp(\eta_i) t_i), \quad i = 1, \dots, n \\ \exp(\eta_i) | \alpha, \beta &\sim \text{Ga}(\alpha, \beta), \quad i = 1, \dots, n \end{aligned}$$

and

$$(26) \quad \begin{aligned} y_i | \nu_i &\sim \text{Poisson}\left(\frac{t_i}{\nu_i}\right), \quad i = 1, \dots, n \\ \frac{1}{\nu_i} | \alpha, \beta &\sim \text{Ga}(\alpha, \beta), \quad i = 1, \dots, n, \end{aligned}$$

with the same distributions for α and β as in parametrisation (8). Let $\tilde{x} = x^{n+i \rightarrow \eta_i}$ and $\dot{x} = x^{n+i \rightarrow \nu_i}$. We now have

$$(27) \quad \begin{aligned} \tilde{\pi}_{\eta_i}(\tilde{x}) &= \Gamma(\exp(\eta_i); \alpha, \beta) = \Gamma(\theta_i; \alpha, \beta) = \pi_{\theta_i}(x) \\ \dot{\pi}_{\nu_i}(\dot{x}) &= 1 - \Gamma\left(\frac{1}{\nu_i}; \alpha, \beta\right) = 1 - \Gamma(\theta_i; \alpha, \beta) = 1 - \pi_{\theta_i}(x). \end{aligned}$$

For parametrisation (8) the lifted likelihood of θ_i (see (10)) is proportional to a $\text{Ga}(y_i + 1, t_i)$ density for θ_i , so

$$(28) \quad \psi_{\theta_i}(x) = \Gamma(\theta_i; y_i + 1, t_i).$$

For parametrisations (25) and (26) the lifted likelihoods of η_i and ν_i are

$$(29) \quad \begin{aligned} \tilde{l}_{\eta_i}(\tilde{x}) &\propto \exp(\eta_i y_i - t_i \exp(\eta_i)) \\ \dot{l}_{\nu_i}(\dot{x}) &\propto \left(\frac{1}{\nu_i}\right)^{y_i} \exp\left(-\frac{t_i}{\nu_i}\right), \end{aligned}$$

which means that $\tilde{\psi}_{\eta_i}(\tilde{x}) = \Gamma(\exp(\eta_i); y_i, t_i)$ for $y_i > 0$ and $\dot{\psi}_{\nu_i}(\dot{x}) = \Gamma(\frac{1}{\nu_i}; y_i - 1, t_i)$ for $y_i > 1$, and neither $\tilde{\psi}_{\eta_i}(\tilde{x})$ nor $\dot{\psi}_{\nu_i}(\dot{x})$ are the same as $\psi_{\theta_i}(x)$ (or $1 - \psi_{\theta_i}(x)$). Hence $\psi_i(x)$ is not invariant to the transformation from θ_i to η_i or ν_i . Figure 5 shows the local critique plots for (a) θ_i , (b) η_i and (c) ν_i .

The functions $\tilde{l}_{\eta_i}(\tilde{x})$ for $y_i = 0$ and $\dot{l}_{\nu_i}(\dot{x})$ for $y_i < 2$ are not integrable. Because the pumps $i = 2, 7$ and 8 have $y_i = 1$, we used the numerical approach proposed in the previous subsection (with $m = 60$ and the Trapezoid Rule on 1000 subintervals) for estimating $\dot{\psi}_{\nu_i}(\dot{x})$, $i = 2, 7, 8$. These three cases with the lowest failure counts are where we see the biggest differences between $\psi_{\theta_i}(x)$, $\tilde{\psi}_{\eta_i}(\tilde{x})$ and $\dot{\psi}_{\nu_i}(\dot{x})$, with $\dot{\psi}_{\nu_i}(\dot{x})$ for $i = 2$ being the most divergent. Pump 2 is an unusual data point with only one failure despite a relatively long operation time. However, for most i there are actually no substantial visible distinctions between $\psi_{\theta_i}(x)$, $\tilde{\psi}_{\eta_i}(\tilde{x})$ and $\dot{\psi}_{\nu_i}(\dot{x})$, and we draw the same conclusions regarding the lifted likelihoods for the different parametrisations for these pumps. Considering the fact that the posterior mean estimates of the θ_i 's range from 0.06 to 1.98, we have here explored relatively drastic transformations $\eta_i = \log \theta_i$ and $\nu_i = \frac{1}{\theta_i}$. The example therefore reassures us that $\psi_i(x)$ does not seem to be very sensitive to transformations.

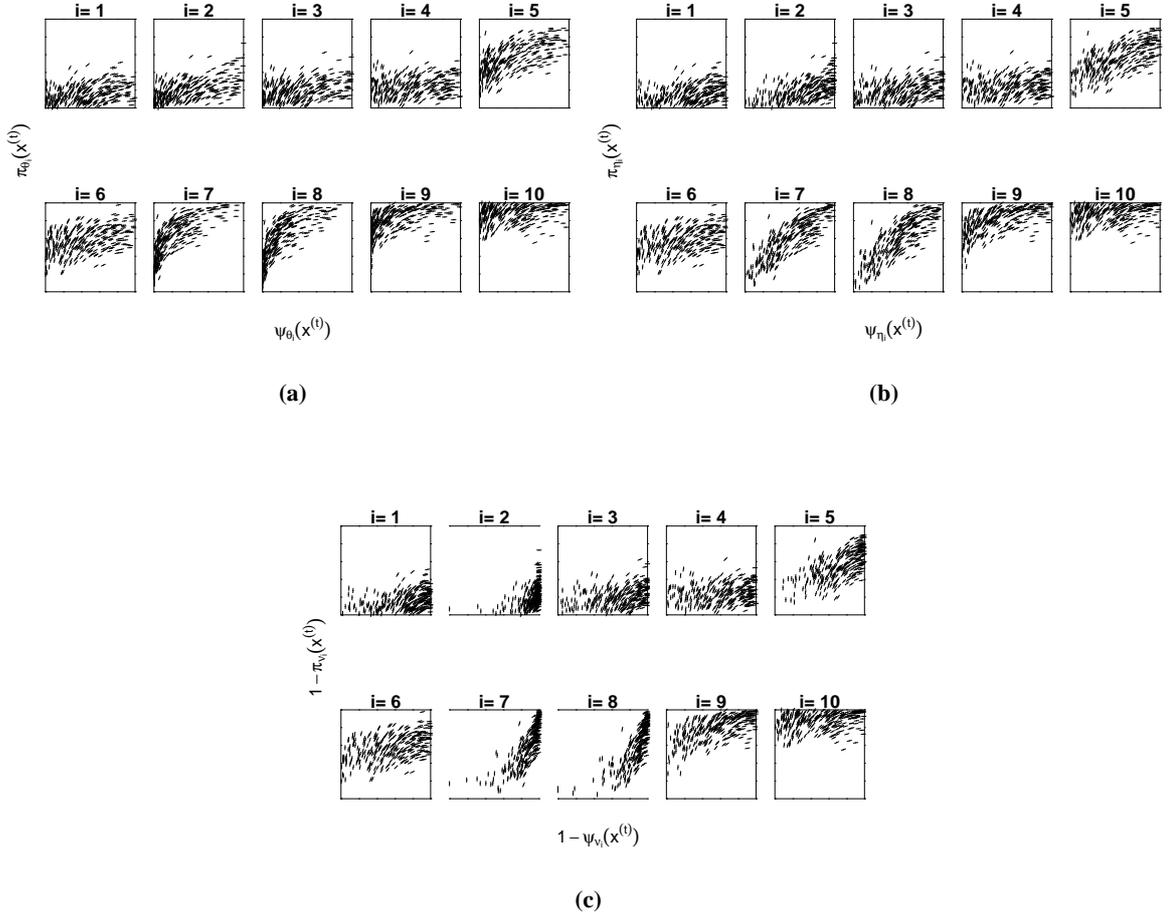


Figure 5: The local critique plots for (a) θ_i in parametrisation (8), (b) η_i in parametrisation (25) and (c) ν_i in parametrisation (26) for the Pumps example ($M = 10000$, results are shown for a random subsample of size 300).

3 Applications

3.1 3-level DAG: Rats

This application illustrates how the local critique plot can reveal local prior or lifted likelihood domination, and identify situations where the posterior is a trade off of tail specifications. It shows local critique plots for the parameters in a 3-level DAG, and is based on the `Rats` example in Spiegelhalter et al. (2004). The weights of $N = 30$ rats were measured weekly over $T = 5$ weeks. Rat i has weight y_{ij} at day t_j . The model is

$$(30) \quad \begin{aligned} y_{ij} | \alpha_i, \beta_i, \sigma_c &\sim \text{N}(\alpha_i + \beta_i(t_j - \bar{t}), \frac{1}{\sigma_c^2}), \quad i = 1, \dots, N, \quad j = 1, \dots, T \\ \alpha_i | \alpha_c, \sigma_\alpha &\sim \text{N}(\alpha_c, \frac{1}{\sigma_\alpha^2}), \quad i = 1, \dots, N \\ \beta_i | \beta_c, \sigma_\beta &\sim \text{N}(\beta_c, \frac{1}{\sigma_\beta^2}), \quad i = 1, \dots, N. \end{aligned}$$

The DAG can be seen in Figure 6. Instead of using the prior distributions from Spiegelhalter et al. (2004), we invent a biologist with strong opinions about what the priors should be. He suggests

$$(31) \quad \begin{aligned} \sigma_c &\sim \text{Unif}(0, U_{\sigma_c}) \\ \alpha_c &\sim \text{N}(\mu_{\alpha_c}, \tau_{\alpha_c}) \\ \sigma_\alpha &\sim \text{Unif}(0, U_{\sigma_\alpha}) \\ \beta_c &\sim \text{N}(\mu_{\beta_c}, \tau_{\beta_c}) \\ \sigma_\beta &\sim \text{Unif}(0, U_{\sigma_\beta}), \end{aligned}$$

with $\mu_{\alpha_c} = 250$, $\mu_{\beta_c} = 1$, $\tau_{\alpha_c} = \tau_{\beta_c} = 1$ and $U_{\sigma_c} = U_{\sigma_\alpha} = U_{\sigma_\beta} = 10$. Let $\tau_c = \frac{1}{\sigma_c^2}$, $\tau_\alpha = \frac{1}{\sigma_\alpha^2}$ and $\tau_\beta = \frac{1}{\sigma_\beta^2}$. It is more convenient to derive ψ functions for the precisions rather than for the standard deviations. A uniform distribution $\text{Unif}(0, U)$ on the standard deviation σ is equivalent to the precision $\tau = \frac{1}{\sigma^2}$ having the cumulative distribution function $F(\tau; U) = 1 - \frac{1}{U}\tau^{-1/2}$, $\tau \in [U^{-2}, \infty)$. The π and ψ functions can be seen in Appendix A.1. The corresponding local critique plots can be seen in Figure 7 and Figure 8.

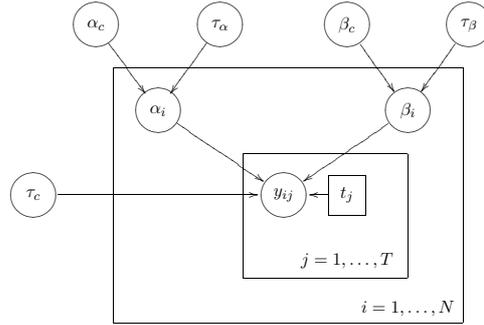


Figure 6: DAG for the `Rats` model.

Many of the local critique plots draw attention to possible conflicts in assumptions. The strongest warning is given by Figure 7 (c), which shows a compromise that imply extreme influence of the tails of both the local prior and lifted likelihood. The marginal posterior distribution for τ_α is a trade off of the outer left tail of the local prior and the outer right tail of the lifted likelihood. In Figure 7 (b), the local prior of α_c is constraining the marginal posterior, and only a very small part of the outer right tail of the lifted likelihood

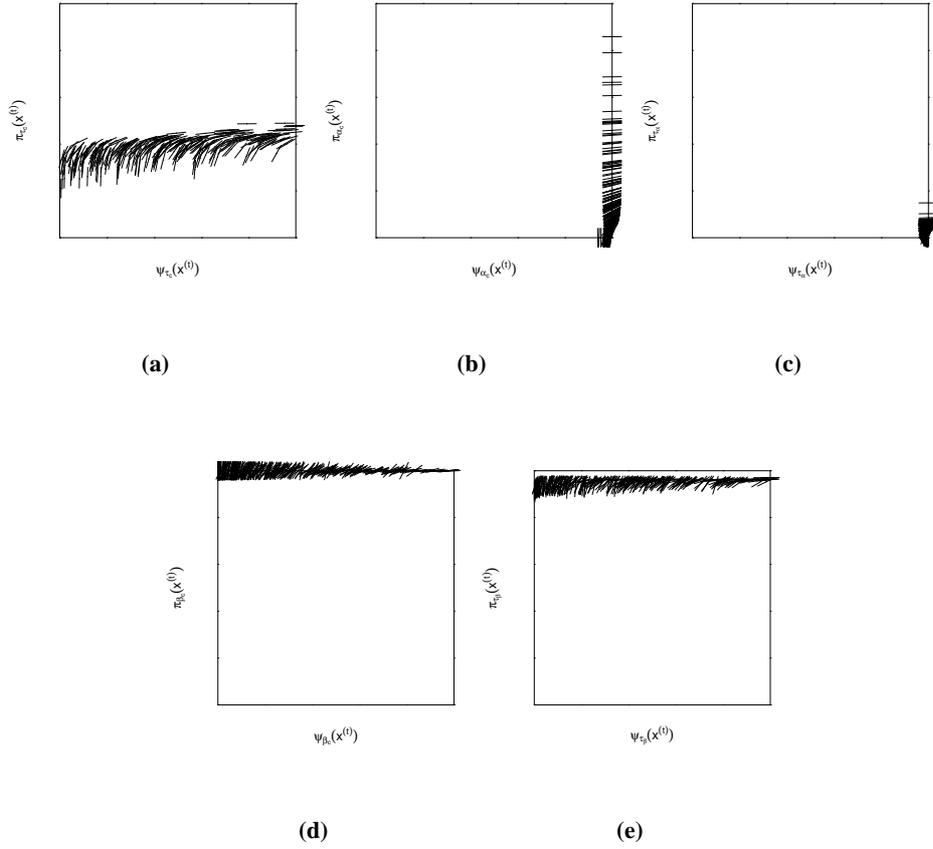
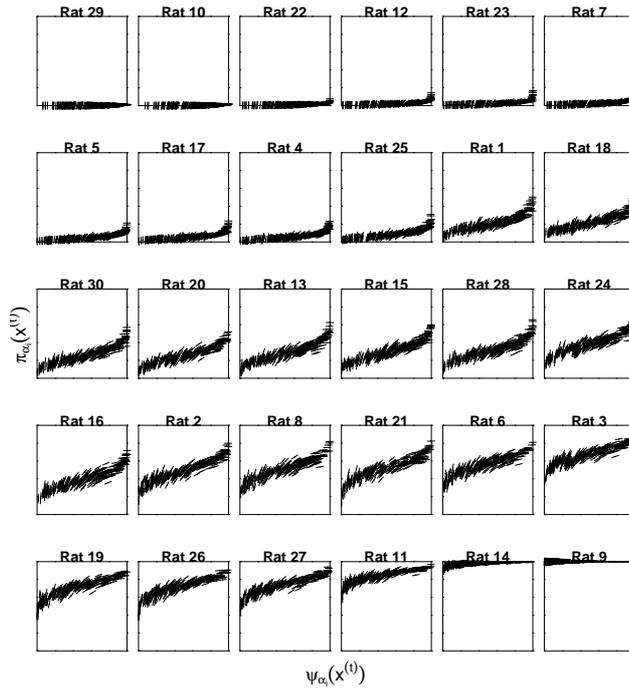
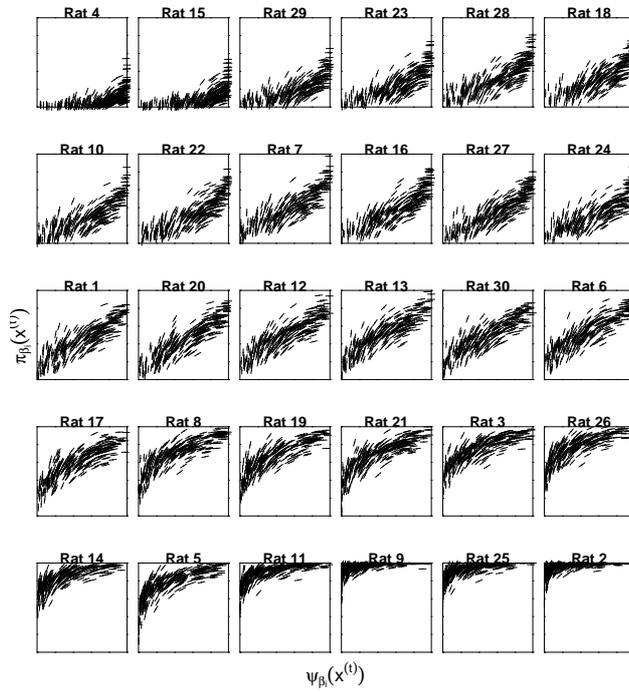


Figure 7: The local critique plots for (a) τ_c , (b) α_c , (c) τ_α , (d) β_c and (e) τ_β for Rats ($M = 10000$, results are shown for a random subsample of size 300).

is used, which generally is a very unsatisfying situation. The local critique plots in Figure 7 (d) and (e) show cases where the lifted likelihoods are constraining the marginal posteriors, and only a small part of the outer right tail of their respective local priors are used. Figure 7 (a) also shows a case where the marginal posterior is dependent on only a small part of the local prior, but here the part of the local prior being used is closer to the centre. It is common, and often desired, for the likelihood to dominate the posterior. However, if we have more justified specifications for the centres of the local priors than for the tails, we may be happy with a local critique plot similar to the one in Figure 7 (a), while a plot similar to those shown in Figure 7 (d) and (e) attracts attention. The local critique plots seen in Figure 8 (a) show that the marginal posteriors for many of the α_i 's are dependent only on the outer left or right tail of their local prior. This mostly concerns the rats with low average weights, for which the corresponding α_i posterior samples are located only in the outer left tail of the local prior. However, most extreme is the plot for the rat with the highest average weight (rat 9), which shows that in addition to being dependent only on the outer right tail of the local prior, the marginal posterior of α_9 depends heavily on the left tail of the lifted likelihood. The same type of situation, to a slightly less serious extent, can be seen in Figure 8 (b) for the β_i 's of the rats with the highest weight gains. For the rats with the lowest weight gains, we can see the opposite situation; the posterior samples of the corresponding β_i 's are located only in the left tail of the local prior and to a large extent in the right tail of the lifted likelihood. Generally, the local critique plots in Figure 8 (b) show that the posterior samples of the β_i 's are more spread out across their local priors than is the case for the α_i 's in Figure 8 (a).



(a)



(b)

Figure 8: The local critique plot for (a) α_i and (b) β_i for Rats. The α_i plots are sorted by increasing order of the average rat weights $\bar{y}_{i\cdot}$, while the β_i plots are sorted by the increasing order of the weight gains $y_{i5} - y_{i1}$ ($M = 10000$, results are shown for a random subsample of size 300).

3.2 2-level DAG: Poisson-Gamma spatial moving average model

This application shows how diagnostic warnings from the local critique plots prompt us to reconsider the statistical model, in particular the way that information from the data is distributed spatially. The application is taken from Best et al. (2000). They used a Poisson-Gamma spatial moving average model to assess the association between exposure to NO₂ and potentially unobserved spatial factors, and the rate of respiratory illness in children in the Huddersfield region in England. The data is from the European multicentre study SAVIAH. The Huddersfield region was partitioned into a grid of $I = 605$ cells A_i of $750\text{m} \times 750\text{m}$ each, to which the disease counts were aggregated (y_i) and long-term average population (N_i , in hundreds) and excess NO₂ concentration (Z_i) estimated. Another partition was done of a larger rectangle that covers all of Huddersfield as well as a surrounding buffer zone of 2km into $J = 96$ areas B_j of $3\text{km} \times 3\text{km}$, representing areas associated with latent spatial factors. The model Best et al. (2000) used is

$$(32) \quad \begin{aligned} y_i &\sim \text{Poisson}(N_i \cdot (\beta_0 + \beta_1 Z_i + \beta_2 \sum_j k_{ij} \gamma_j)), \quad i = 1, \dots, I \\ \beta_0 &\sim \text{Ga}(\alpha_0, \tau_0) \\ \beta_1 &\sim \text{Ga}(\alpha_1, \tau_1) \\ \beta_2 &\sim \text{Ga}(\alpha_2, \tau_2) \\ \gamma_j &\sim \text{Ga}(\alpha_\gamma, \tau_\gamma), \quad j = 1, \dots, J. \end{aligned}$$

The DAG for this model can be seen in Figure 9.

We use here the k_{ij} matrix from Thomas et al. (2004, Example 6), which is of the form

$$(33) \quad k_{ij} = \frac{1}{2\pi\rho^2|B_j|} \int_{B_j} \exp\left(-\frac{|a_i - b_j|^2}{2\rho^2}\right) db_j,$$

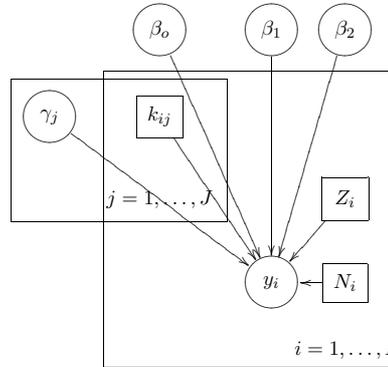


Figure 9: DAG for the SAVIAH model.

where a_i is the centre of cell A_i , $|a_i - b_j|$ is the Euclidean distance between a_i and the location b_j within the latent risk area B_j , and $|B_j|$ is the area of B_j . Here, k_{ij} is a kernel function integrated over all distances between the centre of cell A_i and all locations in the latent risk area B_j , divided by the area of B_j . In Best et al. (2000), k_{ij} is simply the kernel evaluated for the distance between the centre of A_i and the centre of B_j . For computational reasons, Best et al. (2000) fix the value of the scale ρ . They experimented with several different values in the range 0-15 km, but settled on $\rho \equiv 1$ km, which they found to be most consistent with the data. In order to avoid aggregation inconsistencies, Best et al. (2000) chose the identity link, which

gives a linear dose-response relationship.

The model given by (32) has the same joint distribution of data and parameters as the model

$$\begin{aligned}
(34) \quad & y_{i(J+1)} \sim \text{Poisson}(N_i\beta_0), \quad i = 1, \dots, I \\
& y_{i(J+2)} \sim \text{Poisson}(N_i\beta_1 Z_i), \quad i = 1, \dots, I \\
& y_{ij} \sim \text{Poisson}(N_i\beta_2 k_{ij}\gamma_j), \quad i = 1, \dots, I, \quad j = 1, \dots, J \\
& y_i = \sum_{j=0}^{J+2} y_{ij},
\end{aligned}$$

where $\{y_{ij}\}$ are augmented data. The local priors for β_0 , β_1 , β_2 and $\{\gamma_j\}_{j=1}^J$ are the same as in (32). This second way of expressing the model is useful when setting up the full conditionals. The π and ψ functions for the parameters can be seen in Appendix A.2.

The fixed parameters of the prior distributions of β_0 , β_1 and β_2 were chosen by Best et al. (2000) so that there was 80% prior probability that the disease counts attributed to each of three risk factors (baseline β_0 , NO2 related $\beta_1 Z_i$, unobserved spatial $\beta_2 \sum_j k_{ij}\gamma_j$) would lie between one tenth and ten times a nominal equal attribution. The choice of $\alpha_0 = \alpha_1 = \alpha_2 = 0.575$ gives a ratio of 100 for the 90th and 10th percentiles, while $\tau_0 = \frac{3\alpha_0}{\bar{Y}}$, $\tau_1 = \frac{3\alpha_1 \bar{Z}}{\bar{Y}}$, $\tau_2 = \frac{3\alpha_2}{\bar{Y}}$ in addition lead to prior means for β_0 , β_1 and β_2 corresponding to each of the three risk categories contributing with one third of the overall disease rate $\bar{Y} = \frac{\sum_i y_i}{\sum_i N_i}$. \bar{Z} is the population weighted average excess NO2 concentration. The parameters of the prior distributions of $\{\gamma_j\}_{j=1, \dots, J}$ were fixed so that the prior mean of γ_j would be $|B_j|$ and the prior variance reflected the prior belief of moderate spatial variability. This was achieved by setting $\alpha_\gamma = |B_j|/\text{km}^2$ and $\tau_\gamma = 1/\text{km}^2$. The data we have used (taken from Thomas et al., 2004, Example 6) are randomly perturbed compared to the ones used in Best et al. (2000). The local critique plots can be seen in Figures 10 and 11. We see that the samples of β_0 , β_1 and β_2 are distributed well across their respective lifted likelihoods. The marginal posterior distributions of β_0 and β_1 are using almost all of the local priors, except for the right-hand tails. The marginal posterior distribution of β_2 is using almost exclusively the right-hand tail of its local prior.

The local critique plots of the γ_j 's in Figure 11 are laid out in accordance with the location of the corresponding latent risk areas B_j , $j = 1, \dots, J$. The plots for the γ_j 's close to the edges give warnings about their lifted likelihoods. All of the posterior samples of these γ_j 's are concentrated in a small part of the left tail of their lifted likelihoods. An examination of $\psi_{\gamma_j}(x)$ and the data reveals that this is due to the fact that the lifted likelihoods for these γ_j 's are extremely vague. This is caused by very small values of $k_{ij}, \forall i$ for these latent risk areas and thus small values of $\sum_{i=1}^I N_i k_{ij}$. The reason for the small k_{ij} -values is that the correlation length $\rho = 1$ km is relatively short, and that the peripheral latent risk areas are actually partially or completely outside the Huddersfield region, which has an irregular shape. Hence, there is almost no information in the data about these γ_j 's, and the posterior samples are dominated by the local prior information. We verified this explanation by experimenting with a longer correlation length, namely $\rho = 15$ km. This alternative k_{ij} matrix resulted in local critique plots for the γ_j 's with good spread across the whole plotting regions for all j (plots not shown), illustrating that a long correlation length distributes the information in the data to all the latent risk areas. Hence, Figure 11 is an example of local critique plots that give warnings which make us reconsider the modelling, but where a reexamination of the problem reveals that there is a reasonable explanation consistent with the model and data.

Many of the non-peripheral γ_j 's have local critique plots that cover the whole of the plotting region, i.e. all of the local priors and lifted likelihoods are being used. Some of the non-peripheral plots are concentrated

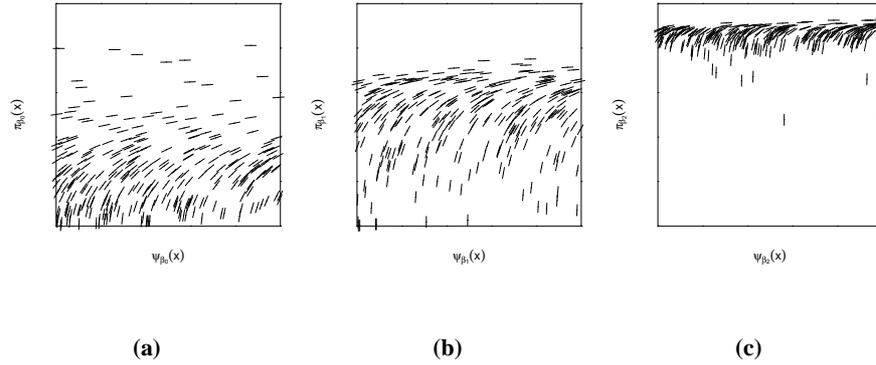


Figure 10: The local critique plots for (a) β_0 , (b) β_1 and (c) β_2 ($M = 20000$, results are shown for a random subsample of size 300).

in the upper-left (e.g. and $j = 20$ and $j = 62$) or the lower-right corner (e.g. $j = 69$). It seems that the local priors of γ_{20} and γ_{62} are restricting their posterior samples to be of lower values than the lifted likelihoods suggest. They have the highest posterior mean and median of all the γ_j 's. Conversely, for γ_{69} it seems that the local prior is restricting the posterior samples to be higher than the lifted likelihood suggest. γ_{69} has the lowest posterior mean and median of all the γ_j 's. The "gaps" in some of the local critique plots in Figure 11 are due to the fact that the augmented data y_{ij} are discrete.

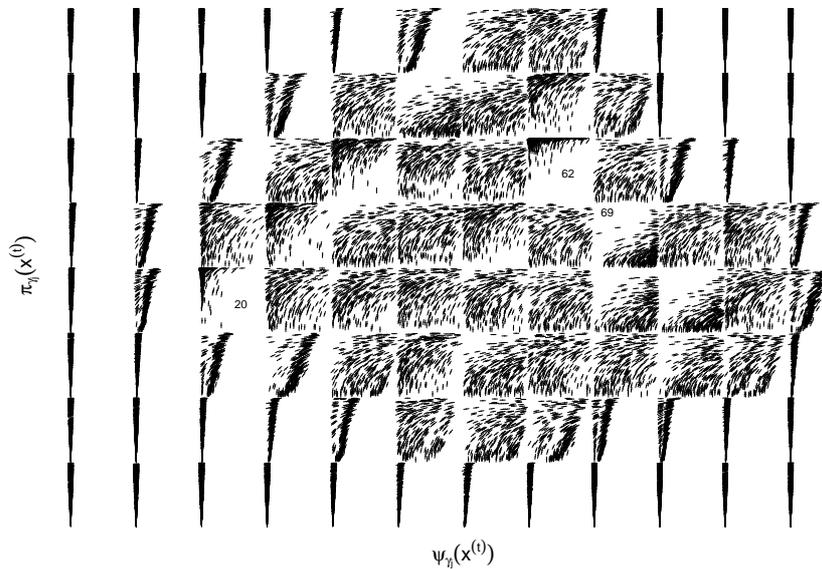


Figure 11: The local critique plots for γ_j ($M = 20000$, results are shown for a random subsample of size 300). The plots for the latent risk areas are laid out according to the respective locations.

3.3 DAG combined with MRF: Larynx cancer

In a Markov Random Field (MRF), neighbours influence the posterior distribution of a variable through the local prior. This application shows how the local critique plot is able to identify conflict between the local prior and the lifted likelihood of a variable, which partly can be traced back to the information provided by the neighbours of the variable. We apply a modification of the model introduced in Besag et al. (1991) for disease mapping. They proposed a model involving a Gaussian Intrinsic Conditional Auto Regression (CAR) prior for spatial random effects combined with an unstructured Normal prior for independent random effects. Green and Richardson (2002) applied a slightly altered version of this model for data on larynx cancer mortality in France for the period 1986-1993. We analyse the same data, which were taken from Rezvani et al. (1997). Initially, we apply the following model

$$\begin{aligned}
 (35) \quad & y_i | c, u_i, v_i \sim \text{Poisson} \left(\exp(c + u_i + v_i) E_i \right), \quad i = 1, \dots, I \\
 & u_i | \tau_u, u_{-i} \sim N(\bar{u}_i, n_i \tau_u), \quad i = 1, \dots, I \\
 & v_i | \tau_v \sim N(0, \tau_v), \quad i = 1, \dots, I \\
 & c \sim \text{Uniform}(-\infty, \infty) \\
 & \tau_u \sim \text{Ga}(\alpha_u, \beta_u) \\
 & \tau_v \sim \text{Ga}(\alpha_v, \beta_v),
 \end{aligned}$$

where y_i is the observed number of disease cases, E_i the (known) estimated expected count based on the population size in area i , n_i the number of neighbours of region i and $\bar{u}_i = \frac{1}{n_i} \sum_{j \in \text{ne}(i)} u_j$. The graph for this model can be seen in Figure 12. We use WinBUGS and its distribution `car.normal` (as prior on u_i) to perform the posterior MCMC simulations. Because `car.normal` imposes the constraint $\sum_{i=1}^I u_i = 0$, it makes sense to include the intercept variable c . Using `car.normal` requires an improper uniform prior on c . We chose $\alpha_u = \beta_u = \alpha_v = \beta_v = 0.01$, as proposed in the GeoBUGS manual (Thomas et al., 2004).

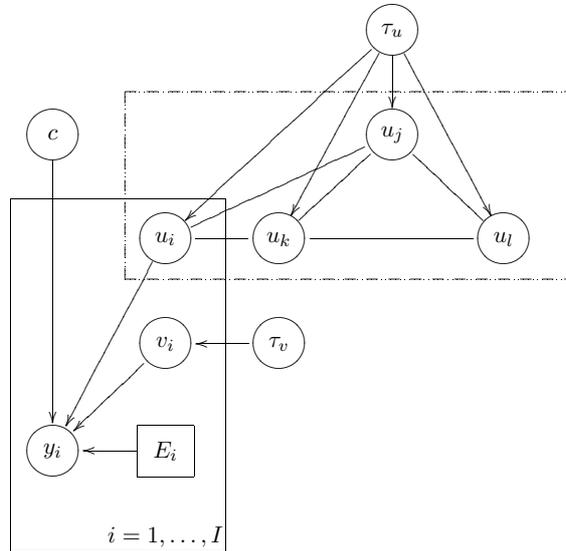


Figure 12: The graph for the Larynx model.

Because the intercept c has an improper prior and hence $\pi_c(x)$ does not exist, c is not of interest in this context. In the following we therefore disregard it. As shown in Section 2.5, the parametrisation in the

distribution for y_i makes the lifted likelihoods of u_i and v_i non-integrable for $y_i = 0$. Instead of u_i and v_i we therefore use the parameters $a_i = \exp(u_i)$ and $b_i = \exp(v_i)$, which then have log-Normal distributions

$$(36) \quad \begin{aligned} y_i | c, a_i, b_i &\sim \text{Poisson}(\exp(c) a_i b_i E_i), \quad i = 1, \dots, I \\ a_i | \tau_u, a_{-i} &\sim \text{logN}(\bar{u}_i, n_i \tau_u), \quad i = 1, \dots, I \\ b_i | \tau_v &\sim \text{logN}(0, \tau_v), \quad i = 1, \dots, I. \end{aligned}$$

The π and ψ functions can be seen in Appendix A.3. The local critique plots for τ_u and τ_v for this model (Larynx1) can be seen in Figure 13.

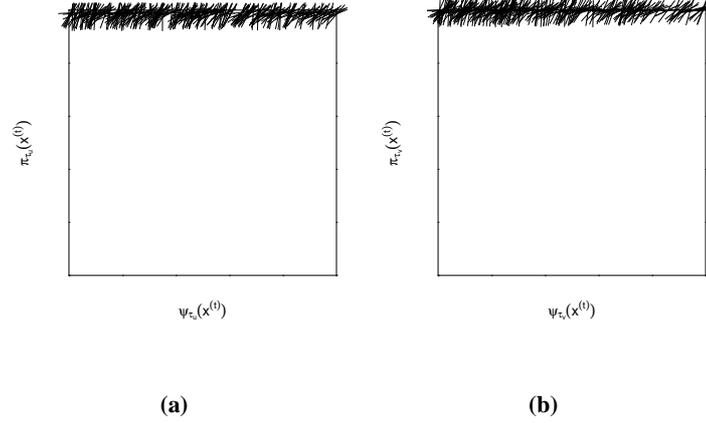


Figure 13: The local critique plots for (a) τ_u and (b) τ_v for Larynx1 ($M = 10000$, results are shown for a random subsample of size 300).

The posterior samples of τ_u and τ_v are pressing against the upper edge of the plots, i.e. they are located in the right hand tail of their prior specifications. This reflects that the marginal posterior distributions for τ_u and τ_v are using only small parts of the respective outer tails of the non-informative local priors. Because the local prior distributions are relatively non-informative, and because we see that the samples are well spread out across the lifted likelihoods, this may well be reasonable.

The local critique plots for the b_i 's show approximately the same features as the local critique plots for the a_i 's, we therefore show only the latter. They can be seen in Figure 14, laid out according to the location of the regions. Some regions are moved slightly in order for all the plots to be visible. Region 20 (Corse) has no neighbours, hence $u_{20} \equiv 0$ and the local critique plot collapses to a single point, therefore Corse is not included in the plot. The values of y_i/E_i can be seen in Table 1 in Appendix A.3. We see that regions with low values of y_i/E_i have local critique plots for a_i with mass gathered towards the lower right corner. Conversely regions with high values of y_i/E_i have local critique plots with mass gathered towards the upper left corner. The strength of these effects are affected by the y_i/E_i values of the neighbours. Consider for example the regions 63 (Puy-de-Dôme), 75 (Paris) and 94 (Val-de-Marne). They have approximately the same y_i/E_i values (1.3), but the a_i local critique plot for Puy-de-Dôme shows that the posterior samples are located far out in the left hand tail of the lifted likelihood, while Paris and Val-de-Marne have posterior samples which cover almost all of the respective lifted likelihoods. The neighbours of Puy-de-Dôme all have y_i/E_i values below 1, while all but one of the neighbours of Paris and Val-de-Marne have values greater than or equal to 1.3. Looking at the regions 50 (Manche) and 51 (Marne) we see that the respective a_i local critique plots are very similar, but Manche has a much lower y_i/E_i value (0.08) than Marne (0.34). This

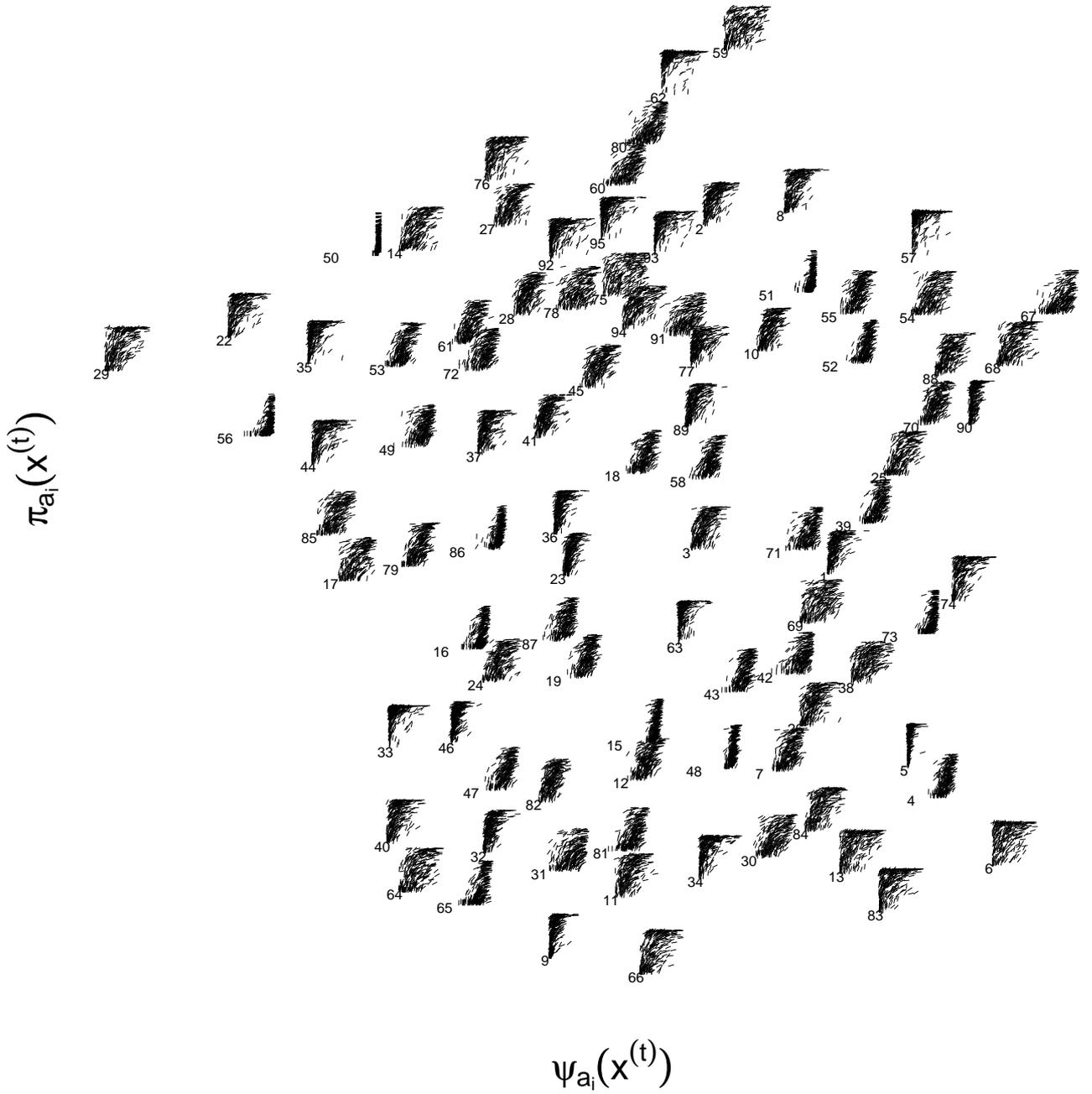


Figure 14: The local critique plot for a_i for Larynx1 ($M = 10000$, results are shown for a random subsample of size 300).

is also due to different neighbour behaviour, the neighbours of Manche have generally lower y_i/E_i values than those of Marne.

The strength of the spatial dependence is controlled by the precision τ_u . The larger the precision τ_u , the greater the spatial dependence and the more the neighbours of region i , through the local prior of a_i , influence the posterior samples of a_i . Thus, for models that lead to smaller posterior values of τ_u than model `Larynx1`, we expect less regions to have local critique plots with gatherings in the upper left or lower right corners than what is the case in Figure 14. We illustrate this in the following subsection by considering an alternative local prior specification for τ_u , which is more informative and encourage τ_u to be smaller, hence implying less spatial dependence. We also consider an alternative local prior for τ_v .

3.3.1 Larynx2

The alternative prior choices are

$$(37) \quad \begin{aligned} \tau_u &\sim \log N(\mu, \rho) \\ \tau_v &\sim \log N(\mu, \rho), \end{aligned}$$

where $\mu = 0$ and $\rho = 5$. We now have new π functions for τ_u and τ_v (see Appendix A.3), while ψ_{τ_u} and ψ_{τ_v} , as well as the π and ψ functions for the other parameters, remain unchanged. The local critique plots for τ_u and τ_v for this model can be seen in Figure 15, and the local critique plots for the a_i 's can be seen in Figure 16.

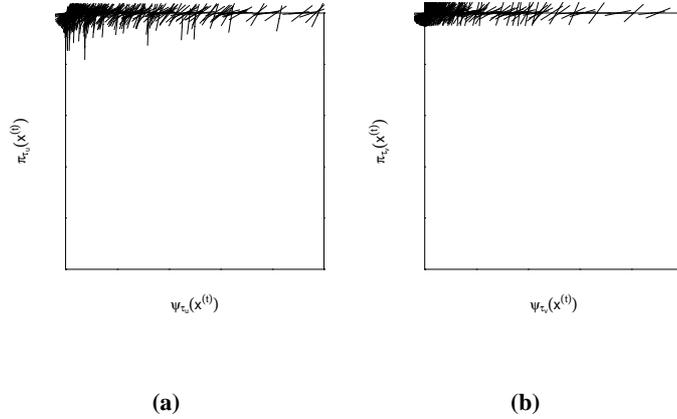


Figure 15: The local critique plots for (a) τ_u and (b) τ_v for `Larynx2` ($M = 10000$, results are shown for a random subsample of size 300).

The marginal posteriors of τ_u and τ_v are also now using only the outer right tails of their local priors, but in addition we see that the samples now are more concentrated towards the left tails of their lifted likelihoods than what was the case for `Larynx1` (Figure 13). This can be explained by the fact that we in `Larynx2` are imposing more informative local priors on τ_u and τ_v than in `Larynx1`. This causes less of the lifted likelihood information to be used. At the same time the marginal posteriors depend only on the outer right tail of the informative local priors. Hence, the local critique plots in Figure 15 show that there seems to be reason to reconsider the modelling in `Larynx2`.

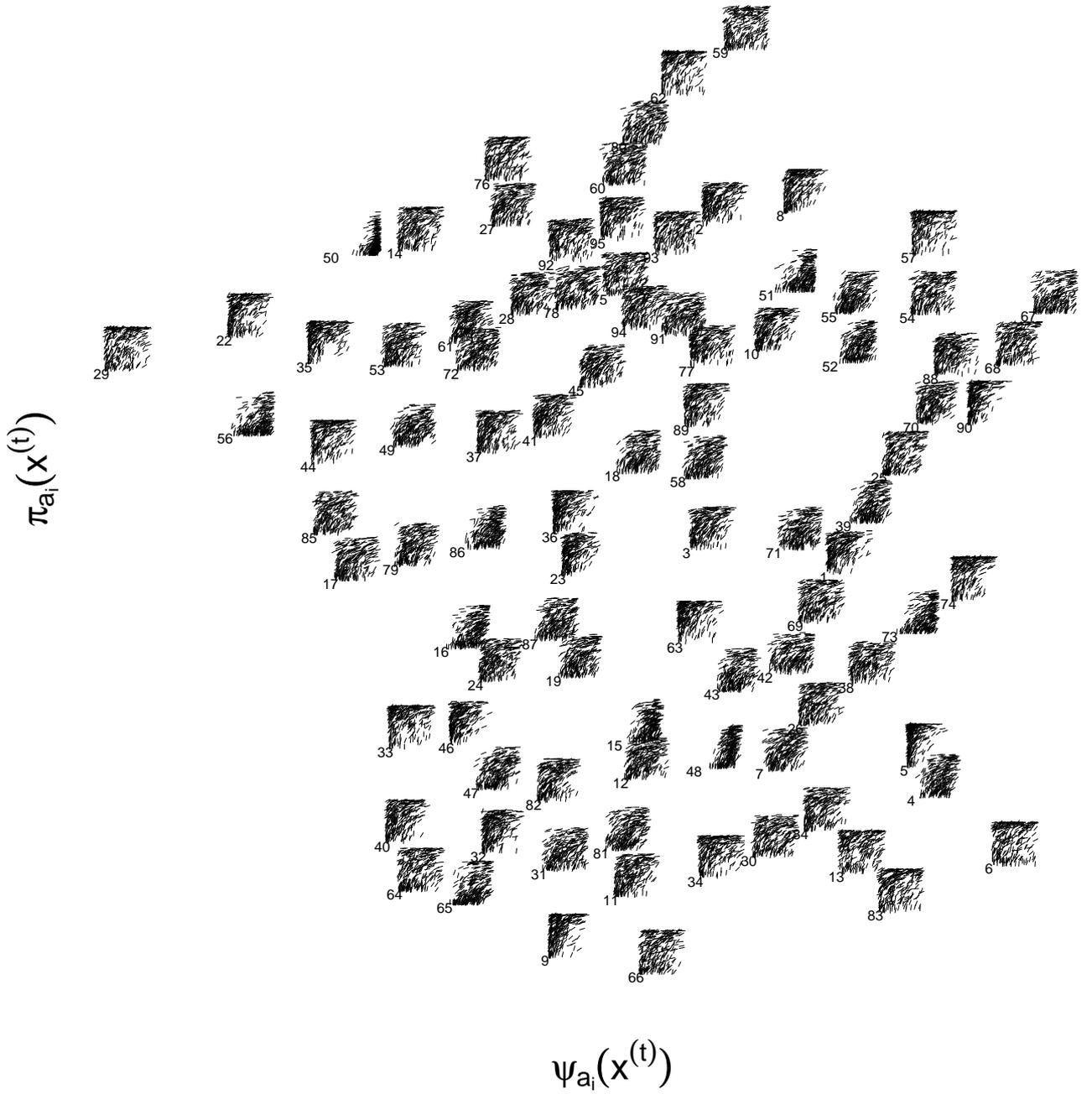


Figure 16: The local critique plot for a_i for Larynx2 ($M = 10000$, results are shown for a random subsample of size 300).

As expected, since `Larynx2` leads to smaller posterior values of τ_u than `Larynx1`, and hence less spatial dependence, a general distinction between Figure 16 and Figure 14 is that for `Larynx2` (Figure 16), the local priors are not constraining the posterior samples as much as for `Larynx1` (Figure 14), and most of the local priors and lifted likelihoods of the a_i 's are being used.

4 Conclusions

In this paper we have proposed a graphical diagnostic for chain graph models called the local critique plot. It investigates conflict locally at the node level between the information coming from the parents and neighbours (through the local prior) and the information coming from the children and co-parents (through the lifted likelihood). By visualising the link between the local prior, the lifted likelihood and the marginal posterior distribution of the parameter, the local critique plot is able to diagnose local conflict between the different sources of information. Specific, local choices made by the modeller that are influential for the posterior analysis of that particular parameter can be identified.

Through applications involving pure DAG models and one application combining a DAG with a MRF, we have illustrated the use and various features of the local critique plot. For example, we have seen situations where the local critique plot reveals that the marginal posterior distribution of a parameter is dominated by the local prior, and hence only a small part of the lifted likelihood is used. In the more common, but generally less problematic, case of lifted likelihood domination, the local critique plot identifies situations where the local prior should be checked for mis-specification, e.g. because only a small part of one of the tails is used. We have also seen situations where the posterior is a result of a trade off of tail specifications.

The lifted likelihood dimension in the local critique plot is not invariant to transformation of the parameter, but the lack of invariance does not seem to be too severe. The applications have shown cases where the local critique plot gives a warning about specific modelling choices, but where after a closer examination of model and data the choices in question are justified. We stress that a warning given by the local critique plot is not saying that the current model is wrong or proposing a better model, but rather that the modeller should re-examine his or her choices to make sure that they are justifiable.

For lifted likelihoods that are not integrable or do not have analytically tractable integrals, we have proposed a numerical integration scheme that is computationally fast, and thus suitable for performing at each simulation step. Hence, the local critique plot can be derived for all parameters in Bayesian hierarchical models characterised by chain graphs, and is easy to implement as a by-product of a posterior simulation, e.g. MCMC. It is not, however, a diagnostic of the coding of the posterior simulation.

The local critique plot is an informal graphical diagnostic, intended for a visual examination of possible conflict. Future extensions of this diagnostic may include theory on formal evaluations of specific features, with the specification of suitable discrepancy measures guided by the local critique plot. When performing the corresponding formal tests, there are some possible pitfalls, e.g. it is important to avoid double use of data, as is the case for the posterior predictive p -values. Also, potential multiple testing issues must be handled carefully.

Acknowledgements

The authors are very grateful to Arnaldo Frigessi, Jørund Gåsemyr, Bent Natvig and Sylvia Richardson for helpful comments and discussions.

References

- Bayarri, M. J. and Berger, J. O. (1999). Quantifying surprise in the data and model verification. In Bernardo, M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 6*, pages 53–82. Oxford University Press.
- Bayarri, M. J. and Berger, J. O. (2000). p -values for composite null models (with discussion). *J. Amer. Statist. Assoc.*, 95:1127–1142.
- Bayarri, M. J. and Castellanos, M. E. (2007). Bayesian checking of the second levels of hierarchical models. *Statist. Sci.*, 22:322–343.
- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43:1–59.
- Best, N. G., Ickstadt, K., Wolpert, R. L., and Briggs, D. J. (2000). Combining models of health and exposure data: the SAVIAH study. In Elliott, P., Wakefield, J., Best, N., and Briggs, D., editors, *Spatial Epidemiology: Methods and Applications*, pages 393–414. Oxford: Oxford University Press.
- Box, G. E. P. (1980). Sampling and Bayes’ inference in scientific modelling and robustness. *J. R. Statist. Soc., Ser. A*, 143:383–430.
- Chaloner, K. (1994). Residual analysis and outliers in Bayesian hierarchical models. In Freeman, P. R. and Smith, A. F. M., editors, *Aspects of uncertainty: A tribute to D. V. Lindley*, chapter 10, pages 149–157. Wiley.
- Dahl, F. A., Gåsemyr, J., and Natvig, B. (2007). A robust conflict measure of inconsistencies in Bayesian hierarchical models. *Scand. J. Statist.* doi: 10.1111/j.1467-9469.2007.00560.x.
- Dey, D. K., Gelfand, A. E., Swartz, T. B., and Vlachos, P. K. (1998). A simulation-intensive approach for checking hierarchical models. *Test*, 7:325–346.
- Gaver, D. P. and O’Muircheartaigh, I. G. (1987). Robust empirical bayes analyses of event rates. *Technometrics*, 29:1–15.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, 85:398–409.
- Gelman, A., Meng, X. L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica*, 6:733–760.
- George, E. I., Makov, U. E., and Smith, A. F. M. (1993). Conjugate likelihood distributions. *Scand. J. Statist.*, 20:147–156.
- Green, P. J. and Richardson, S. (2002). Hidden Markov models and disease mapping. *J. Am. Statist. Ass.*, 97:1055–1070.
- Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *J. R. Statist. Soc., Ser. B*, 29:83–100.
- Hjort, N. L., Dahl, F. A., and Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *J. Am. Statist. Assoc.*, 101:1157–1174.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
- Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.
- Marshall, E. C. and Spiegelhalter, D. J. (2003). Approximate cross-validators predictive checks in disease mapping models. *Stat. Med.*, 22:1649–1660.
- Marshall, E. C. and Spiegelhalter, D. J. (2007). Identifying outliers in Bayesian hierarchical models: a simulation-based approach. *Bayesian Analysis*, 2:409–444.

- Meng, X. L. (1994). Posterior predictive p -values. *Ann. Statist.*, 22:1142–1160.
- O’Hagan, A. (2003). HSSS model criticism. In Green, P. J., Richardson, S., and Hjort, N. L., editors, *Highly Structured Stochastic Systems*, pages 423–444. Oxford: Oxford University Press.
- Rezvani, A., Mollié, A., Doyon, F., and Sancho-Garnier, H. (1997). Atlas de la Mortalité par Cancer en France, Période 1986-1993. *Paris: Editions INSERM*.
- Rubin, D. B. (1984). Bayesian justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, 12:1151–1172.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2004). *WinBUGS Examples Volume 1*. WinBUGS Version 1.4.1.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B*, 64:583–616.
- Thomas, A., Best, N., Lunn, D., Arnold, R., and Spiegelhalter, D. (2004). *GeoBUGS User Manual*. Version 1.2.

Appendix

A π and ψ functions

A.1 Rats

The π and ψ functions for the parameters in the `Rats` example in Section 3.1 are

$$\begin{aligned}
 \pi_{\alpha_i}(x) &= \Phi(\alpha_i; \alpha_c, \tau_\alpha) & \psi_{\alpha_i}(x) &= \Phi(\alpha_i; y_i, T \cdot \tau_c), \quad i = 1, \dots, N \\
 \pi_{\beta_i}(x) &= \Phi(\beta_i; \beta_c, \tau_\beta) & \psi_{\beta_i}(x) &= \Phi(\beta_i; \frac{s_{ty}}{s_{tt}}, s_{tt} \cdot \tau_c), \quad i = 1, \dots, N \\
 \pi_{\tau_c}(x) &= F(\tau_c; U_{\sigma_c}) & \psi_{\tau_c}(x) &= \Gamma(\tau_c; \frac{NT}{2} + 1, \sum_{i=1}^N \sum_{j=1}^T (y_{ij} - \mu_{ij})^2 / 2) \\
 \pi_{\alpha_c}(x) &= \Phi(\alpha_c; \mu_{\alpha_c}, \tau_{\alpha_c}) & \psi_{\alpha_c}(x) &= \Phi(\alpha_c; \bar{\alpha}, N \cdot \tau_\alpha) \\
 \pi_{\tau_\alpha}(x) &= F(\tau_\alpha; U_{\sigma_\alpha}) & \psi_{\tau_\alpha}(x) &= \Gamma(\tau_\alpha; \frac{N}{2} + 1, \sum_{i=1}^N (\alpha_i - \alpha_c)^2 / 2) \\
 \pi_{\beta_c}(x) &= \Phi(\beta_c; \mu_{\beta_c}, \tau_{\beta_c}) & \psi_{\beta_c}(x) &= \Phi(\beta_c; \bar{\beta}, N \cdot \tau_\beta) \\
 \pi_{\tau_\beta}(x) &= F(\tau_\beta; U_{\sigma_\beta}) & \psi_{\tau_\beta}(x) &= \Gamma(\tau_\beta; \frac{N}{2} + 1, \sum_{i=1}^N (\beta_i - \beta_c)^2 / 2),
 \end{aligned}
 \tag{38}$$

where $s_{tt} = \sum_{j=1}^T (t_j - \bar{t})^2$, $s_{ty} = \sum_{j=1}^T (t_j - \bar{t}) \cdot (y_{ij} - y_i)$ and $\mu_{ij} = \alpha_i + \beta_i(t_j - \bar{t})$.

A.2 SAVIAH

The π and ψ functions for the parameters in `SAVIAH1` in Section 3.2 are

$$\begin{aligned}
 \pi_{\beta_0}(x) &= \Gamma(\beta_0; \alpha_0, \tau_0) & \psi_{\beta_0}(x) &= \Gamma(\beta_0; y_{\cdot(J+1)} + 1, \sum_{i=1}^I N_i) \\
 \pi_{\beta_1}(x) &= \Gamma(\beta_1; \alpha_1, \tau_1) & \psi_{\beta_1}(x) &= \Gamma(\beta_1; y_{\cdot(J+2)} + 1, \sum_{i=1}^I N_i Z_i) \\
 \pi_{\beta_2}(x) &= \Gamma(\beta_2; \alpha_2, \tau_2) & \psi_{\beta_2}(x) &= \Gamma(\beta_2; y_{\cdot\cdot} + 1, \sum_{i=1}^I \sum_{j=1}^J N_i k_{ij} \gamma_j) \\
 \pi_{\gamma_j}(x) &= \Gamma(\gamma_j; \alpha_\gamma, \tau_\gamma) & \psi_{\gamma_j}(x) &= \Gamma(\gamma_j; y_{\cdot j} + 1, \beta_2 \sum_{i=1}^I N_i k_{ij}), \quad j = 1, \dots, J,
 \end{aligned}
 \tag{39}$$

where $y_{\cdot j} = \sum_{i=1}^I y_{ij}$ and $y_{\cdot\cdot} = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$.

A.3 Larynx

The π and ψ functions for the parameters in Larynx1 in Section 3.3 are

$$\begin{aligned}
 \pi_{a_i}(x) &= \Omega\left(a_i; \bar{u}_i, n_i \tau_u\right) & \psi_{a_i}(x) &= \Gamma\left(a_i; y_i + 1, \exp(c) b_i E_i\right), \quad i = 1, \dots, I \\
 \pi_{b_i}(x) &= \Omega\left(b_i; 0, \tau_v\right) & \psi_{b_i}(x) &= \Gamma\left(b_i; y_i + 1, \exp(c) a_i E_i\right), \quad i = 1, \dots, I \\
 (40) \quad \pi_{\tau_u}(x) &= \Gamma\left(\tau_u; \alpha_v, \beta_v\right) & \psi_{\tau_u}(x) &= \Gamma\left(\tau_u; \frac{I-m}{2} + 1, \sum_{i \sim i'} (u_i - u_{i'})^2 / 2\right) \\
 \pi_{\tau_v}(x) &= \Gamma\left(\tau_v; \alpha_u, \beta_u\right) & \psi_{\tau_v}(x) &= \Gamma\left(\tau_v; \frac{I}{2} + 1, \sum_{i=1}^I v_i^2 / 2\right),
 \end{aligned}$$

where m is the number of "islands", which, because all the regions except Corse are connected, here equals 2.

For Larynx2 in Section 3.3.1, the π functions for τ_u and τ_v are

$$\begin{aligned}
 (41) \quad \pi_{\tau_u}(x) &= \Omega\left(\tau_u; \mu, \rho\right) \\
 \pi_{\tau_v}(x) &= \Omega\left(\tau_v; \mu, \rho\right),
 \end{aligned}$$

The rest of the π and ψ functions are the same as in (40).

Data

Larynx

Table 1 contains the y_i/E_i values for the Larynx application, based on the data reported by Rezvani et al. (1997).

Table 1: Larynx data

Region	$\frac{y_i}{E_i}$	Region	$\frac{y_i}{E_i}$	Region	$\frac{y_i}{E_i}$	Region	$\frac{y_i}{E_i}$
1	1.19	25	0.79	49	0.51	73	0.37
2	1.32	26	1.09	50	0.08	74	1.14
3	0.93	27	0.94	51	0.34	75	1.3
4	0.56	28	0.92	52	0.41	76	1.26
5	2.12	29	0.85	53	0.59	77	1.29
6	1.46	30	0.86	54	0.82	78	0.95
7	0.66	31	0.66	55	0.62	79	0.57
8	1.2	32	1.14	56	0.32	80	0.89
9	1.42	33	1.24	57	1.29	81	0.61
10	0.99	34	1.22	58	0.56	82	0.91
11	1.01	35	1.07	59	1.25	83	1.5
12	0.73	36	1.13	60	0.87	84	1.14
13	1.29	37	1.01	61	0.67	85	0.63
14	0.8	38	0.93	62	1.6	86	0.31
15	0.22	39	0.47	63	1.29	87	0.48
16	0.32	40	1.12	64	0.87	88	0.93
17	0.68	41	0.97	65	0.46	89	1.15
18	0.57	42	0.58	66	1.09	90	1.35
19	0.54	43	0.52	67	0.66	91	0.9
20	1	44	0.93	68	0.86	92	1.57
21	0.8	45	0.91	69	0.83	93	1.64
22	0.98	46	1.23	70	0.71	94	1.32
23	1.08	47	0.57	71	0.59	95	1.58
24	0.76	48	0	72	0.56		