Convergence Assessment for Reversible Jump MCMC Samplers

John M. Castelloe*, Dale L. Zimmerman

SAS Institute Inc. and Department of Statistics and Actuarial Science, University of Iowa

Abstract

A new method for assessing convergence of a reversible jump MCMC sampler is presented, along with examples demonstrating both its sensitivity and specificity. The method is an extension of the popular technique of Gelman and Rubin (1992) in two ways. First, the one-way ANOVA paradigm applied to chains is extended to a two-way paradigm by treating the sampler's parameter space as a factor, an idea similar to one explored by Brooks and Giudici (2000) but addressed more appropriately as an unbalanced ANOVA. Second, the technique is extended from univariate to multivariate following the suggestions of Brooks and Gelman (1998): several parameters can be monitored simultaneously and conservatively via worst-case scalar functions.

Keywords: analysis of variance, convergence, convergence diagnostic, Markov chain Monte Carlo, mixing

1 Introduction

Before conducting inference using output from a Markov chain Monte Carlo sampler (Hastings, 1970; Metropolis et al., 1953; Geman and Geman, 1984; Gelfand and Smith, 1990), the output should be analyzed to determine a point at which the sampler has "converged" to the proper limiting distribution. A thorough review of Markov chain Monte Carlo (MCMC) convergence assessment techniques is provided by Cowles and Carlin (1996) and Mengersen, Robert, and Guihenneuc-Jouyaux (1998). No existing technique, however, can be applied directly to output from a reversible jump MCMC (RJMCMC) sampler (Green, 1995). RJMCMC is essentially a random sweep Metropolis-Hastings method adapted for general state spaces. Typical MCMC methods (e.g. the Gibbs sampler and Metropolis-Hastings algorithm) apply only to situations in which the dimension of the parameter space is fixed. However, in many important applications the dimension of the parameter space varies across sweeps. For example, in analysis of mixture models with unknown number of

---

*E-mail: john.castelloe@sas.com; Address: SAS Institute Inc., SAS Campus Dr., Cary, NC, 27513

components (Richardson and Green, 1997), a different parameter space is needed for each possible value for $k = \#\{\text{mixture components}\}$, since an increase in k requires an additional set of parameters to characterize the new component. RJMCMC is an ideal tool for such applications, and its primary drawback has been the lack of suitable convergence assessment methods. We review existing approaches and then derive a new method to assess convergence of an RJMCMC sample. Our new method is an extension of the popular technique of Gelman and Rubin (1992), both from MCMC to RJMCMC (to encompass all of the parameter spaces) and from univariate to multivariate (to monitor several parameters simultaneously).

There are two distinct aspects of convergence of an MCMC sampler to consider:

1. Are the samples being generated from the correct distribution?
2. Has the entire parameter space been traversed?

It is difficult to rigorously verify either condition; a generally accepted strategy is to run several chains started at over-dispersed values. If at some point all chains are generating samples from approximately the same distribution, then this distribution is presumed to be the correct one (a justifiable assumption when the Markov chain is designed properly). If the starting values are appropriately over-dispersed, then it is also likely that the parameter space has been thoroughly traversed as well.

Most methods described in Cowles and Carlin (1996) and Mengersen, Robert, and Guihenneuc-Jouyaux (1998) are univariate, considering only one parameter at a time. Currently, the two most popular types are those developed by Geweke (1991) and Gelman and Rubin (1992). Geweke (1991) proposes comparing (univariate) sample means of a parameter computed from different parts of a chain, using variance estimates adjusted for autocorrelation. Gelman and Rubin (1992) propose an analysis of variance (ANOVA) type approach in which several chains are run, and the ratio of a pooled variance estimate and a within-chain variance estimate, similar to the comparison between total mean-square and error mean square in a one-way ANOVA with "chain" being the factor, is calculated. The idea is that if the two variances are comparable, then the chains are probably realizations from a common distribution, presumably the correct limiting distribution. This method depends on the absence of other significant factors, but in RJMCMC, the indicator of parameter space (which from now on we refer to as "model") could be considered a factor in this paradigm. Some parameters may not retain the same meaning across models, and even those that do may vary considerably in different models.

Neither of the two popular methods mentioned above (nor any others that the authors are aware of) are sufficient to detect lack of convergence *within* models. Extensions to Geweke's technique do not appear to be feasible, since output from a RJMCMC sampler for a given model consists of a series of uninterrupted sequences separated by visits to other models, and thus an autocorrelation would need to be assessed in each of these sequences. It is an extension of Gelman and Rubin's method, both from univariate to multivariate *and* 1-way-ANOVA to 2-way-ANOVA, that we develop in section 4. First we discuss some other extensions of their technique which are relevant.

2 Brooks and Gelman's MCMC Convergence Diagnostic

Brooks and Gelman (1998) introduce several different versions of Gelman and Rubin's convergence diagnostic and suggest monitoring both numerator and denominator, not just a ratio. One of the versions is multivariate in the sense of providing an upper bound of an analogous convergence diagnostic computed for a set of scalar parameters.

The univariate approach of Gelman and Rubin requires running $C > 1$ chains of a MCMC sampler (with T sweeps each, say) with over-dispersed starting values. A number m of successive (overlapping) "batches" of increasing length (multiples of a base batch length b) of the output are analyzed from each chain. Let $\theta_c^{(qb+1)}, \ldots, \theta_c^{(2qb)}$, denote the q^{th} batch of length qb, from chain c for a scalar parameter θ, where $c \in \{1, \ldots, C\}$. Successive batches for $q = 1, \ldots, \frac{T}{b}$ are used. Brooks and Gelman (1998) propose monitoring $\widehat{V}^{(q)}(\theta)$, $W^{(q)}(\theta)$ and $\frac{\widehat{V}^{(q)}(\theta)}{W^{(q)}(\theta)}$ (which they call the *potential scale reduction factor*, or PSRF), defined below, computed for each batch.

Defining $\bar{\theta}_c^{(\cdot)}$ and $\bar{\theta}_\cdot^{(\cdot)}$ as

$$\bar{\theta}_c^{(\cdot)} = \frac{1}{qb} \sum_{t=qb+1}^{2qb} \theta_c^{(t)} \quad \text{and} \quad \bar{\theta}_\cdot^{(\cdot)} = \frac{1}{qbC} \sum_{c=1}^{C} \sum_{t=qb+1}^{2qb} \theta_c^{(t)},$$

the quantities of interest are defined as follows:

$$\widehat{V}^{(q)}(\theta) = \frac{qb-1}{qb} W^{(q)}(\theta) + \left(1 + \frac{1}{C}\right) B(\theta)/(qb)$$

and

$$W^{(q)}(\theta) = \frac{1}{C(qb-1)} \sum_{c=1}^{C} \sum_{t=qb+1}^{2qb} \left(\theta_c^{(t)} - \bar{\theta}_c^{(\cdot)}\right)^2$$

where

$$B(\theta)/(qb) = \frac{1}{C-1} \sum_{c=1}^{C} \left(\bar{\theta}_c^{(\cdot)} - \bar{\theta}_\cdot^{(\cdot)}\right)^2.$$

The value of $\widehat{V}^{(q)}(\theta)$ should be larger than $W^{(q)}(\theta)$ for small q, since the starting values are over-dispersed; they may approach a common value as q increases, indicating that the variation is homogeneous across chains.

It may happen that the numerator and denominator happen to fluctuate together but yield a ratio close to 1, so Brooks and Gelman (1998) recommend monitoring these individually in addition to the ratio. They mention that, provided the starting values are appropriately over-dispersed, the settling of $\frac{\widehat{V}^{(q)}(\theta)}{W^{(q)}(\theta)}$ to a neighborhood of 1, *and* of $\widehat{V}^{(q)}(\theta)$ and $W^{(q)}(\theta)$ approximately to a common value for $q \geq q_0$, are generally adequate reasons to justify inferences based on posterior means and variances of the collection of samples $\{\theta^{(q_0 b+1)}, \theta^{(q_0 b+2)}, \ldots\}$. This situation often suggests additionally that the chains are following the same distribution, but they warn that only approximate equivalence of the first 2 moments across chains has been established. It is difficult to determine how close to 1 is "close enough": they cite a cutoff of 1.2 as a rule of thumb in one of their examples.

The multivariate version for a vector $\boldsymbol{\theta}$ of parameters is defined analogously, estimating posterior variance-covariance matrices instead of scalar variances:

Defining $\bar{\boldsymbol{\theta}}_c^{(\cdot)}$ and $\bar{\boldsymbol{\theta}}_\cdot^{(\cdot)}$ as

$$\bar{\boldsymbol{\theta}}_c^{(\cdot)} = \frac{1}{qb} \sum_{t=qb+1}^{2qb} \boldsymbol{\theta}_c^{(t)} \quad \text{and} \quad \bar{\boldsymbol{\theta}}_\cdot^{(\cdot)} = \frac{1}{qbC} \sum_{c=1}^{C} \sum_{t=qb+1}^{2qb} \boldsymbol{\theta}_c^{(t)},$$

the multivariate convergence diagnostics are given by

$$\widehat{V}^{(q)}(\boldsymbol{\theta}) = \frac{qb-1}{qb} W^{(q)}(\boldsymbol{\theta}) + \left(1 + \frac{1}{C}\right) B(\boldsymbol{\theta})/(qb)$$

and

$$W^{(q)}(\boldsymbol{\theta}) = \frac{1}{C(qb-1)} \sum_{c=1}^{C} \sum_{t=qb+1}^{2qb} \left(\boldsymbol{\theta}_c^{(t)} - \bar{\boldsymbol{\theta}}_c^{(\cdot)}\right) \left(\boldsymbol{\theta}_c^{(t)} - \bar{\boldsymbol{\theta}}_c^{(\cdot)}\right)'$$

where

$$B(\boldsymbol{\theta})/(qb) = \frac{1}{C-1} \sum_{c=1}^{C} \left(\bar{\boldsymbol{\theta}}_c^{(\cdot)} - \bar{\boldsymbol{\theta}}_\cdot^{(\cdot)}\right) \left(\bar{\boldsymbol{\theta}}_c^{(\cdot)} - \bar{\boldsymbol{\theta}}_\cdot^{(\cdot)}\right)'.$$

The multivariate PSRF (MPSRF) is then defined as a maximum root statistic-type measure of distance between $\widehat{V}^{(q)}(\boldsymbol{\theta})$ and $W^{(q)}(\boldsymbol{\theta})$:

$$MPSRF(\boldsymbol{\theta}) = \max_{a \in \Re^p} \frac{a' \widehat{V}^{(q)}(\boldsymbol{\theta}) a}{a' W^{(q)}(\boldsymbol{\theta}) a},$$

where p is the dimension of θ. They proceed to prove that $MPSRF(\boldsymbol{\theta})$ can be represented in terms of the maximum eigenvalue of $\left[W^{(q)}(\boldsymbol{\theta})\right]^{-1} \widehat{V}^{(q)}(\boldsymbol{\theta})$, and that it provides an upper bound on the collection of univariate PSRF's, $\frac{\widehat{V}^{(q)}(\theta_i)}{W^{(q)}(\theta_i)}$, where θ_i is the i^{th} scalar component of $\boldsymbol{\theta}$. We now present these results in generic notation.

Lemma 1 *For two non-singular, positive definite and symmetric $p \times p$ matrices M and N,*

$$\max_{a \in \Re^p} \frac{a' M a}{a' N a} = \lambda,$$

where λ is the largest eigenvalue of $N^{-1}M$.

Proof: See Mardia, Kent, and Bibby (1979, Theorem A.9.2). $\square$

Lemma 2 *Let M and N be two non-singular, positive definite and symmetric $p \times p$ matrices, and denote the diagonal elements as $\{m_1, \ldots, m_p\}$ and $\{n_1, \ldots, n_p\}$, respectively. Then*

$$\max_{a \in \Re^p} \frac{a'Ma}{a'Na} \geq \max_{i \in \{1,\ldots,p\}} \frac{m_i}{n_i}.$$

Proof: Let $\mathbf{i}_j$ denote a $p \times 1$ vector of zeroes with the j^{th} entry replaced by 1. Then

$$\max_{a \in \Re^p} \frac{a'Ma}{a'Na} \geq \max_{j \in \{1,\ldots,p\}} \frac{\mathbf{i}_j' M \mathbf{i}_j}{\mathbf{i}_j' N \mathbf{i}_j} = \max_{i \in \{1,\ldots,p\}} \frac{m_i}{n_i}. \quad \square$$

Note that the collection $\{v_1, \ldots, v_p\}$ of diagonal elements of the multivariate version of $\widehat{V}$ are equivalent to the univariate versions, and that the same holds for the diagonal elements $\{w_1, \ldots, w_p\}$ of W. Thus Lemma 2 establishes that Brooks and Gelman's MPSRF is an upper bound of the unvariate PSRF's. They suggest monitoring this MPSRF, and also $f\left(\widehat{V}^{(q)}(\boldsymbol{\theta})\right)$ and $f\left(W^{(q)}(\boldsymbol{\theta})\right)$ for some real-valued function $f(\cdot)$, such as the determinant.

3 Brooks and Giudici's RJMCMC Convergence Diagnostic

Brooks and Giudici (2000) introduce the first proposed method, a univariate one, specifically designed for RJMCMC convergence assessment. The basic idea is to compute various decompositions of the estimated variance of a collection of samples of a scalar parameter from C different chains. Two factors determine the decompositions: "model" (the indicator of the different parameter spaces) and "chain." The scalar parameter chosen must have the same meaning across all models. They claim that the decompositions correspond to three pairs of variance estimates, with each member of a pair estimating the same quantity. Thus they propose following the method of Brooks and Gelman (1998) by monitoring each of these 3 pairs and the 3 ratios they produce.

Brooks and Giudici do not specify how batches should be chosen for analysis. For simplicity of notation, we will consider calculations for one batch only. Suppose $C > 1$ chains of a RJMCMC sampler are run. Let θ be a scalar parameter in the chain (with equivalent interpretation across models), T denote the batch size, and M denote the total number of different models (different parameter spaces) visited by any chain for this batch. Define θ^r_{cm} as the r^{th} value of θ occurring in chain c and model m. Also define R_{cm} as the number of times model m occurs in chain c and $R_{\cdot m}$ as the number of times model m occurs across chains. Note that $R_{c\cdot} = T$ and the total number of sweeps in the batch over all chains is CT. Brooks and Giudici (2000) define the following quantities (note: the subscripts on the left-hand side are parts of the names, and do not correspond to values of indices on the right-hand side):

$$
\begin{align*}
\widehat{V}(\theta) &= \frac{1}{CT-1}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2 \\
W_{\mathrm{c}}(\theta) &= \frac{1}{C}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\frac{\left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{c\cdot}\right)^2}{T-1} \\
W_{\mathrm{m}}(\theta) &= \frac{1}{M}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\frac{\left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{\cdot m}\right)^2}{R_{\cdot m}-1} \\
W_{\mathrm{m}}W_{\mathrm{c}}(\theta) &= \frac{1}{CM}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\frac{\left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{cm}\right)^2}{R_{cm}-1} \\
B_{\mathrm{m}}(\theta) &= \sum_{m=1}^{M}\frac{\left(\bar{\theta}^{\cdot}_{\cdot m} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2}{M-1} \tag{1} \\
B_{\mathrm{m}}W_{\mathrm{c}}(\theta) &= \sum_{c=1}^{C}\sum_{m=1}^{M}\frac{\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot}\right)^2}{C(M-1)} \tag{2}
\end{align*}
$$

where

$$
\begin{align*}
\bar{\theta}^{\cdot}_{cm} &= \frac{1}{R_{cm}}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm} \\
\bar{\theta}^{\cdot}_{c\cdot} &= \frac{1}{T}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm} \\
\bar{\theta}^{\cdot}_{\cdot m} &= \frac{1}{R_{\cdot m}}\sum_{c=1}^{C}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm} \\
\bar{\theta}^{\cdot}_{\cdot\cdot} &= \frac{1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm}.
\end{align*}
$$

(Note: we have corrected two obvious typographical errors in the definitions of W_{m} and W_{c}).

Brooks and Giudici claim the following:

1. Both $\widehat{V}(\theta)$ and $W_{\mathrm{c}}(\theta)$ should well approximate the true variation of θ under the stationary distribution of the Markov chain (and this comparison is essentially the original Gelman and Rubin comparison)

2. Both $W_{\mathrm{m}}(\theta)$ and $W_{\mathrm{m}}W_{\mathrm{c}}$ should well approximate the true mean within-model variance

3. Both $B_{\mathrm{m}}(\theta)$ and $B_{\mathrm{m}}W_{\mathrm{c}}$ should well approximate the true between-model variance.

It is true that, in the case of equal R_{cm} counts, these 6 quantities correspond to the descriptions they attach using ANOVA terminology. However, in the case of unequal R_{cm} counts, the meanings of the quantities are unclear. In general, the R_{cm} counts will be dramatically different, as some models are less likely than others and hence visited infrequently. Brooks and Giudici encounter this situation in their own example: the second and third comparisons break down when *one* of the chains visits a rare model *once* late in the sequence. It is easy to see why this occurs: the comparisons are based on *unweighted* sample variances of means, allowing imprecise sample means from rare models to heavily influence their values. While it may be useful in some situations to have such diagnostics to detect rare model visits, we do not feel that this satisfies the definition of a *convergence* diagnostic. It is perfectly fine for some models to be more unlikely than others. We reconsider Brooks and Giudici's apparent initial motives and develop a strategy from scratch by considering two-way unbalanced ANOVA models with the more appropriate *weighted* sample variances.

4 An Alternative Strategy for RJMCMC Convergence Assessment

In this section we design a convergence diagnostic especially for RJMCMC situations in which different parameter spaces ("models") are indexed by some parameter in the chain. Our convergence diagnostic utilizes multiple chains and detects the following:

1. variation between chains (i.e., the target of the original Gelman and Rubin diagnostic: variation that is not homogeneous across chains),
2. an interaction between models and chains (i.e., between-model variation that differs from one chain to another), and
3. significant differences in the frequencies of model visits from one chain to another.

Any one of these three conditions would indicate that the chains are not living in the same stationary distribution, and hence that convergence has not occurred.

4.1 Forms of Variation Estimators

Suppose we have a RJMCMC sampler which produces output of a parameter vector Θ, with some $k \in \Theta$ indexing "model" and $\boldsymbol{\theta} \subset \Theta$ a vector of parameters which retain the same meaning across models ($k \notin \boldsymbol{\theta}$). Let the output of $\boldsymbol{\theta}$ be represented as $\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots\right)$. Suppose $C > 1$ chains of this sampler are run for the same number of sweeps. For simplicity of notation, we will consider output from one *batch* of size qb only, i.e.,

$$\left(\boldsymbol{\theta}_1^{(qb+1)}, \ldots, \boldsymbol{\theta}_1^{(2qb)}\right), \ldots, \left(\boldsymbol{\theta}_C^{(qb+1)}, \ldots, \boldsymbol{\theta}_C^{(2qb)}\right)$$

for some q and base batch size b. We now represent this collection in a more convenient notation (as in section 3), which we describe completely below.

Let

$$\begin{aligned}
\boldsymbol{\theta} &= \text{vector of parameters retaining same interpretation} && (3)\\
&\quad\ \text{across models} \\
\theta &= \text{arbitrary scalar component of } \boldsymbol{\theta} && (4)\\
C &= \text{number of chains} && (5)\\
T &= \text{batch size (this many sweeps per chain)} && (6)\\
M &= \text{number of distinct models visited by any chain} && (7)\\
\boldsymbol{\theta}^{r}_{cm} &= \text{value of } \boldsymbol{\theta} \text{ for } r^{\text{th}} \text{ occurrence of} && (8)\\
&\quad\ \text{model } m \text{ in chain } c \\
R_{cm} &= \text{number of times model } m \text{ occurred in chain } c && (9)\\
R_{\cdot m} &= \sum_{c=1}^{C} R_{cm} && (10)\\
\bar{\boldsymbol{\theta}}^{\cdot}_{cm} &= \frac{1}{R_{cm}} \sum_{r=1}^{R_{cm}} \boldsymbol{\theta}^{r}_{cm} && (11)\\
\bar{\boldsymbol{\theta}}^{\cdot}_{c\cdot} &= \frac{1}{T} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \boldsymbol{\theta}^{r}_{cm} && (12)\\
\bar{\boldsymbol{\theta}}^{\cdot}_{\cdot m} &= \frac{1}{R_{\cdot m}} \sum_{c=1}^{C} \sum_{r=1}^{R_{cm}} \boldsymbol{\theta}^{r}_{cm} && (13)\\
\bar{\boldsymbol{\theta}}^{\cdot}_{\cdot\cdot} &= \frac{1}{CT} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \boldsymbol{\theta}^{r}_{cm}. && (14)
\end{aligned}$$

Our convergence diagnostic is based on the following estimates of variation: (note: the subscripts on the left-hand side are parts of the names, and do not correspond to values of indices on the right-hand side):

$$\begin{aligned}
\widehat{V}(\theta) &= \frac{1}{CT-1} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^{2} && (15)\\
W_{\mathrm{c}}(\theta) &= \frac{1}{C(T-1)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{c\cdot}\right)^{2} && (16)\\
W_{\mathrm{m}}(\theta) &= \frac{1}{CT-M} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{\cdot m}\right)^{2} && (17)\\
W_{\mathrm{m}}W_{\mathrm{c}}(\theta) &= \frac{1}{C(T-M)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^{r}_{cm} - \bar{\theta}^{\cdot}_{cm}\right)^{2} && (18)
\end{aligned}$$

Note that these quantities may be interpreted as total variation ($\widehat{V}$), variation within chains (W_{c}), variation within models (W_{m}), and variation within models and chains ($W_{\mathrm{m}}W_{\mathrm{c}}$). The first of two comparisons we will use involves $\widehat{V}$ and W_{c}, which are defined in the same way as Brooks and Giudici (2000), and correspond (except for minor differences in multiplicative factors) to the original Gelman and Rubin diagnostic. The second involves W_{m} and $W_{\mathrm{m}}W_{\mathrm{c}}$, which are defined differently so as to correspond meaningfully to elements of appropriate ANOVA models. We establish these correpondences, for both pairs of variation estimates, in section 4.2.

4.2 Interpretation from an ANOVA Perspective

The output from the RJMCMC sampler can be considered as a collection of observations from a factorial design, in which the factors are "chain" and/or "model." An analysis of variance (ANOVA) can be used to assess the significance of factors and interactions. The primary exception to the usual assumptions of ANOVA approaches is that the samples are not independent. However, we shall see that certain quantities constructed from ANOVA features are still useful in suggesting and interpreting our convergence diagnostics. Consider the three ANOVA models defined in Tables 1 – 3.

ANOVA 1

$$\boldsymbol{\theta}^r_{cm} = \mu + \alpha_c + e^r_{cm(1)}$$

where: $\alpha_c \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\mathrm{ch}})$

$e^r_{cm(1)} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\mathrm{er(ch)}})$

Source	df	SS
chain	$C-1$	$T\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$
error(chain)	$C(T-1)$	$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{c\cdot}\right)^2$
total	$CT-1$	$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$

Table 1: ANOVA 1: One-way ANOVA with factor chain (random), balanced.

We represent model as a "fixed" factor and chain as "random," which is certainly debatable. However, basically the same conclusions are reached regardless of how the factors are treated (differing only in description of effects and minor coefficient changes). For example, if model were treated as random, effects would be described in terms of σ^2_{mo}, not the individual effects $\{\beta_m\}$. If chain were treated as random, effects

ANOVA 2

$\boldsymbol{\theta}^r_{cm} = \mu + \beta_m + e^r_{cm(2)}$

where: $\sum_{m=1}^{M} \beta_m = 0$

$e^r_{cm(2)} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\text{er(mo)}})$

$\text{``}\sigma^2_{\text{mo}}\text{''} = \frac{1}{M-1} \sum_{m=1}^{M} \beta_m{}^2$

Source	df	SS
model	$M - 1$	$\sum_{m=1}^{M} R_{\cdot m} \left(\bar{\theta}^{\cdot}_{\cdot m} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$
error(model)	$CT - M)$	$\sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{\cdot m}\right)^2$
total	$CT - 1$	$\sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$

Table 2: ANOVA 2: One-way ANOVA with factor model (fixed), unbalanced.

would be described in terms of $\{\alpha_c\}$ instead of σ^2_{ch}.

Winer (1971, pp. 212 and 403) establishes the expressions for degrees of freedom entries. All terms which have the same notation in the three ANOVA's (e.g., $\mu, \alpha_c, \beta_m, \sigma^2_{\text{ch}}$) are *equivalent*. The error terms $(e^r_{cm(1)}, e^r_{cm(2)}, e^r_{cm(3)})$ are labeled differently because they are in general *not* equivalent for the three models. In comparing entries in the ANOVA's with (15) – (18), it is clear that

$$\widehat{V} \equiv \text{MS}_{\text{tot}} \text{ for ANOVA 1,} \tag{19}$$

$$W_{\text{c}} \equiv \text{MS}_{\text{er(ch)}} \text{ for ANOVA 1,} \tag{20}$$

$$W_{\text{m}} \equiv \text{MS}_{\text{er(mo)}} \text{ for ANOVA 2}, \mathit{and} \tag{21}$$

$$W_{\text{m}} W_{\text{c}} \equiv \text{MS}_{\text{er(ch}\times\text{mo)}} \text{ for ANOVA 3,} \tag{22}$$

where "MS" denotes mean-square. We can of course *not* claim that an ANOVA model is a realistic description of the output from parallel chains of a RJMCMC sampler, since the assumptions of independence and normality in general do not hold. However, the effects of dependence are likely to be at least approximately cancelled out since we are focusing on ratios of mean-squares. The convergence diagnostics of Gelman and Rubin (1992); Brooks and Gelman (1998); and Brooks and Giudici (2000) all make this same implicit assumption. Furthermore, we will not rely on approximate normality for inferences. Thus we will proceed by considering the sampler output as occurring approximately according to an ANOVA model (not specifying yet which one(s), but using Tables 1 – 3 as appropriate).

ANOVA 3

$$\boldsymbol{\theta}^r_{cm} = \mu + \alpha_c + \beta_m + (\alpha\beta)_{cm} + e^r_{cm(3)}$$

where:

$$\alpha_c \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\text{ch}})$$

$$\textstyle\sum_{m=1}^{M} \beta_m = 0$$

$$(\alpha\beta)_{cm} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\text{ch}\times\text{mo}})$$

$$e^r_{cm(3)} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_{\text{er(ch}\times\text{mo)}})$$

$$\text{“}\sigma^2_{\text{mo}}\text{”} = \tfrac{1}{M-1} \textstyle\sum_{m=1}^{M} {\beta_m}^2$$

Source	df	SS
chain	$C-1$	$T\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$
model	$M-1$	$\sum_{m=1}^{M} R_{\cdot m}\left(\bar{\theta}^{\cdot}_{\cdot m} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$
chain×model	$(C-1)(M-1)$	$\sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$
error(chain×model)	$C(T-M)$	$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right)^2$
total	$CT-1$	$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2$

Table 3: ANOVA 3: Two-way ANOVA with factors model (fixed), chain (random) and chain×model interaction (random, unrestricted), balanced across chain only.

Derivations of expected mean-squares for the three ANOVA models (shown in section 9) reveal that the expected values of (15) – (18) under ANOVA assumptions are given as follows:

$$E\widehat{V} = \sigma^2_{\text{er(ch)}} + \left[\frac{(C-1)T}{CT-1}\right]\sigma^2_{\text{ch}} \tag{23}$$

$$EW_{\text{c}} = \sigma^2_{\text{er(ch)}} \tag{24}$$

$$EW_{\text{m}} = \sigma^2_{\text{er(ch}\times\text{mo)}} \quad + \tag{25}$$

$$\left[\frac{(C-1)T}{CT-M} + \frac{1}{C^2(CT-M)}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{\left(CR_{cm} - R_{\cdot m}\right)^2}{R_{\cdot m}}\right]\sigma^2_{\text{ch}} + \tag{26}$$

$$\left[\frac{2}{C^2(CT-M)T}\sum_{m=1}^{M}\left\{\sum_{c=1}^{C}\left(CR_{cm} - R_{\cdot m}\right)^2\right\}\beta^2_m\right] + \tag{27}$$

$$\left[\frac{CT}{CT-M} + \frac{-1}{CT-M}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^2_{cm}}{R_{\cdot m}} \quad + \right.$$

$$\left.\frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}\left(CR_{cm} - R_{\cdot m}\right)\frac{R^2_{cm}}{R_{\cdot m}}\right]\sigma^2_{\text{ch}\times\text{mo}} \tag{28}$$

$$EW_{\text{m}}W_{\text{c}} = \sigma^2_{\text{er(ch}\times\text{mo)}} \tag{29}$$

If the set of within-chain model frequencies is equivalent for all chains (i.e., $R_{cm} = \frac{R_{\cdot m}}{C}\ \forall c, m$), then (25) – (28) simplifies to

$$EW_{\text{m}} = \sigma^2_{\text{er(ch}\times\text{mo)}} + \left[\frac{(C-1)T}{CT-M}\right]\sigma^2_{\text{ch}} + \left[\frac{(C-1)T}{CT-M}\right]\sigma^2_{\text{ch}\times\text{mo}}. \tag{30}$$

For large T (and any $\{R_{cm}\}$),

$$E\widehat{V} \ \approx\ \sigma^2_{\text{er(ch)}} + \left[\frac{C-1}{C}\right]\sigma^2_{\text{ch}} \tag{31}$$

and

$$\begin{aligned}
EW_{\text{m}} \quad &\approx \quad \sigma^2_{\text{er(ch}\times\text{mo)}} + \left[\frac{C-1}{C}\right]\sigma^2_{\text{ch}} + && (32)\\
& \left[\frac{1}{C^2(CT-M)}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{(CR_{cm}-R_{\cdot m})^2}{R_{\cdot m}}\right]\sigma^2_{\text{ch}} + && (33)\\
& \left[\frac{2}{C^3T^2-C^2MT}\sum_{m=1}^{M}\left\{\sum_{c=1}^{C}(CR_{cm}-R_{\cdot m})^2\right\}\beta_m^2\right] + && (34)\\
& \left[1+\left(\frac{-1}{CT-M}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^2}{R_{\cdot m}}\right)+\right.\\
& \left.\left(\frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}(CR_{cm}-R_{\cdot m})\frac{R_{cm}^2}{R_{\cdot m}}\right)\right]\sigma^2_{\text{ch}\times\text{mo}}. && (35)
\end{aligned}$$

If additionally the set of within-chain model frequencies is equivalent for all chains ($R_{cm} = \frac{R_{\cdot m}}{C}\ \forall c, m$), then (32) – (35) simplifies to (for large T):

$$EW_{\text{m}} \ \approx\ \sigma^2_{\text{er(ch}\times\text{mo)}} + \left[\frac{C-1}{C}\right]\sigma^2_{\text{ch}} + \left[\frac{C-1}{C}\right]\sigma^2_{\text{ch}\times\text{mo}}. \tag{36}$$

Notice that (26) and (27), in the presence of chain and model effects, respectively, increase as the model frequencies across chains, R_{cm}, deviate more from the frequencies $\frac{R_{\cdot m}}{C}$ that would occur if the set of within-chain model frequencies were equivalent for all chains.

The multiple of $\sigma^2_{\text{ch}\times\text{mo}}$ in expression (28) can be characterized as follows. Let

$$X = X_1 + X_2 + X_3$$

where

$$\begin{aligned}
X_1 &= \frac{CT}{CT-M}\\
X_2 &= \frac{-1}{CT-M}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^2}{R_{\cdot m}}\\
X_3 &= \frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}(CR_{cm}-R_{\cdot m})\frac{R_{cm}^2}{R_{\cdot m}}.
\end{aligned}$$

The ranges of X_2 and X_3 can be determined from consideration of two extreme cases,

$$A. \qquad R_{cm} = \frac{R_{\cdot m}}{C} \ \forall c, \quad \text{and}$$

$$B. \qquad R_{cm} = \begin{cases} R_{\cdot m}, & \text{for } c = c'_m \\ 0, & \text{for } c \neq c'_m \end{cases} \qquad \text{for some } \{c'_1, \ldots, c'_M\},$$

to be

$$\frac{-CT}{CT-M} \leq X_2 \leq \frac{-T}{CT-M}$$

and

$$0 \leq X_3 \leq \frac{2(C-1)}{CT} \sum_{m=1}^{M} R_{\cdot m}^2 \quad .$$

Also, X is strictly positive because (i) if case A holds, then $X = \frac{(C-1)T}{CT-M}$, and (ii) if case A does not hold, then X_3 is strictly positive. In general, X increases (although not necessarily monotonically) as the set of within-cell model frequencies becomes less homogeneous across chains.

Thus we can conclude the following about the ratios $\frac{E\widehat{V}}{EW_{\mathrm{c}}}$ and $\frac{EW_{\mathrm{m}}}{EW_{\mathrm{m}}W_{\mathrm{c}}}$:

1. $\frac{E\widehat{V}}{EW_{\mathrm{c}}} \geq 1$, with $\frac{E\widehat{V}}{EW_{\mathrm{c}}} = 1$ indicating the absence of a chain effect. The greater $\frac{E\widehat{V}}{EW_{\mathrm{c}}}$, the stronger the chain effect, with each term in the numerator and denominator stabilizing as $T \rightarrow \infty$ and thus preserving the validity of the magnitude as $T \rightarrow \infty$.

2. $\frac{EW_{\mathrm{m}}}{EW_{\mathrm{m}}W_{\mathrm{c}}} \geq 1$, with $\frac{EW_{\mathrm{m}}}{EW_{\mathrm{m}}W_{\mathrm{c}}} = 1$ indicating:

    (a) the absence of a chain effect, *and*

    (b) the absence of a chain×model interaction, *and*

    (c) either (i) no model effect or (ii) equality of the set of within-chain model frequencies across chains, or both.

    The greater the violation of any combination of these three criteria (2a)–(2c), the larger $\frac{EW_{\mathrm{m}}}{EW_{\mathrm{m}}W_{\mathrm{c}}}$ becomes. The relative weights of the three criteria as $T \rightarrow \infty$ (i.e., the sensitivity of the ratio to violations of each of the three criteria) are not yet fully understood. We can at least reason by (36) that when the set of within-chain model frequencies are somewhat homogeneous across chains (i.e., $CR_{cm} \approx R_{\cdot m} \ \forall c, m$), then the ratio has approximately equal sensitivity to (2a) and (2b), and so either a significant chain effect or chain×model interaction should be detected.

These properties clearly suggest the design of a convergence diagnostic based on the two ratios $\frac{\widehat{V}}{W_{\mathrm{c}}}$ and $\frac{W_{\mathrm{m}}}{W_{\mathrm{m}}W_{\mathrm{c}}}$. We suggest the use of both ratios, because it may help to narrow down the cause of any violations of convergence. In section 4.4, we show the exact form of the diagnostic technique we propose.

Expressions analogous to (1) and (2), represented with the proper degrees of freedom terms in the denominators, yield expected mean-squares that do not appear to be useful for comparison purposes. Further research is needed to determine what a ratio based on analogues of (1) and (2) would actually detect.

4.3 Multivariate Version

Define the following multivariate versions of (15) – (18):

$$\widehat{V}(\boldsymbol{\theta}) = \frac{1}{CT-1}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}} \left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{\cdot\cdot}\right)\left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{\cdot\cdot}\right)' \tag{37}$$

$$W_{\mathrm{c}}(\boldsymbol{\theta}) = \frac{1}{C(T-1)}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}} \left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{c\cdot}\right)\left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{c\cdot}\right)' \tag{38}$$

$$W_{\mathrm{m}}(\boldsymbol{\theta}) = \frac{1}{CT-M}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}} \left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{\cdot m}\right)\left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{\cdot m}\right)' \tag{39}$$

$$W_{\mathrm{m}}W_{\mathrm{c}}(\boldsymbol{\theta}) = \frac{1}{C(T-M)}\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}} \left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{cm}\right)\left(\boldsymbol{\theta}^{r}_{cm} - \bar{\boldsymbol{\theta}}^{\cdot}_{cm}\right)' \tag{40}$$

4.4 The Convergence Assessment Strategy

Define the following set of potential scale reduction factors, for a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, using (15)–(18) and (37)–(40):

$$PSRF_1(\theta_i) = \frac{\widehat{V}(\theta_i)}{W_{\mathrm{c}}(\theta_i)} \tag{41}$$

$$PSRF_2(\theta_i) = \frac{W_{\mathrm{m}}(\theta_i)}{W_{\mathrm{m}}W_{\mathrm{c}}(\theta_i)} \tag{42}$$

$$MPSRF_1(\boldsymbol{\theta}) = \text{maximum eigenvalue of } [W_{\mathrm{c}}(\boldsymbol{\theta})]^{-1}\,\widehat{V}(\boldsymbol{\theta}) \tag{43}$$

$$MPSRF_2(\boldsymbol{\theta}) = \text{maximum eigenvalue of } [W_{\mathrm{m}}W_{\mathrm{c}}(\boldsymbol{\theta})]^{-1}\,W_{\mathrm{m}}(\boldsymbol{\theta}). \tag{44}$$

By Lemmas 1 and 2, we have that

$$MPSRF_1(\boldsymbol{\theta}) \geq \max_i PSRF_1(\theta_i) \text{ and } MPSRF_2(\boldsymbol{\theta}) \geq \max_i PSRF_2(\theta_i). \tag{45}$$

Our convergence assessment technique consists of the following steps:

RJMCMC Convergence Assessment: Implement the following procedure as a convergence assessment technique for RJMCMC applied to a model with parameters Θ, using (15)–(18), (37)–(40) and (41)–(44):

1. Identify a parameter $k \in \Theta$ which is an indicator of "model" and select a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \subset \Theta$ consisting of quantities which have the same interpretation across k (but with $k \notin \boldsymbol{\theta}$).

2. Simulate $C > 1$ chains of equal length T via RJMCMC, with over-dispersed starting values.

3. Choose a base batch size b (Brooks and Gelman (1998) recommend, for example, $b \approx \frac{T}{20}$).

4. Let the notation $S^{(q)}(\cdot)$ represent a statistic S computed for the q^{th} batch

$$(\boldsymbol{\theta}_1^{(qb+1)}, \ldots, \boldsymbol{\theta}_1^{(2qb)}), \ \ldots \ , (\boldsymbol{\theta}_C^{(qb+1)}, \ldots, \boldsymbol{\theta}_C^{(2qb)})$$

   For batches $q = 1, \ldots, \frac{T}{b}$, do the following:

   (a) Plot $MPSRF_1{}^{(q)}(\boldsymbol{\theta})$ vs. q and $MPSRF_2{}^{(q)}(\boldsymbol{\theta})$ vs. q (separately or together).

   (b) Plot the maximum eigenvalues of $\widehat{V}^{(q)}(\boldsymbol{\theta})$ and $W_{\text{c}}{}^{(q)}(\boldsymbol{\theta})$ together vs. q.

   (c) Plot the maximum eigenvalues of $W_{\text{m}}{}^{(q)}(\boldsymbol{\theta})$ and $W_{\text{m}}W_{\text{c}}{}^{(q)}(\boldsymbol{\theta})$ together vs. q.

   (d) Optionally plot $PSRF_1{}^{(q)}(\theta_i)$ vs. q and $PSRF_2{}^{(q)}(\theta_i)$ vs. q.

   (e) Optionally plot the numerator and denominator of $PSRF_1{}^{(q)}(\theta_i)$ together vs. q.

   (f) Optionally plot the numerator and denominator of $PSRF_2{}^{(q)}(\theta_i)$ together vs. q.

5. Determine q_0 such that for $q \geq q_0$ the plots in Step 4a have settled close to 1, *and* the plots in Step 4b have settled approximately to a common value, *and* the plots in Step 4c have settled approximately to a common value.

6. Discard the first $q_0 b$ sweeps from each chain, and then pool the remaining ones together to use for inference.

We prefer the maximum eigenvalue to the determinant for monitoring individual matrices, since it is on the same scale as the univariate variance estimates and hence can conveniently be displayed in the same plot. The method can be performed on more than one parameter vector $\boldsymbol{\theta}$. It may be useful to use a collection of related sets of scalar parameters in order to target which sets are mixing faster than others. The purpose of the $MPSRF$ is to provide a safe (conservative) alternative to the monitoring of a large number of scalar parameters individually. However, the individual scalar parameters can still be monitored (Steps 4d–4f), providing more detailed information.

5 Example: Convergence Assessment for a Bivariate Normal Mixture with Unknown Number of Components

We demonstrate the use of our proposed convergence assessment technique on a situation discussed in Castelloe (1999) and Castelloe (1998). The research goal is to estimate the features of clusters in a Poisson cluster processes, modeled as a bivariate normal mixture with common but arbitrary covariance. The number and location of cluster centers and allocation information are unobserved. In the estimation of properties of mixture components, proper adjustment for uncertainty about the number of components is difficult, but an RJMCMC approach provides a solution.

5.1 Setup

Consider a Poisson cluster process on a region $A \in \Re^2$ with bivariate normal dispersion of offspring around parents, where the only observed data is the offspring locations $Y = [\mathbf{y}_1, ..., \mathbf{y}_n]'$. This process can be represented as a bivariate normal mixture with k components (i.e., clusters), equal mixing proportions, and common covariance matrix. The component means (i.e., cluster centers) $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k]'$ are distributed uniformly and independently on A. Component allocations (i.e., parent identifiers) are given by $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)' \overset{\text{i.i.d.}}{\sim} \text{Mult}\left(1, \frac{1}{k}\mathbf{1}\right)$. The displacements of offspring from parents, $\mathbf{y}_1 - \boldsymbol{\mu}_{\mathbf{z}_1}, \ \cdots \ , \mathbf{y}_n - \boldsymbol{\mu}_{\mathbf{z}_n}$, are independently and identically distributed as $N(\mathbf{0}, \boldsymbol{\Sigma})$, conditional on being confined to A.

The primary goal is to estimate $\boldsymbol{\Sigma}$, which characterizes the common shape, orientation, and volume of the (elliptical) clusters. The parameter $\boldsymbol{\Sigma}$ can be represented vectorially as the variances and covariance of the offspring displacement coordinates, $\boldsymbol{\sigma} = (\sigma_{11}, \ \sigma_{22}, \ \sigma_{12})'$. A normalized version is

$$(\log \sigma_{11}, \ \log \sigma_{22}, \ z(\rho_{12}))' \text{ where } z(\rho_{12}) = \text{ Fisher's z-transformation of } \sigma_{12}(\sigma_{11}\sigma_{22})^{-\frac{1}{2}}$$

An RJMCMC sampler can be constructed (see Castelloe (1999)) to simulate from the joint posterior distribution of $\{k, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{Z}\}$.

5.2 Choice of Parameters to Monitor

In MCMC convergence assessment it is recommended that, if feasible, all parameters are monitored, and if not, then at least one representative parameter of each "type" is monitored. The output of our RJMCMC sampler consists of $k^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$, $\boldsymbol{\mu}^{(t)}$, and $\mathbf{Z}^{(t)}$ for each sweep t. Firstly, k and $\boldsymbol{\Sigma}$ can be monitored easily, as these parameters retain the same meaning from sweep to sweep. Label-switching complicates efforts to monitor individual components of $\boldsymbol{\mu}$ and $\mathbf{Z}$. However, we have devised an approach to monitor a combination of $\boldsymbol{\mu}$ and $\mathbf{Z}$ which *is* identifiable. A certain number of offspring are "marked," and the parent locations of these offspring are tracked from sweep to sweep. We choose to monitor 3 offspring, chosen as:

1. an event near the center of a clearly defined cluster,

2. an event located between 2 clusters that are potential competitors for ownership of this event, and

3. an isolated event that could potentially be the sole member of a cluster, or an outlier in another cluster.

The purpose of these particular choices is to attempt to monitor parent locations that are expected to fluctuate across sweeps in different ways. This approach essentially boils down to monitoring $\boldsymbol{\mu}_{\mathbf{z}_{j_1}}$, $\boldsymbol{\mu}_{\mathbf{z}_{j_2}}$ and $\boldsymbol{\mu}_{\mathbf{z}_{j_3}}$ (6 scalar parameters in all) for 3 chosen offspring j_1, j_2 and j_3. These quantities retain the same meaning from sweep to sweep, and they represent instances of both parent locations and offspring allocations. We emphasize that the choice of offspring to track can be made after the sampler is run, since we are only using the usual sampler output.

To conserve space, results are reported in this paper only for k and $\boldsymbol{\Sigma}$. Results for $\boldsymbol{\mu}_{\mathbf{z}_{j_1}}$, $\boldsymbol{\mu}_{\mathbf{z}_{j_2}}$ and $\boldsymbol{\mu}_{\mathbf{z}_{j_3}}$ are similar, leading to the same conclusions.

5.3 Data Sets

We analyze two point patterns, one chosen for its resulting well-behaved convergence and the other for its slow convergence. The first (referred to as the "redwood pattern" and shown in Figure 1) consists of Redwood seedling locations as given in Strauss (1975). This pattern was deemed to produce favorable convergence properties in Castelloe (1998). The second (referred to as the "hex pattern" and shown in Figure 2) is a realization of a Poisson cluster process with bivariate normal dispersion, constructed to encourage competing explanations of the clustering, and presumably poor convergence. The dispersion parameters for the hex pattern are ${\sigma_{11}}^2 = {\sigma_{22}}^2 = 0.002$ and $\sigma_{12} = 0$. A sampler might "see" one large cluster, or nineteen small circular clusters, or several elongated clusters making 0, 60, or 120-degree angles with the positive x-axis. In Figures 1 and 2, the offspring whose parent locations are to be monitored in convergence assessment are marked as "1", "2" and "3."

5.4 Trace Plots

Before we proceed to the application of our new technique, we present some trace plots for each of the two data sets, showing the value of each scalar parameter chosen to monitor. It is not possible to ascertain "convergence" from such plots, but they can be helpful in revealing any major problems.

Three chains are produced from an RJMCMC sampler for each data set, each of length $20,000$ (runs of length $200,000$ where every 10^{th} sweep is saved).

Figure 3 shows trace plots of k, σ_{11}, σ_{22} and σ_{12} for the entire redwood chain #2, and Figure 4 shows the same for the entire hex chain#2. These trace plots are useful for demonstrating the behavior of the sampler for the full duration of sweeps.

Figures 5 and 6 show “snapshots” of the cluster allocations for several values of k, for redwood chain #2 and hex chain #2. The sweeps used were the last ones with k equal to the given value.

Figures 7 through 9 show trace plots for sweeps 1001-2000 (of the original 200,000) of all three redwood chains. Figures 11 through 13 show the same for all three hex chains. These trace plots are useful for demonstrating differences among the three chains for each data set. Figures 10 and 14 show histograms of k for the same chain portions (sweeps 1001-2000). These histograms provide a more useful view of the differences in the distribution of k among chains.

5.5 Convergence Plots

Figures 15 and 16 show convergence plots for the redwood data and hex pattern, constructed according to the technique in section 4.4 and based on the collection $\{\log \sigma_{11}, \log \sigma_{22}, z(\rho_{12})\}$. Each chain has length $20,000$, resulting from a run of length $200,000$ with every 10^{th} sweep saved.

Figure 17 shows the same displays for an additional set of three chains from an RJMCMC sampler run for $20,000$ sweeps on the hex pattern, with every sweep saved. So, Figure 17 is essentially a replication of Figure 16 zooming in on the first 10% of the sampler sweeps.

6 Discussion

A perusal of the trace plots and histograms for sweeps $1,000 - 2,000$ (Figures 7– 14) is sufficient to strongly doubt convergence in these eary sweeps for the hex pattern, and to moderately doubt convergence for the redwood pattern. For the hex pattern, the trace plots for chains #1 and #2 have clearly not settled down, and the histograms of k reveal that the chains visit models with markedly different frequencies. For the redwood data, the trace plot for chain #2 has not settled down, and the chains differ in their frequency of visiting the $k = 7$ model.

Trace plots (Figures 3 and 4) and histograms for the full $200,000$ sweeps, however, do not convey any indication of lack of convergence for either sampler. More sophisticated diagnostics are needed.

The convergence plots for the redwood pattern (Figure 15) imply that convergence is likely to have occurred by the 13^{th} batch. The MPSRF’s stay below 1.03, and the PSRF’s stay below 1.02. Pairs of max e.v.’s stay very close together throughout, although the pairs for $\log \sigma_{11}$ and $\log \sigma_{22}$ do not settle down to a common value until approximately the 13^{th} batch.

The convergence plots for the original hex pattern (Figure 16) cast serious doubt on convergence. MP-

SRF's for $\widehat{V}$ vs. W_c start out above 2 and surpass 1.2 as late as the 6th batch. PSRF's for $\log \sigma_{22}$ are just below the MPSRF's; thus, $\log \sigma_{22}$ is the component of $\mathbf{\Sigma}$ most responsible for the poor convergence. Pairs of max e.v.'s get closer together over time but stay fairly separated until the last couple batches. The members of each pair never settle to a common value. For $\widehat{V}$ vs. W_c the primary convergence violation is separation, and for W_m vs. $W_\mathrm{m} W_\mathrm{c}$ the main violation is failure to settle to a common value. (Reminder: in the max e.v. plots, a pair comparison always involves two lines with the same linetype; sometimes one of the members of a pair gets close to the multivariate maximum and gives the misleading impression of values converging together).

Plots for the additional hex pattern with all $20,000$ sweeps saved (Figure 17) show even more dramatic lack of convergence, especially with regard to the max e.v.'s of $\widehat{V}$ vs. W_c for $\log \sigma_{22}$.

7 Conclusions and Future Directions

The convergence assessment method outlined in section 4.4 offers a new approach for diagnosing convergence in reversible jump MCMC, an important extension of standard MCMC requiring more complex diagnostics due to the changing parameter space. We have demonstrated both its sensitivity and specificity in examples with two RJMCMC samplers. The plots display relevant warning signals for a sampler intentionally constructed to yield poor convergence, and they indicate well-behaved convergence for a sampler deemed to possess favorable mixing properties. While this of course is not conclusive evidence of the new technique's effectiveness, it is nevertheless a promising demonstration, bolstered by solid theoretical motivation and derivation.

Future research can hone in on the different types of "lack of convergence" such as variation between chains, interaction between models and chains, and variation in model visit frequencies across chains, to see how well the different components of the method detect them. In addition, the asymptotic behavior of the diagnostics for well-behaved samplers can be investigated as the number of sweeps goes to infinity, to determine whether the specificity is maintained.

8 Acknowledgments

Much of this research was conducted while the first author was a student at the Department of Statistics and Actuarial Science, University of Iowa. Thanks to Kate Cowles and Russell Lenth for valuable assistance.

9 Appendix: Derivation of Expectations of Variation Estimates Used in Convergence Assessment

First, we (trivially) have (24) by Winer (1971, p. 163) and (29) by Winer (1971, p. 325).

Proof of (23): Consider ANOVA 1 in Table 1. From Winer (1971, p. 165), we have that

$$E\mathrm{MS}_{\mathrm{ch}} = T\sigma^2_{\mathrm{ch}} + \sigma^2_{\mathrm{er(ch)}}$$

and so

$$\begin{aligned}
E\widehat{V} &= E\mathrm{MS}_{\mathrm{tot}} \\
&= \frac{1}{CT-1} E\mathrm{SS}_{\mathrm{tot}} \\
&= \frac{1}{CT-1} E\left[\mathrm{SS}_{\mathrm{ch}} + \mathrm{SS}_{\mathrm{er(ch)}}\right] \\
&= \frac{1}{CT-1}\left[(C-1)E\mathrm{MS}_{\mathrm{ch}} + C(T-1)E\mathrm{MS}_{\mathrm{er(ch)}}\right] \\
&= \sigma^2_{\mathrm{er(ch)}} + \frac{(C-1)T}{CT-1}\sigma^2_{\mathrm{ch}} \quad \square
\end{aligned}$$

Proof of (25): Consider ANOVA 2 and ANOVA 3 in Tables 2 – 3. The quantity $\mathrm{SS}_{\mathrm{er(mo)}}$ in ANOVA 2 can be re-written as follows:

$$\begin{aligned}
\mathrm{SS}_{\mathrm{er(mo)}} &= \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{\cdot m}\right)^2 \\
&= \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left[\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right) + \left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right) + \left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right)\right]^2 \\
&= \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right)^2 + \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2 + \\
&\quad \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2 + \\
&\quad 2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right)\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right) + \\
&\quad 2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right)\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right) + \\
&\quad 2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right).
\end{aligned}$$

Each term can then be simplified:

$$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^r_{cm} - \bar{\theta}^{\cdot}_{cm}\right)^2 = \mathrm{SS}_{\mathrm{er(ch\times mo)}},$$

$$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2=\mathrm{SS}_{\mathrm{ch}},$$

$$\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2=\mathrm{SS}_{\mathrm{ch\times mo}},$$

$$\begin{aligned}
&2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^{r}_{cm}-\bar{\theta}^{\cdot}_{cm}\right)\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\\
&\quad= 2\sum_{c=1}^{C}\sum_{m=1}^{M}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left[\sum_{r=1}^{R_{cm}}\left(\theta^{r}_{cm}-\bar{\theta}^{\cdot}_{cm}\right)\right]\\
&\quad= 2\sum_{c=1}^{C}\sum_{m=1}^{M}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)[0]\\
&\quad= 0,
\end{aligned}$$

$$\begin{aligned}
&2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\theta^{r}_{cm}-\bar{\theta}^{\cdot}_{cm}\right)\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\\
&\quad= 2\sum_{c=1}^{C}\sum_{m=1}^{M}\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left[\sum_{r=1}^{R_{cm}}\left(\theta^{r}_{cm}-\bar{\theta}^{\cdot}_{cm}\right)\right]\\
&\quad= 2\sum_{c=1}^{C}\sum_{m=1}^{M}\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)[0]\\
&\quad= 0,
\end{aligned}$$

and finally, using the notation

$$\theta^{\cdot}_{cm}=\sum_{r=1}^{R_{cm}}\theta^{r}_{cm},\ \theta^{\cdot}_{c\cdot}=\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm},\ \theta^{\cdot}_{\cdot m}=\sum_{c=1}^{C}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm},\ \text{ and }\theta^{\cdot}_{\cdot\cdot}=\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\theta^{r}_{cm},$$

$$\begin{aligned}
&2\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\\
&\quad= 2\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left[\sum_{m=1}^{M}\left\{R_{cm}\left(\bar{\theta}^{\cdot}_{cm}-\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot m}+\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\right\}\right]\\
&\quad= 2\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left[\sum_{m=1}^{M}\left(\theta^{\cdot}_{cm}-\frac{1}{M}\theta^{\cdot}_{c\cdot}-\frac{1}{C}\theta^{\cdot}_{\cdot m}+\frac{1}{CM}\theta^{\cdot}_{\cdot\cdot}\right)\right]\\
&\quad= 2\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot}-\bar{\theta}^{\cdot}_{\cdot\cdot}\right)\left[\theta^{\cdot}_{c\cdot}-\theta^{\cdot}_{c\cdot}-\frac{1}{C}\theta^{\cdot}_{\cdot\cdot}+\frac{1}{C}\theta^{\cdot}_{\cdot\cdot}\right]\\
&\quad= 0.
\end{aligned}$$

Therefore,

$$\mathrm{SS}_{\mathrm{er(mo)}}=\mathrm{SS}_{\mathrm{er(ch\times mo)}}+\mathrm{SS}_{\mathrm{ch}}+\mathrm{SS}_{\mathrm{ch\times mo}}.$$

Now we derive the expected mean-squares for chain and chain×model, using terminology from ANOVA 3 (Table 3). Several steps in the derivations use the following ANOVA assumptions (and not always with explicit reference):

$$\begin{aligned}&\sum_{m=1}^{M}\beta_m = 0,\ E\alpha_c = 0,\ E(\alpha\beta)_{cm} = 0,\ Ee^r_{cm} = 0,\ \text{ and}\\ &\text{all }\ \{\alpha_c\}, \{(\alpha\beta)_{cm}\}, \{e^r_{cm}\}\ \text{ are mutually independent.}\end{aligned} \tag{46}$$

For simplicity of notation, let $e^r_{cm} \equiv e^r_{cm(3)}$.

First, $E\mathrm{MS}_{\mathrm{ch}}$:

$$\begin{aligned}
E\mathrm{MS}_{\mathrm{ch}} = {} & \frac{1}{C-1}E\left\{T\sum_{c=1}^{C}\left(\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2\right\}\\
= {} & \frac{1}{C-1}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot\cdot}\right\}^2\\
= {} & \frac{1}{C-1}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\frac{1}{T}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}\left(\mu + \alpha_c + \beta_{m'} + (\alpha\beta)_{cm'} + e^{r'}_{cm'}\right) - \right.\\
& \left.\frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}\left(\mu + \alpha_{c'} + \beta_{m'} + (\alpha\beta)_{c'm'} + e^{r'}_{c'm'}\right)\right\}^2\\
= {} & \frac{1}{C-1}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\left(\mu + \alpha_c + \frac{1}{T}\sum_{m'=1}^{M}R_{cm'}\beta_{m'} + \frac{1}{T}\sum_{m'=1}^{M}R_{cm'}(\alpha\beta)_{cm'} + \right.\right.\\
& \left.\frac{1}{T}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}e^{r'}_{cm'}\right) -\\
& \left(\mu + \frac{1}{C}\sum_{c'=1}^{C}\alpha_{c'} + \frac{1}{CT}\sum_{m'=1}^{M}R_{\cdot m'}\beta_{m'} + \frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}R_{c'm'}(\alpha\beta)_{c'm'} + \right.\\
& \left.\left.\frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}e^{r'}_{c'm'}\right)\right\}^2\\
= {} & \frac{1}{C-1}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\left(\alpha_c - \frac{1}{C}\sum_{c'=1}^{C}\alpha_{c'}\right) + \right.\\
& \left(\frac{1}{T}\sum_{m'=1}^{M}R_{cm'}\beta_{m'} - \frac{1}{CT}\sum_{m'=1}^{M}R_{\cdot m'}\beta_{m'}\right) +\\
& \left(\frac{1}{T}\sum_{m'=1}^{M}R_{cm'}(\alpha\beta)_{cm'} - \frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}R_{c'm'}(\alpha\beta)_{c'm'}\right) +\\
& \left.\left(\frac{1}{T}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}e^{r'}_{cm'} - \frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}}e^{r'}_{c'm'}\right)\right\}^2\\
= {} & \frac{1}{C-1}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\left(\alpha_c - \frac{1}{C}\sum_{c'=1}^{C}\alpha_{c'}\right)^2 + \right.\\
& \left(\frac{1}{CT}\sum_{m'=1}^{M}\left[CR_{cm'} - R_{\cdot m'}\right]\beta_{m'}\right)^2 +
\end{aligned}$$

$$\left(\frac{1}{T}\sum_{m'=1}^{M} R_{cm'}(\alpha\beta)_{cm'} - \frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M} R_{c'm'}(\alpha\beta)_{c'm'}\right)^2 +$$

$$\left.\left(\frac{1}{T}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}} e^{r'}_{cm'} - \frac{1}{CT}\sum_{c'=1}^{C}\sum_{m'=1}^{M}\sum_{r'=1}^{R_{cm}} e^{r'}_{c'm'}\right)^2\right\}$$

(all sums of cross-products are 0, by (46))

$$= \frac{1}{C-1} E \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{(\alpha_c - \bar{\alpha}.)^2 + \right.$$

$$\left(\frac{1}{C^2T^2}\sum_{m'=1}^{M}\left[CR_{cm'} - R_{\cdot m'}\right]^2 \beta^2_{m'}\right) +$$

$$\left(\frac{1}{T^2}\sum_{m'=1}^{M} R^2_{cm'}(\alpha\beta)^2_{cm'} - \frac{2}{CT^2}\sum_{m'=1}^{M} R^2_{cm'}(\alpha\beta)^2_{cm'} + \right.$$

$$\left.\left.\frac{1}{C^2T^2}\sum_{c'=1}^{C}\sum_{m'=1}^{M} R^2_{c'm'}(\alpha\beta)^2_{c'm'}\right) + (\bar{e}^{\cdot}_{c\cdot} - \bar{e}^{\cdot}_{\cdot\cdot})^2\right\}$$

$$= \frac{1}{C-1} E \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{(\alpha_c - \bar{\alpha}.)^2 + \frac{1}{C^2T^2}\sum_{m'=1}^{M}\left[CR_{cm'} - R_{\cdot m'}\right]^2 \beta^2_{m'} + \right.$$

$$\frac{1}{CT^2}\sum_{m'=1}^{M}\left[CR^2_{cm'} - 2R^2_{cm'}\right](\alpha\beta)^2_{cm'} +$$

$$\left.\frac{1}{C^2T^2}\sum_{c'=1}^{C}\sum_{m'=1}^{M} R^2_{c'm'}(\alpha\beta)^2_{c'm'} + (\bar{e}^{\cdot}_{c\cdot} - \bar{e}^{\cdot}_{\cdot\cdot})^2\right\}$$

$$= \frac{1}{C-1}\left\{T(C-1)\sigma^2_{\text{ch}} + \frac{T}{C^2T^2}\sum_{m'=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm'} - R_{\cdot m'})^2\right]\beta^2_{m'} + \right.$$

$$T\frac{C-2}{CT^2}\sum_{m'=1}^{M}\sum_{c=1}^{C} R^2_{cm'} E(\alpha\beta)^2_{cm'} + \frac{1}{CT}\sum_{m'=1}^{M}\sum_{c'=1}^{C} R^2_{c'm'} E(\alpha\beta)^2_{c'm'} +$$

$$\left.T(C-1)\frac{\sigma^2_{\text{er(ch}\times\text{mo)}}}{T}\right\}$$

$$= T\sigma^2_{\text{ch}} + \frac{1}{(C-1)C^2T}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm} - R_{\cdot m})^2\right]\beta^2_m +$$

$$\left[\frac{1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\right]\sigma^2_{\text{ch}\times\text{mo}} + \sigma^2_{\text{er(ch}\times\text{mo)}} \quad .$$

Next, $E\text{MS}_{\text{ch}\times\text{mo}}$:

$$E\text{MS}_{\text{ch}\times\text{mo}} = \frac{1}{(C-1)(M-1)} E\left\{\sum_{c=1}^{C}\sum R_{cm}\left(\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right)^2\right\}$$

$$= \frac{1}{(C-1)(M-1)} E \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\bar{\theta}^{\cdot}_{cm} - \bar{\theta}^{\cdot}_{c\cdot} - \bar{\theta}^{\cdot}_{\cdot m} + \bar{\theta}^{\cdot}_{\cdot\cdot}\right\}^2$$

$$= \frac{1}{(C-1)(M-1)} E \sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\left\{\frac{1}{R_{cm}}\sum_{r'=1}^{R_{cm}}\left(\mu + \alpha_c + \beta_m + (\alpha\beta)_{cm} + e^{r'}_{cm}\right) - \right.$$

$$
\begin{aligned}
& \frac{1}{T} \sum_{m'=1}^{M} \sum_{r'=1}^{R_{cm}} \left(\mu + \alpha_c + \beta_{m'} + (\alpha\beta)_{cm'} + e^{r'}_{cm'}\right) - \\
& \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} \sum_{r'=1}^{R_{cm}} \left(\mu + \alpha_{c'} + \beta_m + (\alpha\beta)_{c'm} + e^{r'}_{c'm}\right) + \\
& \left.\frac{1}{IT} \sum_{c'=1}^{C} \sum_{m'=1}^{M} \sum_{r'=1}^{R_{cm}} \left(\mu + \alpha_{c'} + \beta_{m'} + (\alpha\beta)_{c'm'} + e^{r'}_{c'm'}\right)\right\}^2 \\
= \; & \frac{1}{(C-1)(M-1)} E \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \{(\mu + \alpha_c + \beta_m + (\alpha\beta)_{cm} + \bar{e}^{\cdot}_{cm}) - \\
& \left(\mu + \alpha_c + \frac{1}{T} \sum_{m'=1}^{M} R_{cm'}\beta_{m'} + \frac{1}{T} \sum_{m'=1}^{M} R_{cm'}(\alpha\beta)_{cm'} + \bar{e}^{\cdot}_{c\cdot}\right) - \\
& \left(\mu + \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}\alpha_{c'} + \beta_m + \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}(\alpha\beta)_{c'm} + \bar{e}^{\cdot}_{\cdot m}\right) + \\
& \left.\left(\mu + \frac{1}{C} \sum_{c'=1}^{C} \alpha_{c'} + \frac{1}{CT} \sum_{m'=1}^{M} R_{\cdot m'}\beta_{m'} + \frac{1}{CT} \sum_{c'=1}^{C} \sum_{m'=1}^{M} R_{c'm'}(\alpha\beta)_{c'm'} + \bar{e}^{\cdot}_{\cdot\cdot}\right)\right\}^2 \\
= \; & \frac{1}{(C-1)(M-1)} E \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left\{\left(\frac{1}{C} \sum_{c'=1}^{C} \alpha_{c'} - \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}\alpha_{c'}\right) + \right. \\
& \left(\frac{1}{CT} \sum_{m'=1}^{M} R_{\cdot m'}\beta_{m'} - \frac{1}{T} \sum_{m'=1}^{M} R_{cm'}\beta_{m'}\right) + \\
& \left((\alpha\beta)_{cm} - \frac{1}{T} \sum_{m'=1}^{M} R_{cm'}(\alpha\beta)_{cm'} - \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}(\alpha\beta)_{c'm} + \right. \\
& \left.\left.\frac{1}{CT} \sum_{c'=1}^{C} \sum_{m'=1}^{M} R_{c'm'}(\alpha\beta)_{c'm'}\right) + (\bar{e}^{\cdot}_{cm} - \bar{e}^{\cdot}_{c\cdot} - \bar{e}^{\cdot}_{\cdot m} + \bar{e}^{\cdot}_{\cdot\cdot})\right\}^2 \\
= \; & \frac{1}{(C-1)(M-1)} E \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \left\{\left(\frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}\alpha_{c'} - \frac{1}{C} \sum_{c'=1}^{C} \alpha_{c'}\right)^2 + \right. \\
& \left(\frac{1}{T} \sum_{m'=1}^{M} R_{cm'}\beta_{m'} - \frac{1}{CT} \sum_{m'=1}^{M} R_{\cdot m'}\beta_{m'}\right)^2 + \\
& \left((\alpha\beta)_{cm} - \frac{1}{T} \sum_{m'=1}^{M} R_{cm'}(\alpha\beta)_{cm'} - \frac{1}{R_{\cdot m}} \sum_{c'=1}^{C} R_{c'm}(\alpha\beta)_{c'm} + \right. \\
& \left.\left.\frac{1}{CT} \sum_{c'=1}^{C} \sum_{m'=1}^{M} R_{c'm'}(\alpha\beta)_{c'm'}\right)^2 + (\bar{e}^{\cdot}_{cm} - \bar{e}^{\cdot}_{c\cdot} - \bar{e}^{\cdot}_{\cdot m} + \bar{e}^{\cdot}_{\cdot\cdot})^2\right\} \\
& \quad \text{(all sums of cross-products are 0, by (46))}
\end{aligned}
$$

At this point we simplify the last portion:

$$
\begin{aligned}
& E\left(\bar{e}^{\cdot}_{cm} - \bar{e}^{\cdot}_{c\cdot} - \bar{e}^{\cdot}_{\cdot m} + \bar{e}^{\cdot}_{\cdot\cdot}\right)^2 \\
& \quad = \; E\left[\bar{e}^{\cdot}_{cm} - \frac{1}{M} \sum_{m'=1}^{M} \bar{e}^{\cdot}_{cm'} - \frac{1}{C} \sum_{c'=1}^{C} \bar{e}^{\cdot}_{c'm} + \frac{1}{CM} \sum_{c'=1}^{C} \sum_{m'=1}^{M} \bar{e}^{\cdot}_{c'm'}\right]^2
\end{aligned}
$$

$$\begin{aligned}
= \quad & E\Bigg[(\bar{e}^{\cdot}_{cm})^2 + \frac{1}{M^2}\sum_{m'=1}^{M}(\bar{e}^{\cdot}_{cm'})^2 + \frac{1}{C^2}\sum_{c'=1}^{C}(\bar{e}^{\cdot}_{c'm})^2 + \frac{1}{C^2M^2}\sum_{c'=1}^{C}\sum_{m'=1}^{M}(\bar{e}^{\cdot}_{c'm'})^2 - \\
& \frac{2}{M}(\bar{e}^{\cdot}_{cm})^2 - \frac{2}{C}(\bar{e}^{\cdot}_{cm})^2 + \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2 + \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2 - \frac{2}{CM^2}\sum_{m'=1}^{M}(\bar{e}^{\cdot}_{cm'})^2 - \\
& \frac{2}{C^2M}\sum_{c'=1}^{C}(\bar{e}^{\cdot}_{c'm})^2\Bigg] \\
= \quad & E\Bigg[(\bar{e}^{\cdot}_{cm})^2 + \frac{1}{M}(\bar{e}^{\cdot}_{cm})^2 + \frac{1}{C}\bar{e}^{\cdot}_{cm})^2 + \frac{1}{CM}(\bar{e}^{\cdot}_{cm})^2 - \\
& \frac{2}{M}(\bar{e}^{\cdot}_{cm})^2 - \frac{2}{C}(\bar{e}^{\cdot}_{cm})^2 + \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2 + \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2 - \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2 - \\
& \frac{2}{CM}(\bar{e}^{\cdot}_{cm})^2\Bigg] \\
= \quad & \frac{1}{C^2M^2}\left[(CM)^2 + C^2M + CM^2 + CM - 2C^2M - 2CM^2 + \right. \\
& \left. 2CM + 2CM - 2CM - 2CM\right]E(\bar{e}^{\cdot}_{cm})^2 \\
= \quad & \frac{1}{C^2M^2}\left[(CM)^2 - C^2M - CM^2 + CM\right]\frac{\sigma^2_{\text{er}(\text{ch}\times\text{mo})}}{R_{cm}} \\
= \quad & \frac{(C-1)(M-1)}{R_{cm}CM}\sigma^2_{\text{er}(\text{ch}\times\text{mo})}
\end{aligned}$$

Now, resuming again with $E\text{MS}_{\text{ch}\times\text{mo}}$:

$$\begin{aligned}
& E\text{MS}_{\text{ch}\times\text{mo}} \\
= \quad & \frac{1}{(C-1)(M-1)}E\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{r=1}^{R_{cm}}\Bigg\{\left(\frac{1}{R^2_{\cdot m}C^2}\sum_{c'=1}^{C}\left(CR_{c'm} - R_{\cdot m}\right)^2\alpha^2_{c'}\right) + \\
& \left(\frac{1}{C^2T^2}\sum_{m'=1}^{M}\left(CR_{cm'} - R_{\cdot m'}\right)^2\beta^2_{m'}\right) + \\
& \Bigg((\alpha\beta)^2_{cm} + \frac{1}{T^2}\sum_{m'=1}^{M}R^2_{cm'}(\alpha\beta)^2_{cm'} + \frac{1}{R^2_{\cdot m}}\sum_{c'=1}^{C}R^2_{c'm}(\alpha\beta)_{c'm} + \\
& \frac{1}{C^2T^2}\sum_{c'=1}^{C}\sum_{m'=1}^{M}R^2_{c'm'}(\alpha\beta)^2_{c'm'} - \frac{2}{T}R_{cm}(\alpha\beta)^2_{cm} - \frac{2}{R_{\cdot m}}R_{cm}(\alpha\beta)^2_{cm} + \\
& \frac{2}{CT}R_{cm}(\alpha\beta)^2_{cm} + \frac{2}{TR_{\cdot m}}R^2_{cm}(\alpha\beta)^2_{cm} - \frac{2}{CT^2}\sum_{m'=1}^{M}R^2_{cm'}(\alpha\beta)^2_{cm'} - \\
& \frac{2}{R_{\cdot m}CT}\sum_{c'=1}^{C}R^2_{c'm}(\alpha\beta)_{c'm}\Bigg) + \left(\frac{(C-1)(M-1)}{R_{cm}CM}\sigma^2_{\text{er}(\text{ch}\times\text{mo})}\right)\Bigg\} \\
= \quad & \frac{1}{(C-1)(M-1)}\Bigg\{\left(\sum_{m=1}^{M}R_{\cdot m}\left[\frac{1}{R^2_{\cdot m}C^2}\sum_{c'=1}^{C}\left(CR_{c'm} - R_{\cdot m}\right)^2\sigma^2_{\text{ch}}\right]\right) + \\
& \left(T\sum_{c=1}^{C}\left[\frac{1}{C^2T^2}\sum_{m'=1}^{M}\left(CR_{cm'} - R_{\cdot m'}\right)^2\beta^2_{m'}\right]\right) +
\end{aligned}$$

$$
\begin{aligned}
& \Bigg(CT\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + T\sum_{c=1}^{C}\left[\frac{1}{T^2}\sum_{m'=1}^{M} R^2_{cm'}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] + \\
& \sum_{m=1}^{M} R_{\cdot m}\left[\frac{1}{R^2_{\cdot m}}\sum_{c'=1}^{C} R^2_{c'm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] + CT\left[\frac{1}{C^2T^2}\sum_{c'=1}^{C}\sum_{m'=1}^{M} R^2_{c'm'}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] - \\
& \sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left[\frac{2}{T}R_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] - \sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left[\frac{2}{R_{\cdot m}}R_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] + \\
& \sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left[\frac{2}{CT}R_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] + \sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left[\frac{2}{TR_{\cdot m}}R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] - \\
& T\sum_{c=1}^{C}\left[\frac{2}{CT^2}\sum_{m'=1}^{M} R^2_{cm'}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right] - \sum_{m=1}^{M} R_{\cdot m}\left[\frac{2}{R_{\cdot m}CT}\sum_{c'=1}^{C} R^2_{c'm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\right]\Bigg) + \\
& \left(\sum_{c=1}^{C}\sum_{m=1}^{M} R_{cm}\left[\frac{(C-1)(M-1)}{R_{cm}CM}\sigma^2_{\mathrm{er}(\mathrm{ch}\times\mathrm{mo})}\right]\right)\Bigg\} \\
= \quad & \frac{1}{(C-1)(M-1)}\Bigg\{\left(\frac{1}{C^2}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{\left(CR_{cm}-R_{\cdot m}\right)^2}{R_{\cdot m}}\sigma^2_{\mathrm{ch}}\right) + \\
& \left(\frac{1}{C^2T}\sum_{c=1}^{C}\sum_{m=1}^{M}\left(CR_{cm}-R_{\cdot m}\right)^2\beta^2_m\right) + \\
& \Bigg(CT\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \frac{1}{T}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \\
& \sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^2_{cm}}{R_{\cdot m}}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \frac{1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} - \\
& \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} - 2\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^2_{cm}}{R_{\cdot m}}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \\
& \frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^3_{cm}}{R_{\cdot m}}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} - \\
& \frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}} - \frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\sigma^2_{\mathrm{ch}\times\mathrm{mo}}\Bigg) + \\
& \left((C-1)(M-1)\sigma^2_{\mathrm{er}(\mathrm{ch}\times\mathrm{mo})}\right)\Bigg\} \\
= \quad & \left[\frac{1}{C^2(C-1)(M-1)}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{\left(CR_{cm}-R_{\cdot m}\right)^2}{R_{\cdot m}}\right]\sigma^2_{\mathrm{ch}} + \\
& \frac{1}{C^2T(C-1)(M-1)}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}\left(CR_{cm}-R_{\cdot m}\right)^2\right]\beta^2_m + \\
& \frac{1}{(C-1)(M-1)}\Bigg[CT + \frac{1}{T}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm} + \sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^2_{cm}}{R_{\cdot m}} + \frac{1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm} - \\
& \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm} - 2\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^2_{cm}}{R_{\cdot m}} + \frac{2}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm} + \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R^3_{cm}}{R_{\cdot m}} - \\
& \frac{4}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M} R^2_{cm}\Bigg]\sigma^2_{\mathrm{ch}\times\mathrm{mo}} + \sigma^2_{\mathrm{er}(\mathrm{ch}\times\mathrm{mo})}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \left[\frac{1}{C^2(C-1)(M-1)}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{(CR_{cm}-R_{\cdot m})^2}{R_{\cdot m}}\right]\sigma^2_{\text{ch}} + \\
& \frac{1}{C^2T(C-1)(M-1)}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm}-R_{\cdot m})^2\right]\beta_m^2 + \\
& \frac{1}{(C-1)(M-1)}\left[CT - \frac{C+1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}R_{cm}^2 - \sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^2}{R_{\cdot m}} + \right. \\
& \left. \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^3}{R_{\cdot m}}\right]\sigma^2_{\text{ch}\times\text{mo}} + \sigma^2_{\text{er(ch}\times\text{mo)}}
\end{aligned}
$$

Finally, putting it all together,

$$
\begin{aligned}
EW_{\text{m}} &= E\text{MS}_{\text{er(mo)}} \\
= \quad & \frac{1}{CT-M}E\left[\text{SS}_{\text{er(ch}\times\text{mo)}} + \text{SS}_{\text{ch}} + \text{SS}_{\text{ch}\times\text{mo}}\right] \\
= \quad & \frac{1}{CT-M}\left[C(T-M)E\text{MS}_{\text{er(ch}\times\text{mo)}} + (C-1)E\text{MS}_{\text{ch}} + \right. \\
& \left. (C-1)(M-1)E\text{MS}_{\text{ch}\times\text{mo}}\right] \\
= \quad & \frac{C(T-M)}{CT-M}\sigma^2_{\text{er(ch}\times\text{mo)}} + \\
& \frac{C-1}{CT-M}\left\{T\sigma^2_{\text{ch}} + \frac{1}{(C-1)C^2T}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm}-R_{\cdot m})^2\right]\beta_m^2 + \right. \\
& \left. \left[\frac{1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}R_{cm}^2\right]\sigma^2_{\text{ch}\times\text{mo}} + \sigma^2_{\text{er(ch}\times\text{mo)}}\right\} + \\
& \frac{(C-1)(M-1)}{CT-M}\left\{\left[\frac{1}{C^2(C-1)(M-1)}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{(CR_{cm}-R_{\cdot m})^2}{R_{\cdot m}}\right]\sigma^2_{\text{ch}} + \right. \\
& \frac{1}{C^2T(C-1)(M-1)}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm}-R_{\cdot m})^2\right]\beta_m^2 + \\
& \frac{1}{(C-1)(M-1)}\left[CT - \frac{C+1}{CT}\sum_{c=1}^{C}\sum_{m=1}^{M}R_{cm}^2 - \sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^2}{R_{\cdot m}} + \right. \\
& \left. \left. \frac{2}{T}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{R_{cm}^3}{R_{\cdot m}}\right]\sigma^2_{\text{ch}\times\text{mo}} + \sigma^2_{\text{er(ch}\times\text{mo)}}\right\} \\
= \quad & \left[\frac{C(T-M)+C-1+(C-1)(M-1)}{CT-M}\right]\sigma^2_{\text{er(ch}\times\text{mo)}} + \\
& \frac{1}{CT-M}\left[(C-1)T + \frac{1}{C^2}\sum_{c=1}^{C}\sum_{m=1}^{M}\frac{(CR_{cm}-R_{\cdot m})^2}{R_{\cdot m}}\right]\sigma^2_{\text{ch}} + \\
& \frac{2}{(CT-M)C^2T}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}(CR_{cm}-R_{\cdot m})^2\right]\beta_m^2 + \\
& \frac{1}{CT-M}\left[CT + \sum_{c=1}^{C}\sum_{m=1}^{M}\left(\frac{C-1}{CT} - \frac{C+1}{CT} - \frac{1}{R_{\cdot m}} + \frac{2R_{cm}}{TR_{\cdot m}}\right)R_{cm}^2\right]\sigma^2_{\text{ch}\times\text{mo}}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \sigma^2_{\mathrm{er(ch\times mo)}} + \left[\frac{(C-1)T}{CT-M} + \frac{1}{C^2(CT-M)} \sum_{c=1}^{C}\sum_{m=1}^{M} \frac{\left(CR_{cm} - R_{\cdot m}\right)^2}{R_{\cdot m}} \right] \sigma^2_{\mathrm{ch}} + \\
& \frac{2}{C^3T - C^2MT} \sum_{m=1}^{M} \left[\sum_{c=1}^{C} \left(CR_{cm} - R_{\cdot m}\right)^2 \right] \beta^2_m + \\
& \Bigg[\frac{CT}{CT-M} + \frac{-1}{CT-M} \sum_{c=1}^{C}\sum_{m=1}^{M} \frac{R^2_{cm}}{R_{\cdot m}} + \\
& \frac{2}{CT} \sum_{c=1}^{C}\sum_{m=1}^{M} \left(CR_{cm} - R_{\cdot m}\right) \frac{R^2_{cm}}{R_{\cdot m}} \Bigg] \sigma^2_{\mathrm{ch\times mo}} \; . \quad \square
\end{aligned}
$$

References

Berger, J. O. et al., eds. (1992). *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, April 15-20, 1991.* Oxford University Press, New York.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

Brooks, S. P. and Giudici, P. (2000). MCMC convergence assessment via two-way ANOVA. *Journal of Computational and Graphical Statistics* **9**, 266–285.

Castelloe, J. (1999). Reversible jump Markov chain Monte Carlo analysis of spatial Poisson cluster processes with bivariate normal displacement. *Computing Science and Statistics* **31**, 306–315.

Castelloe, J. M. (1998). *Issues in Reversible Jump Markov Chain Monte Carlo and Composite EM Analysis, Applied to Spatial Poisson Cluster Processes.* Ph. D. thesis, University of Iowa.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**, 883–904.

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

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences. *Statistcal Science* (7), 457–511.

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

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Berger et al. (1992), pp. 169–188.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

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

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, Inc., London.

Mengersen, K. L., Robert, C. P., and Guihenneuc-Jouyaux, C. (1998). MCMC convergence diagnostics: a “reviewww”. CREST-INSEE, Paris (in press).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B, Methodological* **59**, 731–792.

Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–475.

Winer, B. J. (1971). *Statistical Principles in Experimental Design* (2nd ed.). McGraw-Hill Book Company, Inc., New York.

10 Figures

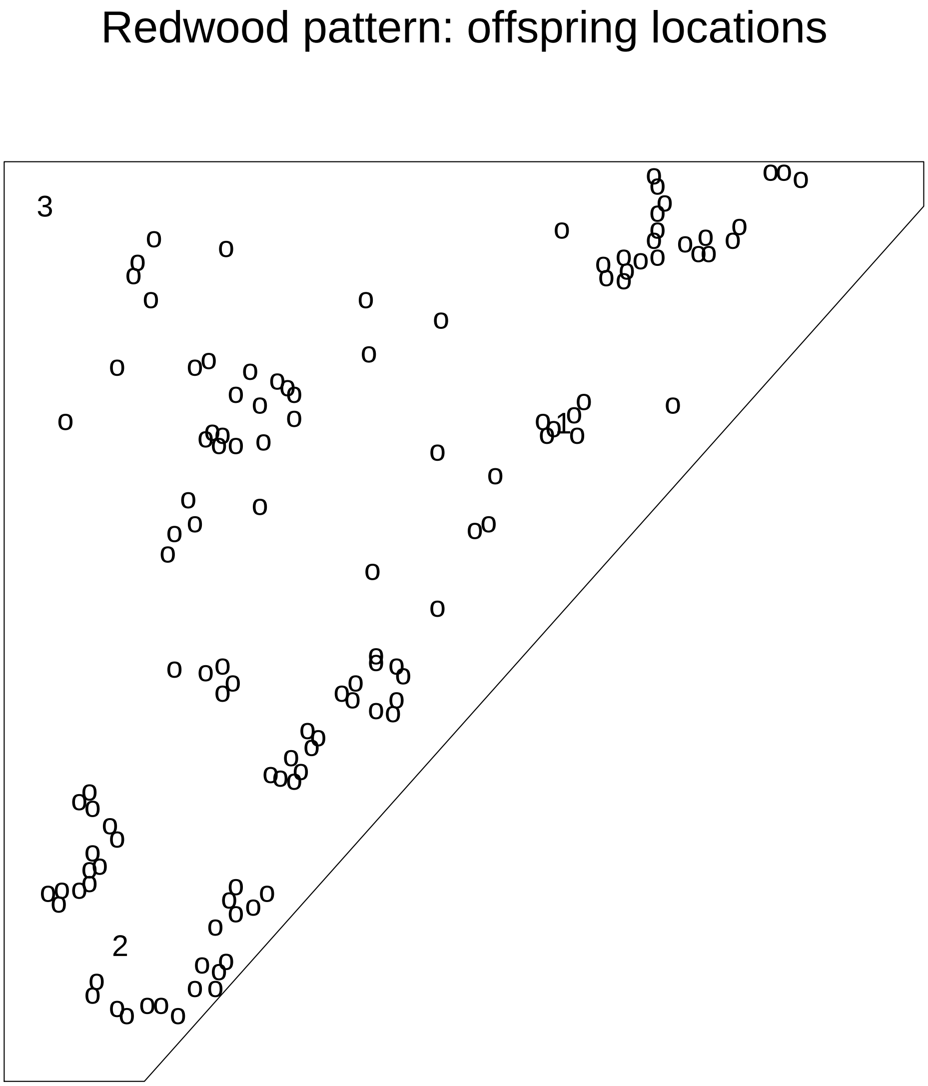


Figure 1: Locations of Redwood seedlings, with tracked offspring marked.

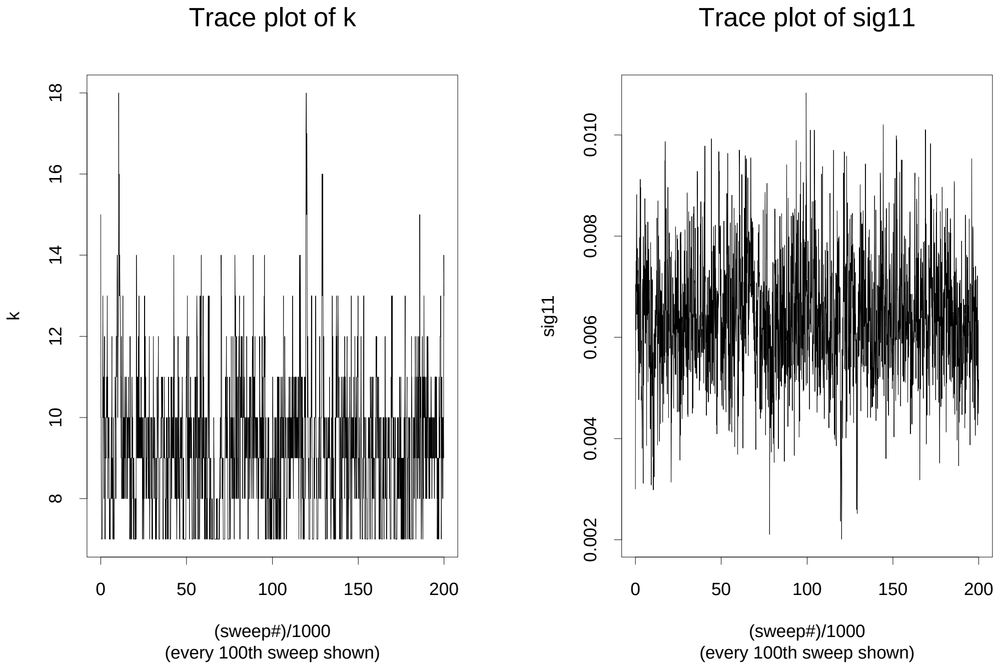

Hex pattern: offspring locations

Hex pattern: parent locations

(a) Locations of offspring, with tracked offspring marked

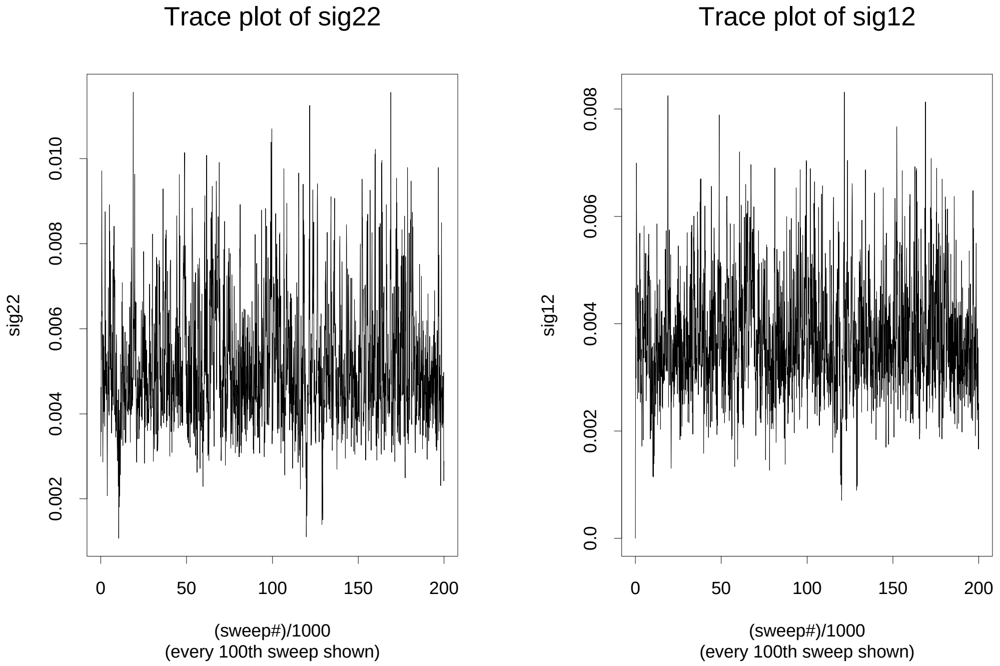
(b) Locations of parent events

Figure 2: Hex pattern



(a) k

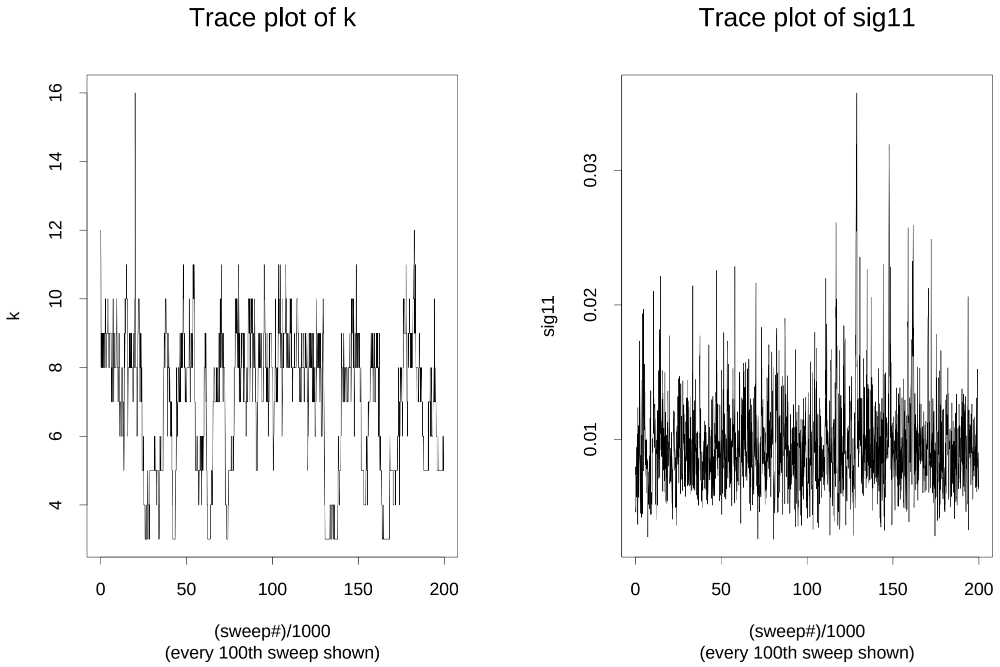
(b) σ_{11}



(c) σ_{22}

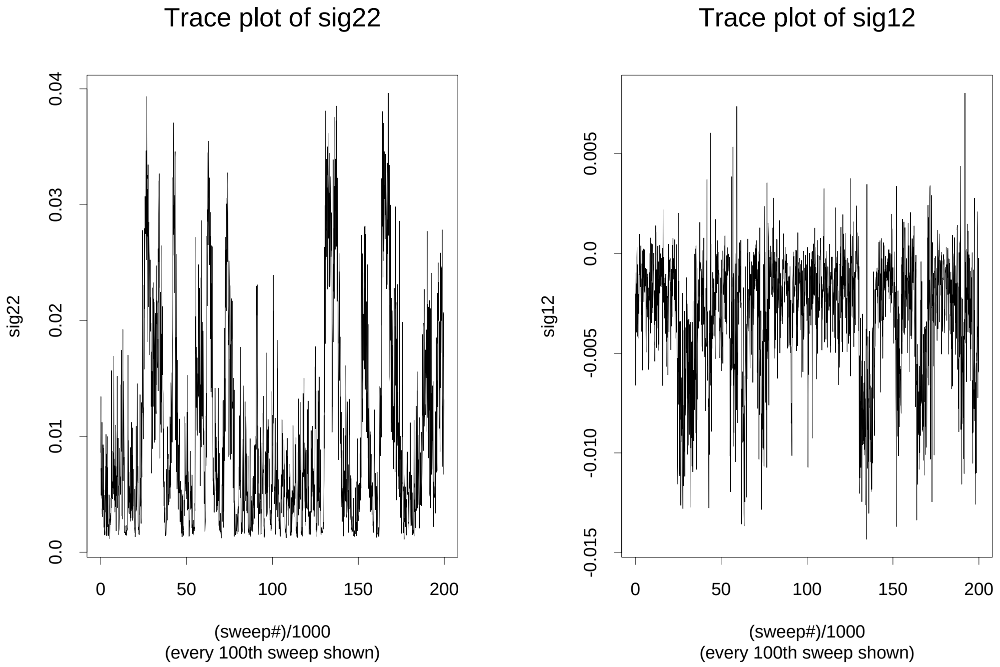
(d) σ_{12}

Figure 3: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for redwood chain #2 using all 200,000 sweeps.



(a) k

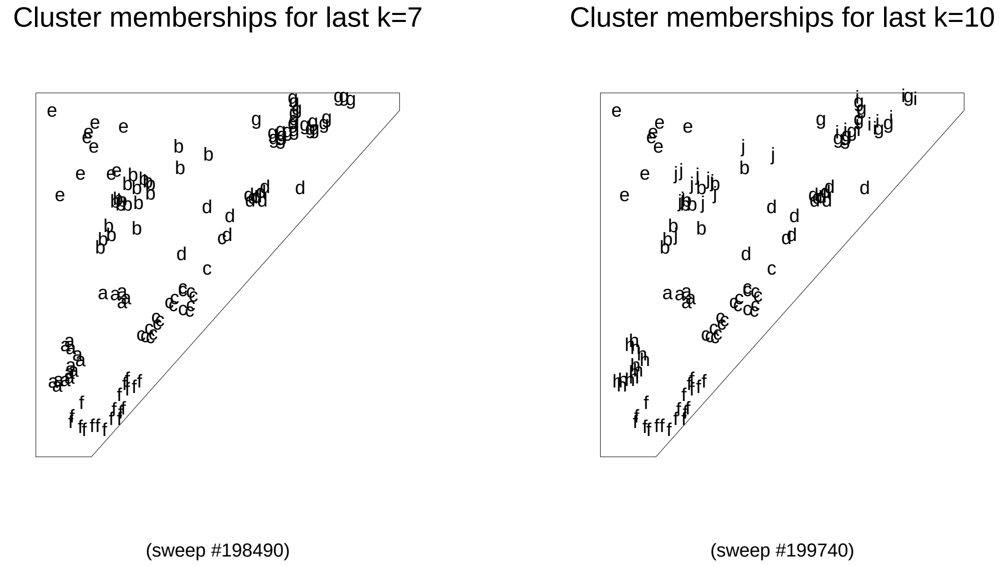
(b) σ_{11}



(c) σ_{22}

(d) σ_{12}

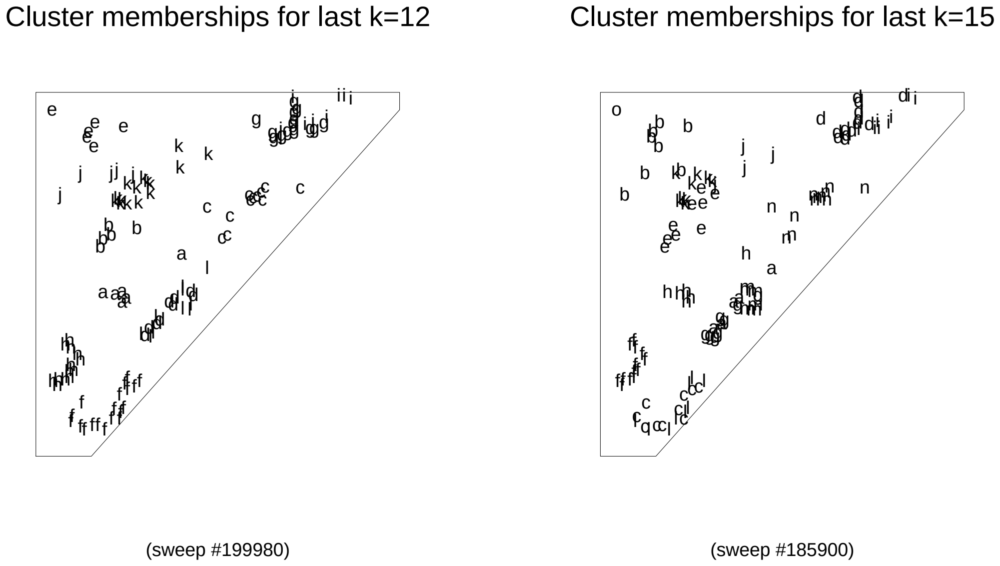
Figure 4: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for hex chain #2 using all 200,000 sweeps.


Cluster memberships for last k=7

Cluster memberships for last k=10

(a) $k = 7$

(b) $k = 10$

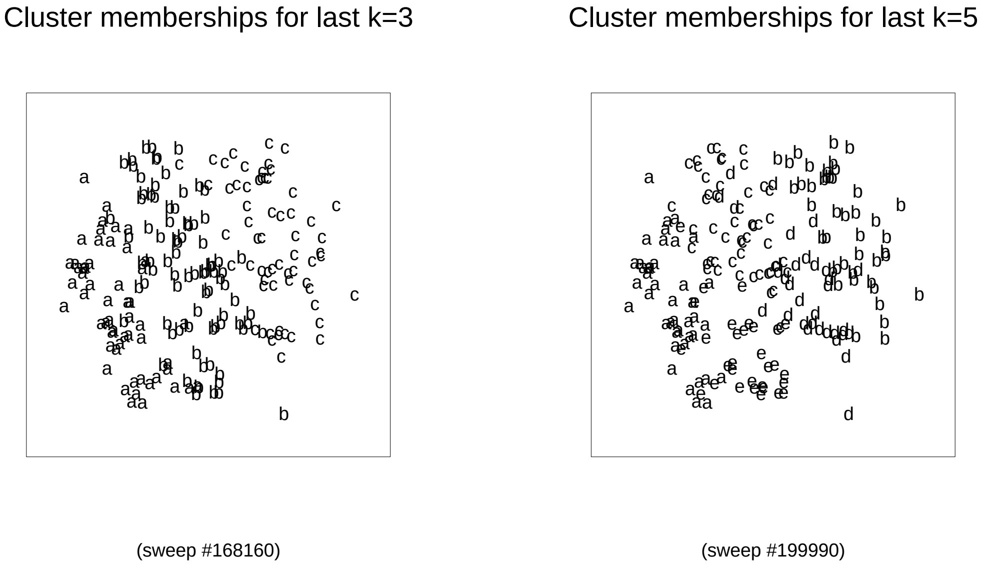

Cluster memberships for last k=12

Cluster memberships for last k=15

(c) $k = 12$

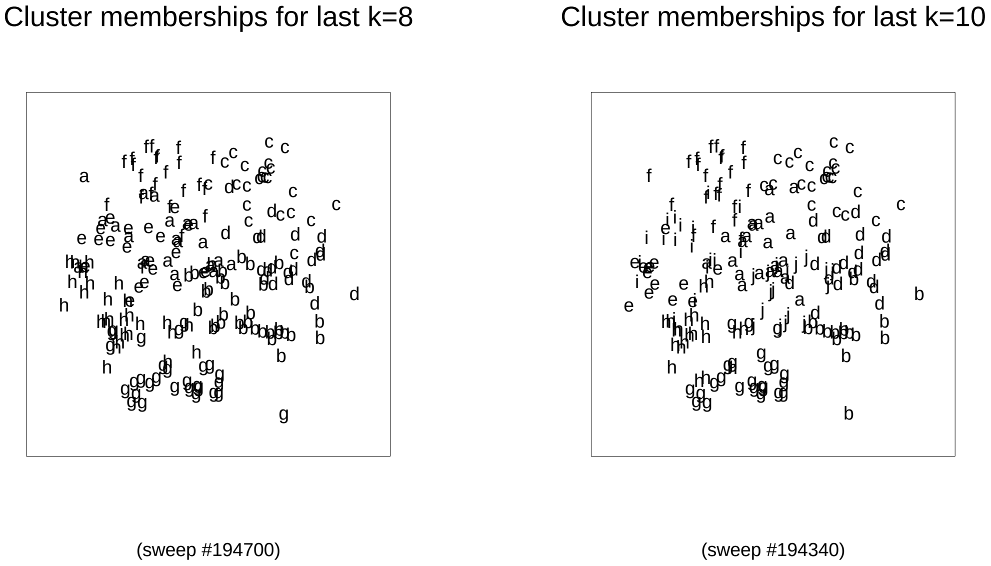
(d) $k = 15$

Figure 5: Sample cluster memberships, redwood data.



(a) $k = 3$

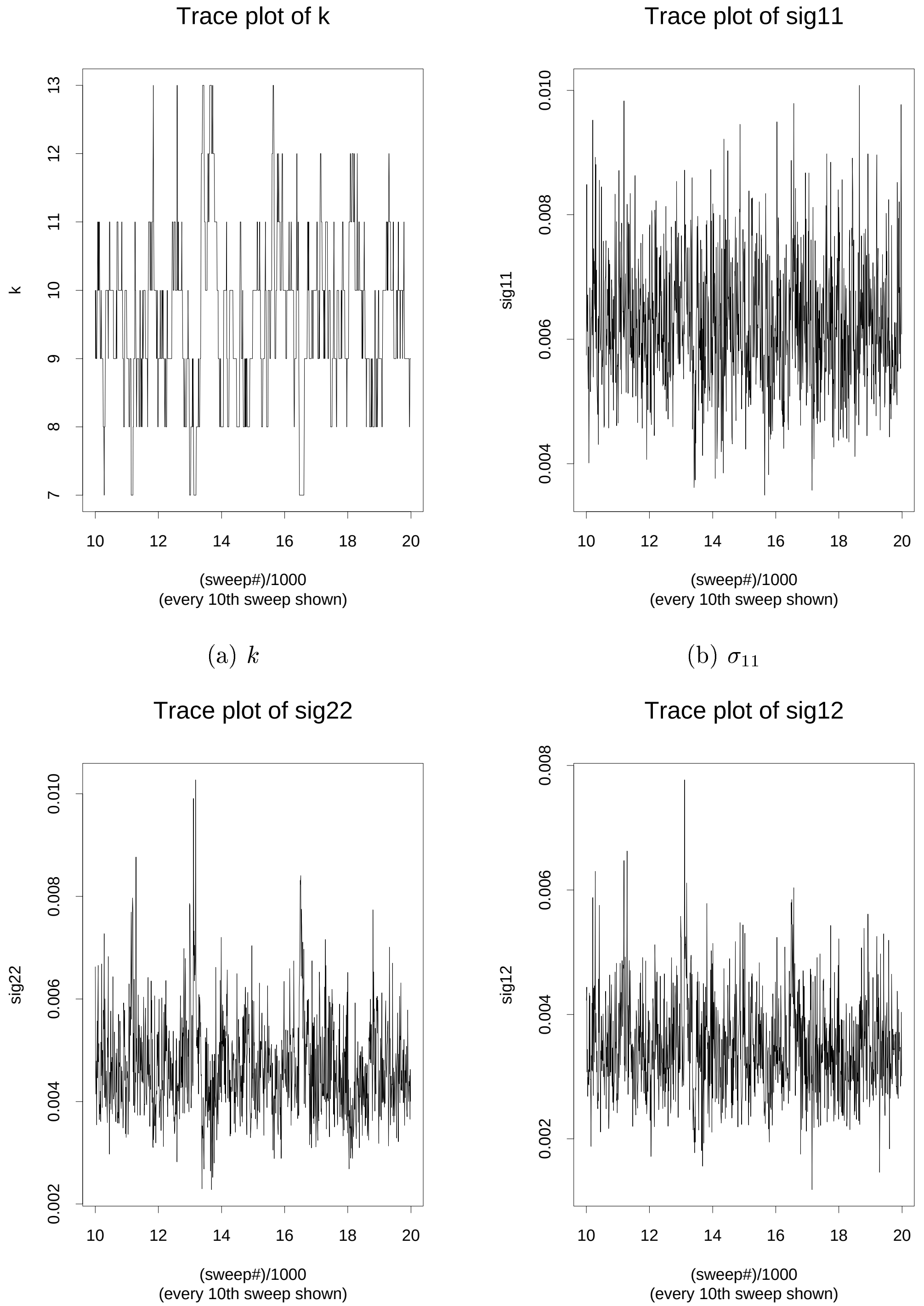
(b) $k = 5$



(c) $k = 8$

(d) $k = 10$

Figure 6: Sample cluster memberships, hex pattern.



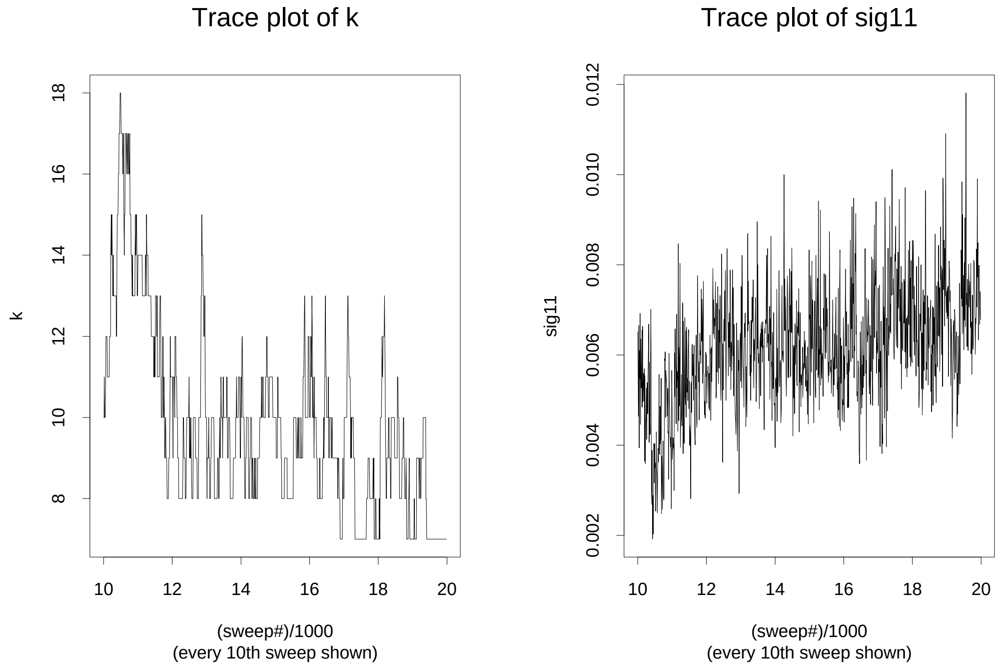
(a) k

(b) σ_{11}

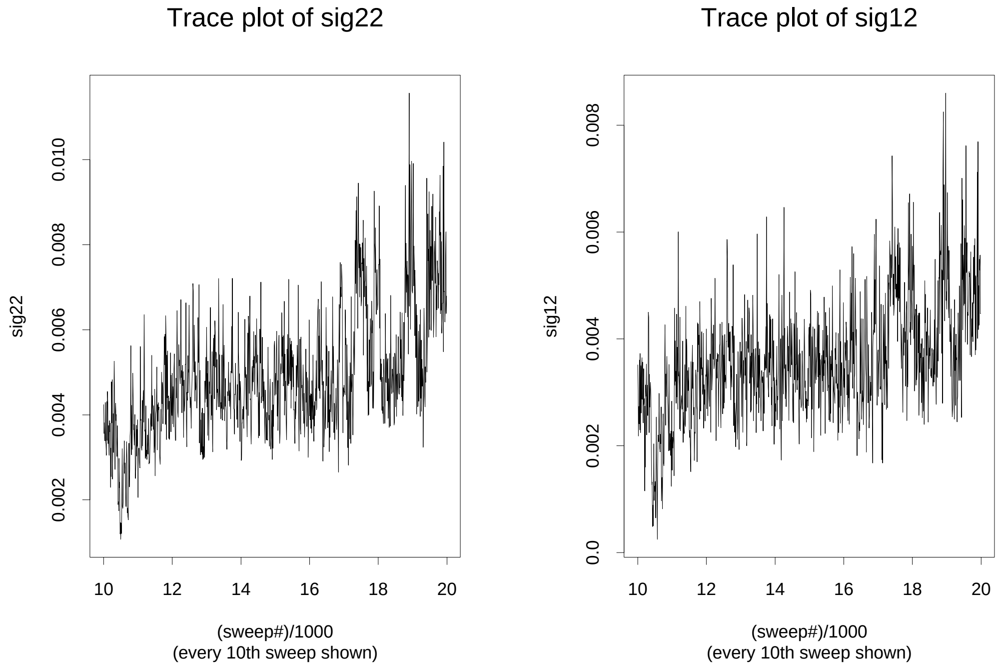
(c) σ_{22}

(d) σ_{12}

Figure 7: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for redwood chain #1 using sweeps 1,000-2,000.

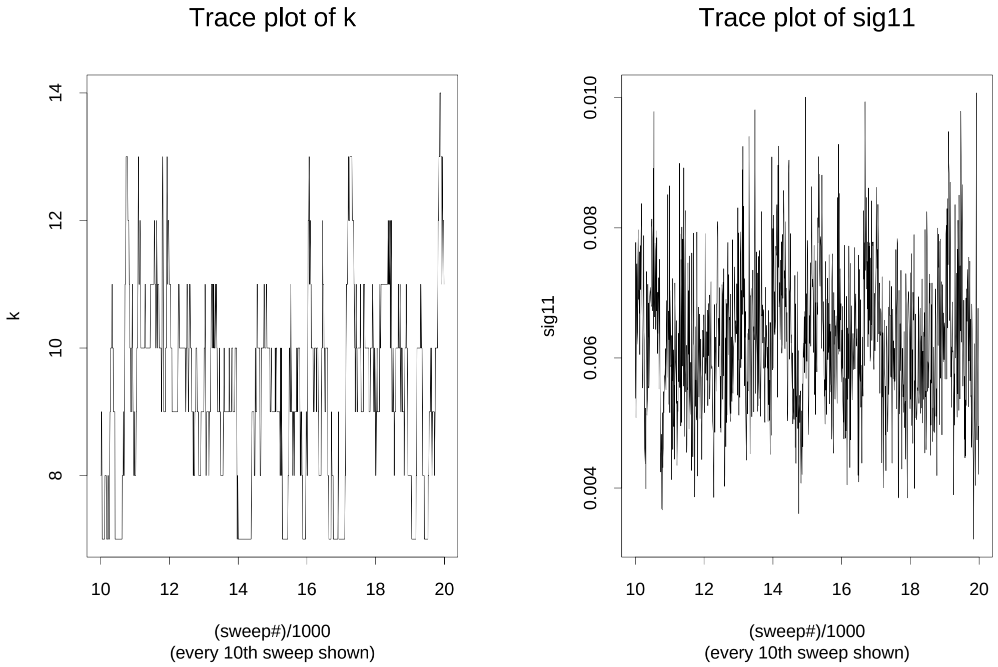


(a) k (b) σ_{11}



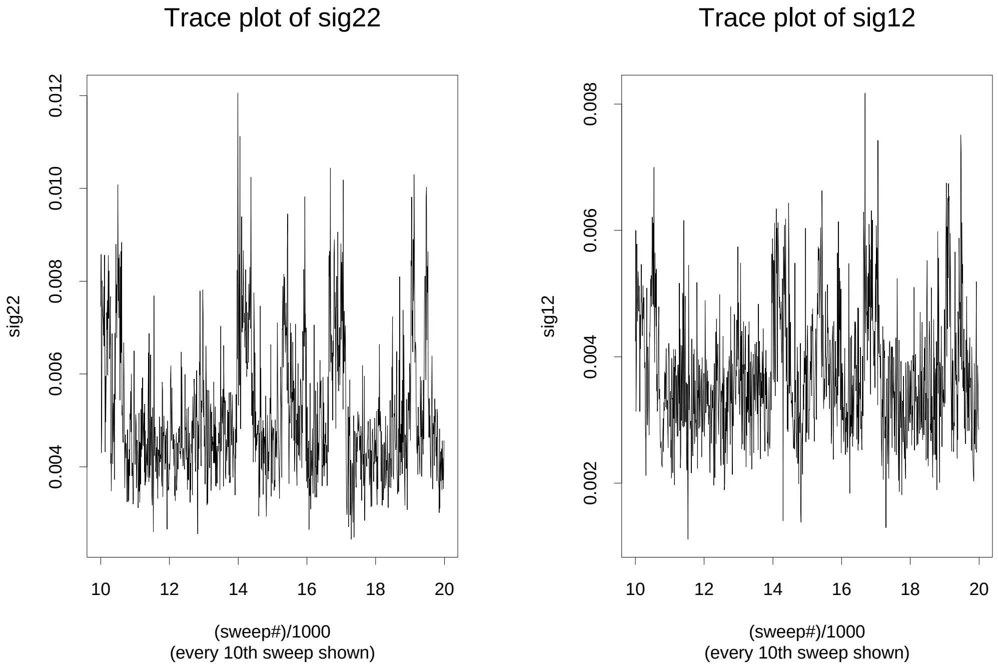
(c) σ_{22} (d) σ_{12}

Figure 8: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for redwood chain #2 using sweeps 1,000-2,000.



(a) k

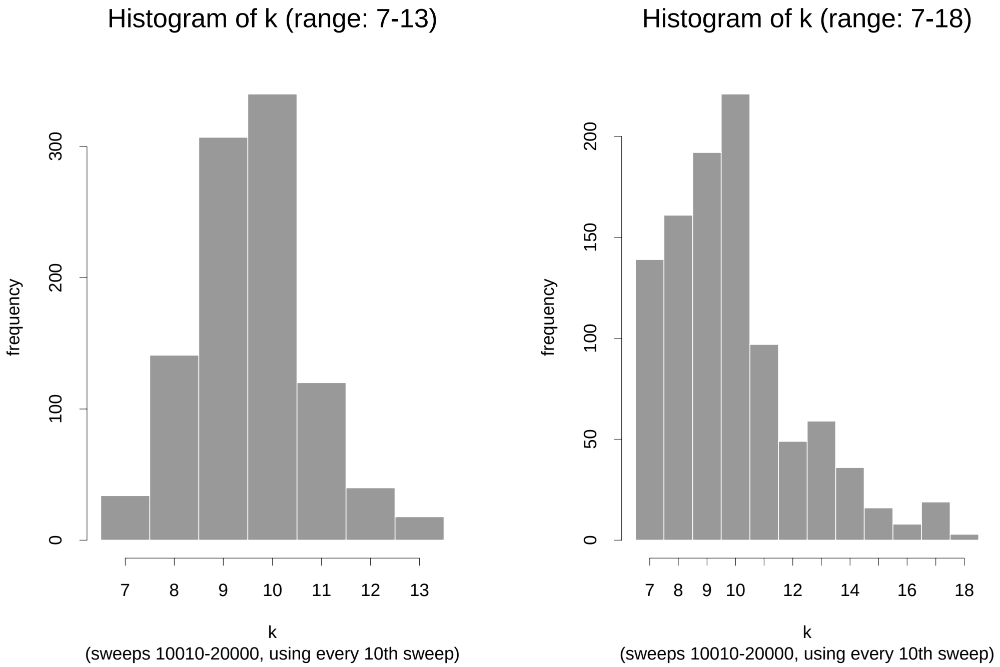
(b) σ_{11}



(c) σ_{22}

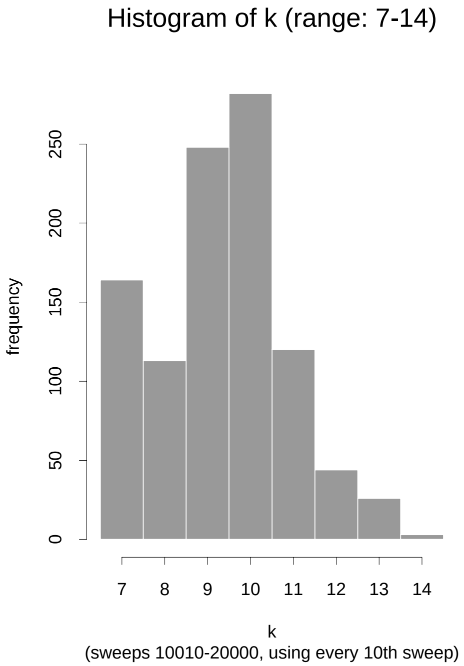
(d) σ_{12}

Figure 9: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for redwood chain #3 using sweeps 1,000-2,000.



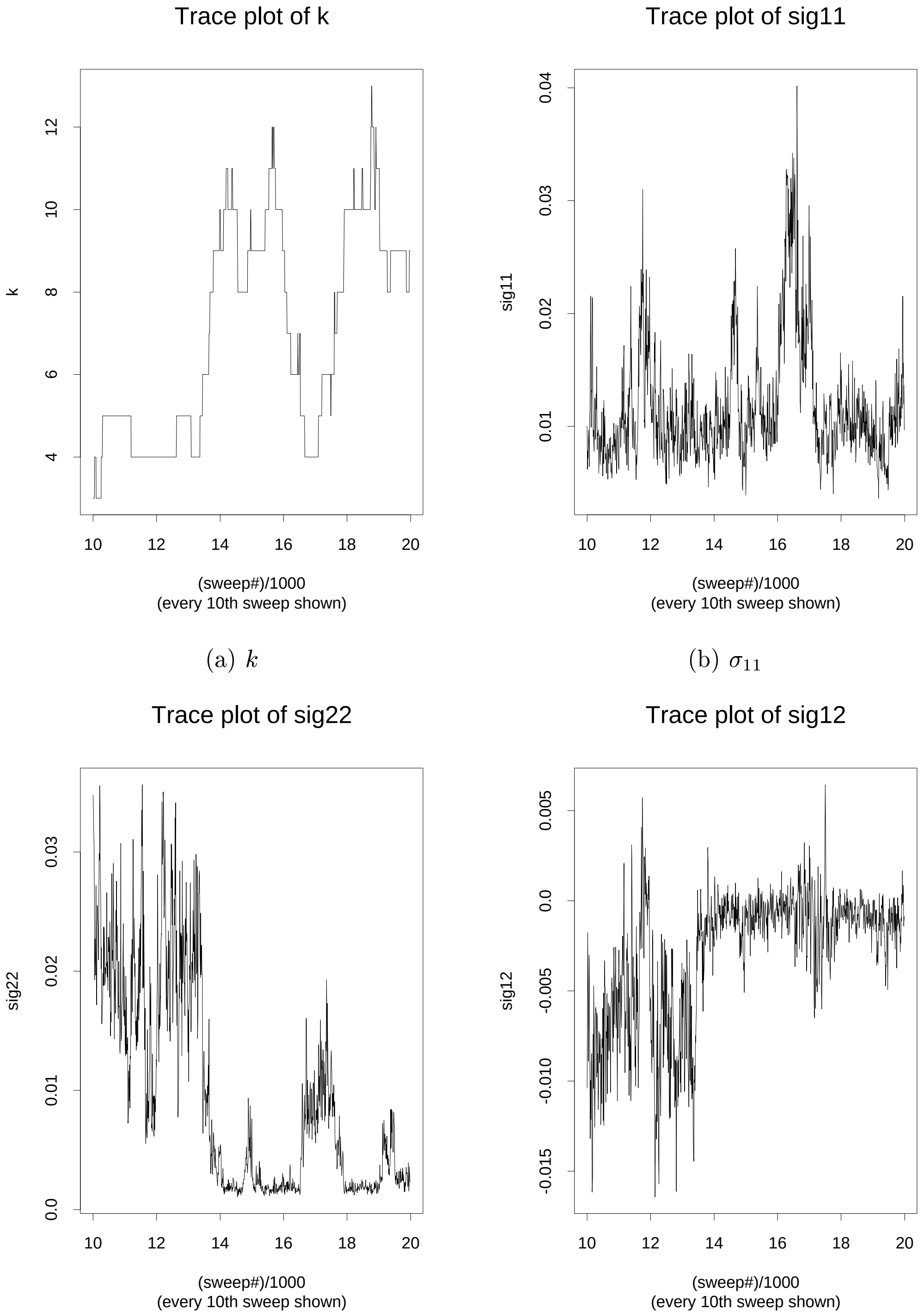
(a) chain #1

(b) chain #2



(c) chain #3

Figure 10: Histogram of k for each redwood chain using sweeps 1,000-2,000.



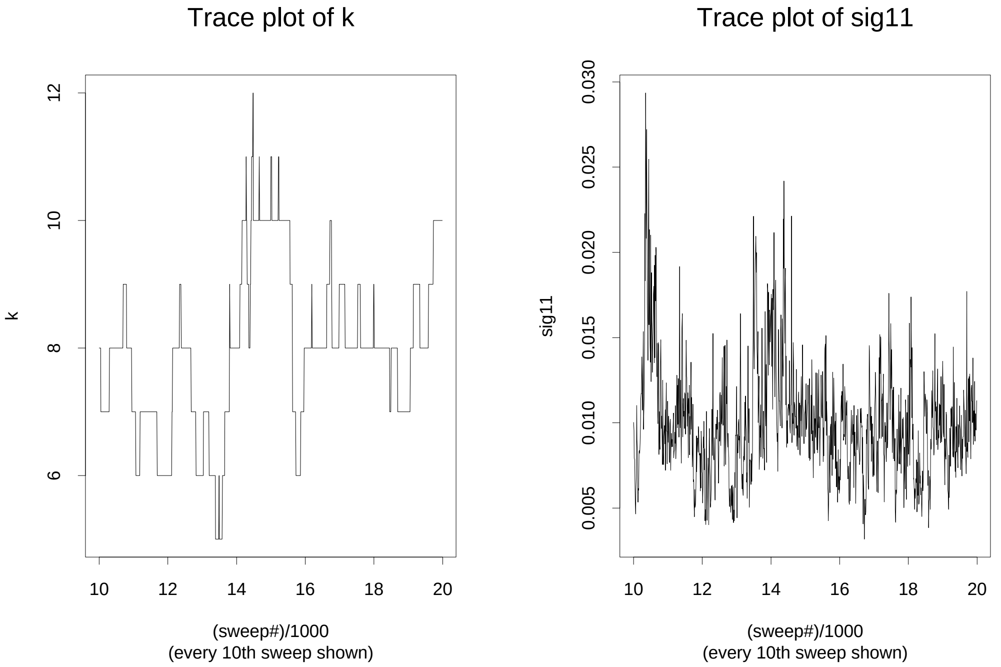
(a) k

(b) σ_{11}

(c) σ_{22}

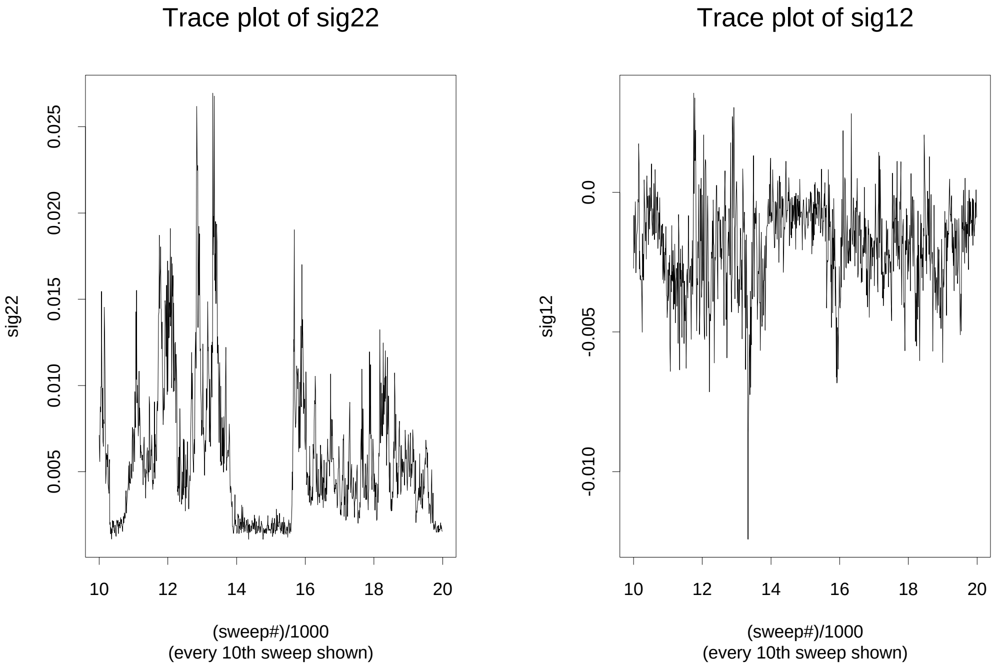
(d) σ_{12}

Figure 11: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for hex chain #1 using sweeps 1,000-2,000.



(a) k

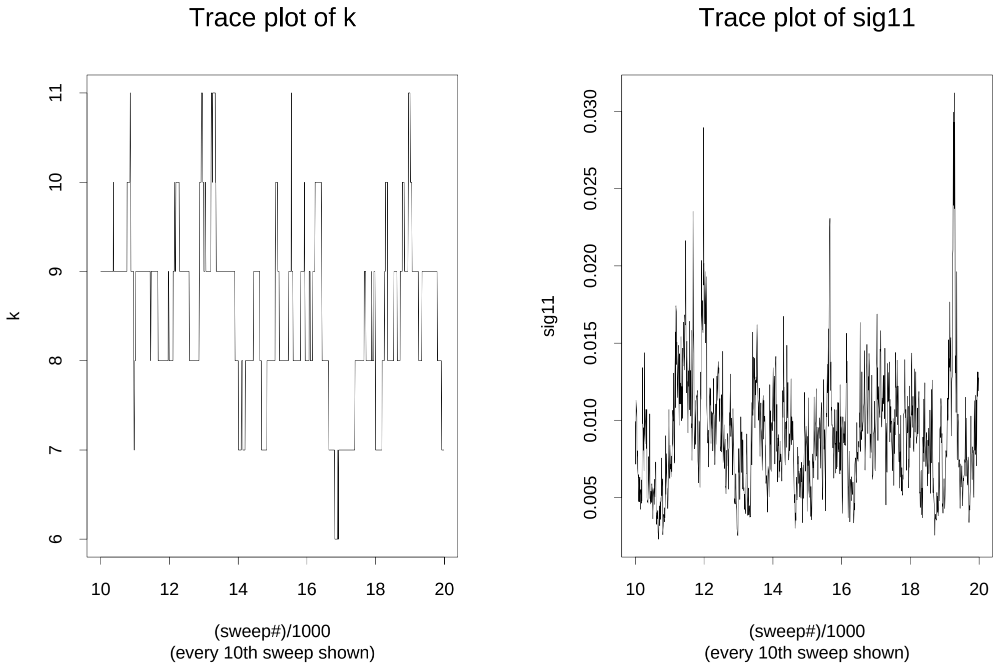
(b) σ_{11}



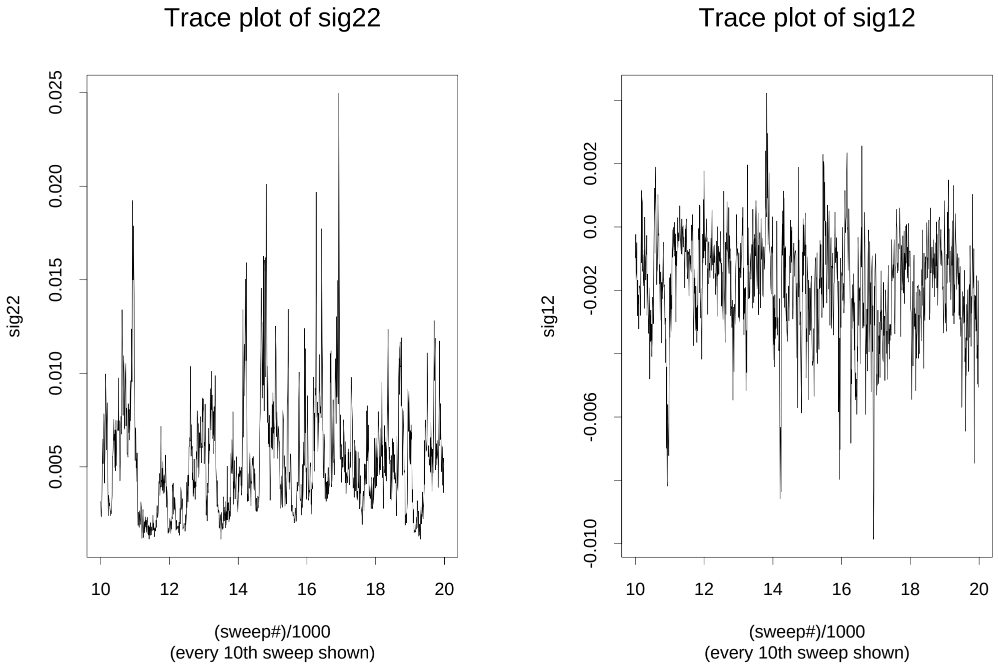
(c) σ_{22}

(d) σ_{12}

Figure 12: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for hex chain #2 using sweeps 1,000-2,000.

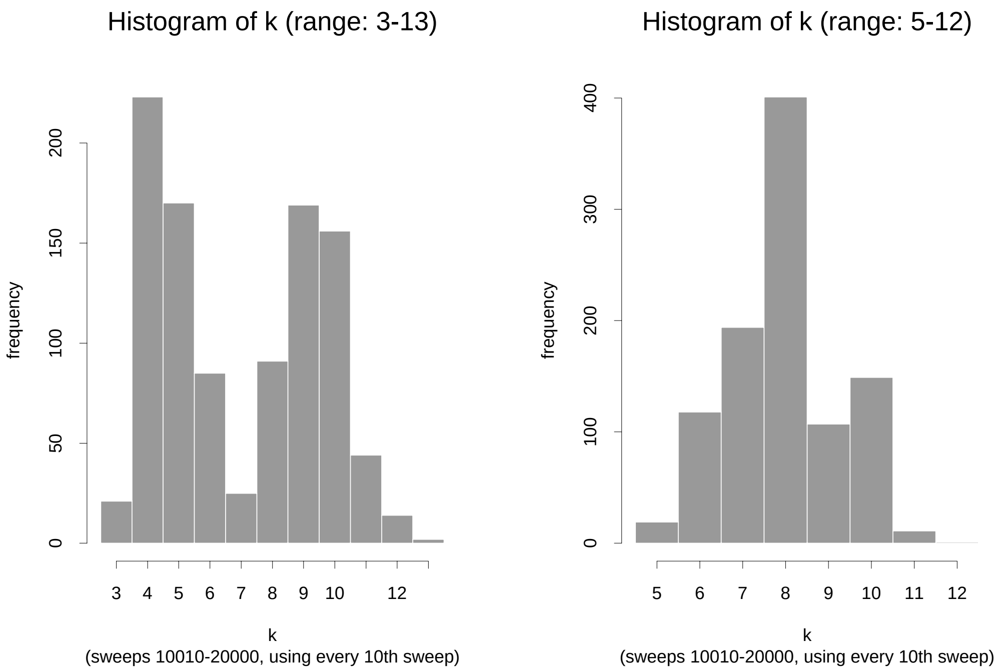


(a) k (b) σ_{11}

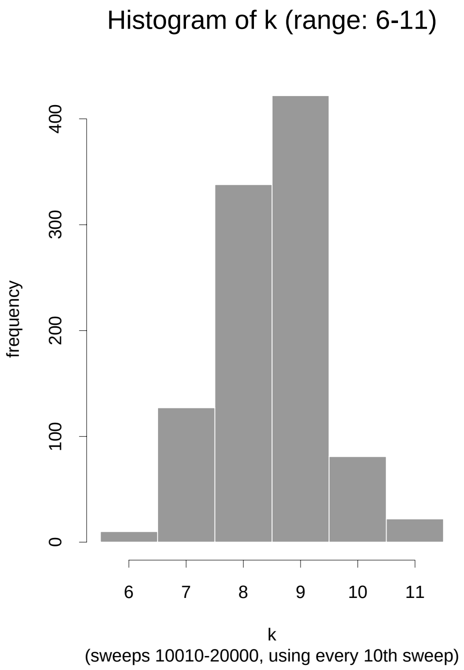


(c) σ_{22} (d) σ_{12}

Figure 13: Trace plots of k, σ_{11}, σ_{22} and σ_{12} for hex chain #3 using sweeps 1,000-2,000.

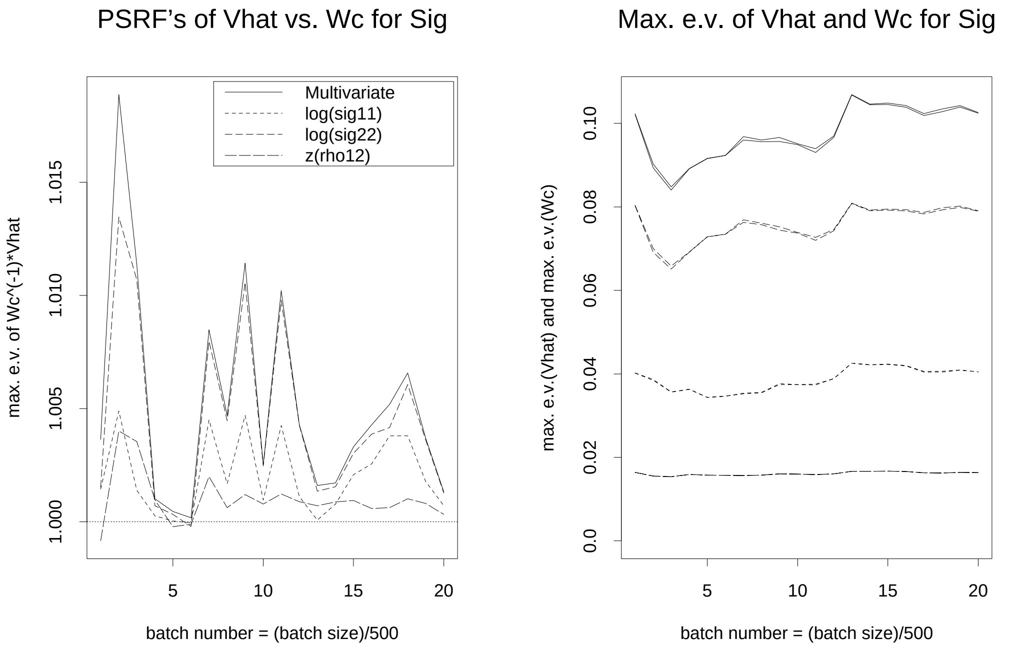


(a) chain #1 (b) chain #2

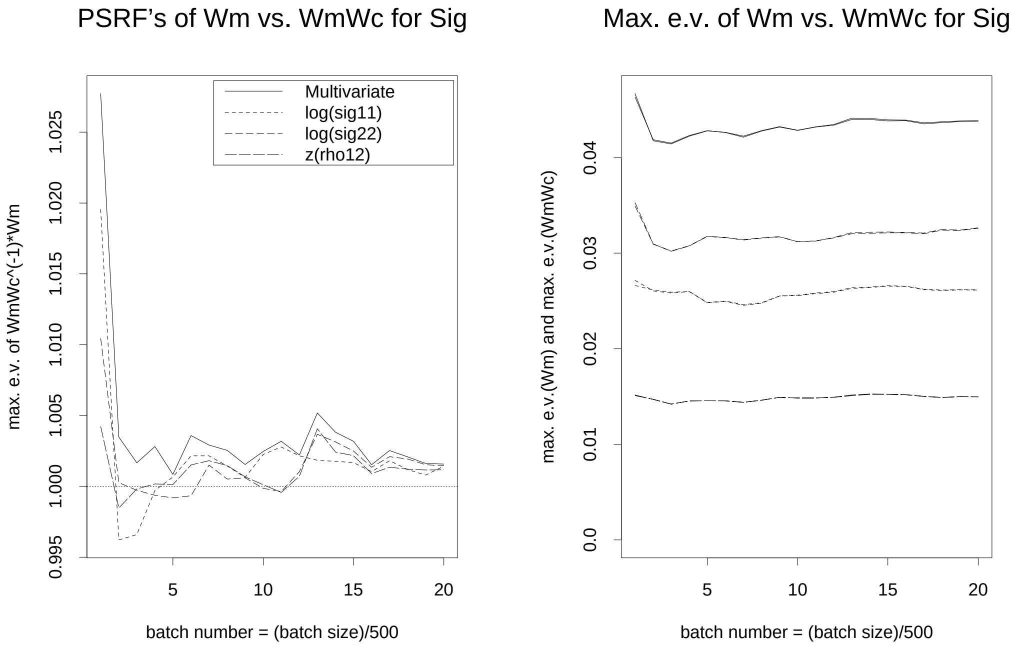


(c) chain #3

Figure 14: Histogram of k for each hex chain using sweeps 1,000-2,000.

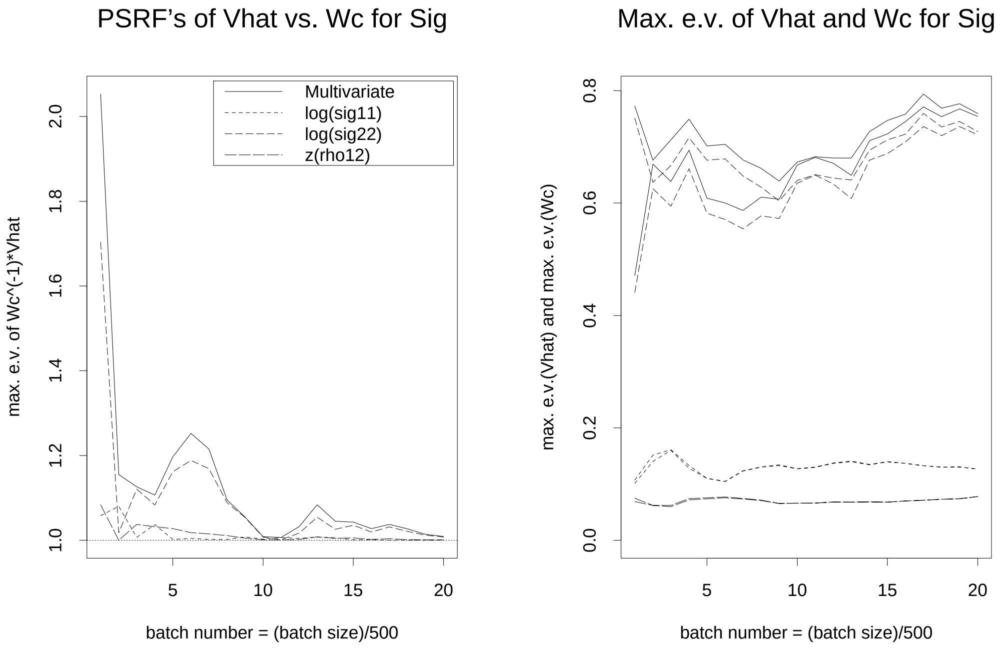


(a) PSRF: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_{c} (b) Max. e.v.: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_{c}

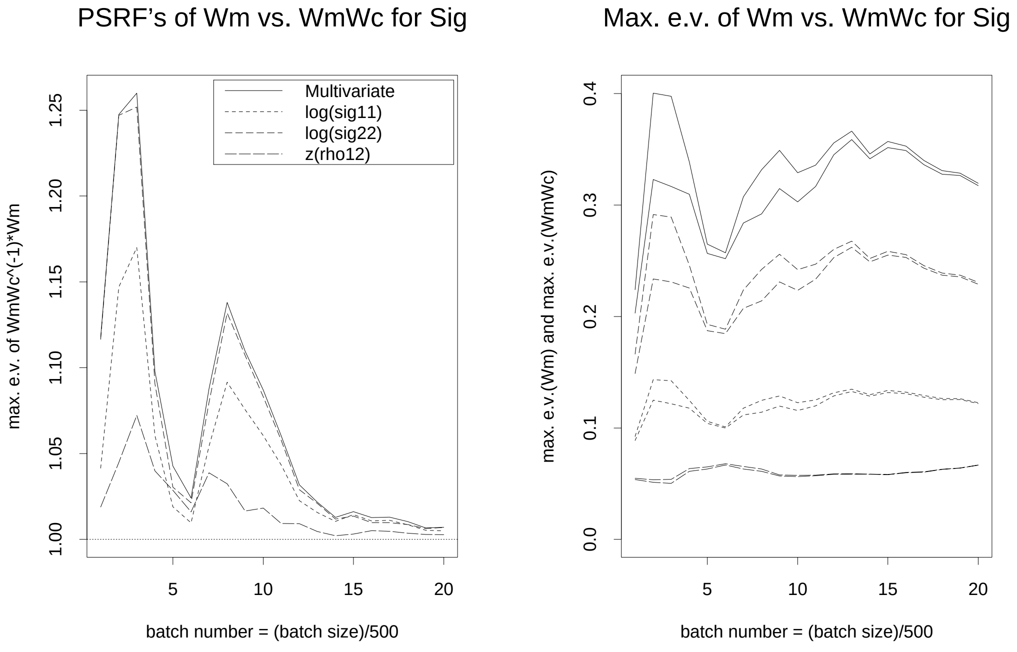


(c) PSRF: $\mathbf{\Sigma}$, W_{m} vs. $W_{\mathrm{m}}W_{\mathrm{c}}$ (d) Max. e.v.: $\mathbf{\Sigma}$, W_{m} vs. $W_{\mathrm{m}}W_{\mathrm{c}}$

Figure 15: Potential scale reduction factor and maximum eigenvalue plots for $(\log\sigma_{11}, \log\sigma_{22}, z(\rho_{12}))$, redwood data.

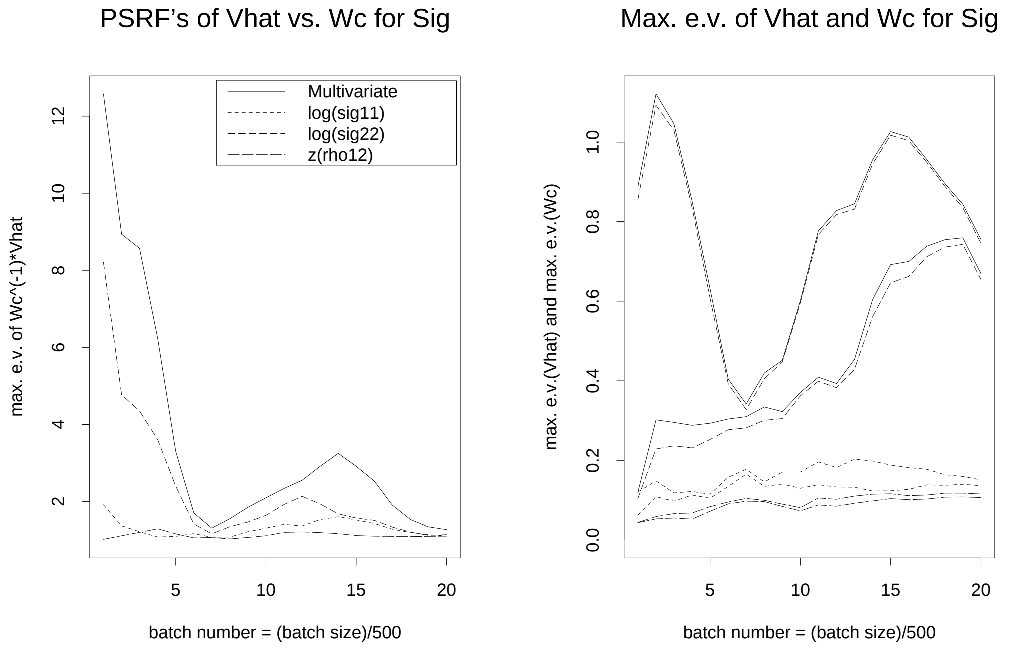


(a) PSRF: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_c (b) Max. e.v.: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_c

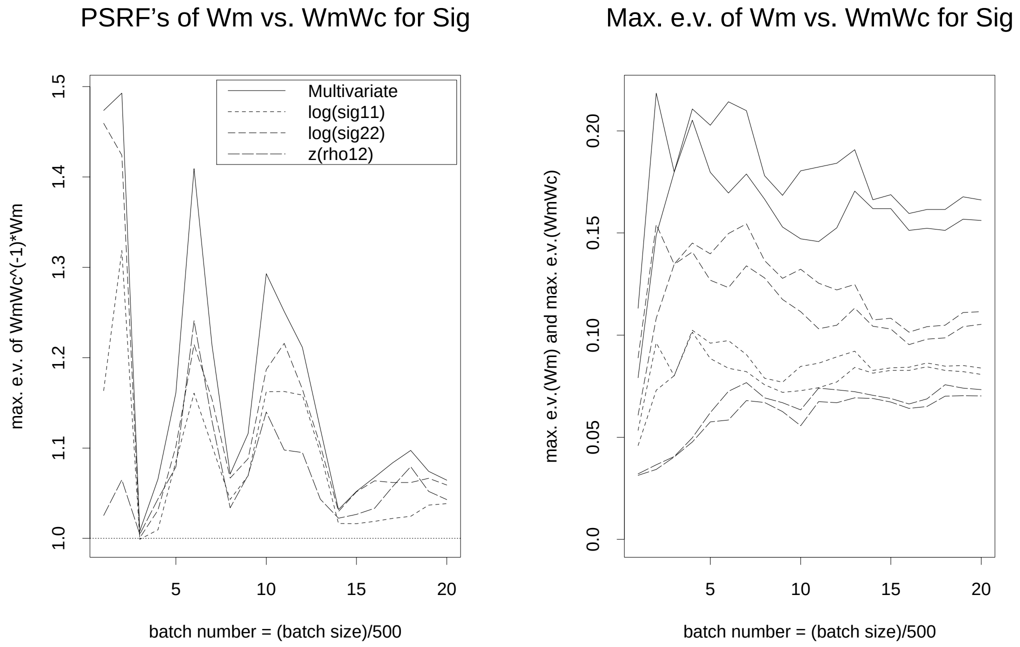


(c) PSRF: $\mathbf{\Sigma}$, W_m vs. $W_\mathrm{m}W_\mathrm{c}$ (d) Max. e.v.: $\mathbf{\Sigma}$, W_m vs. $W_\mathrm{m}W_\mathrm{c}$

Figure 16: Potential scale reduction factor and maximum eigenvalue plots for $(\log\sigma_{11}, \log\sigma_{22}, z(\rho_{12}))$, hex pattern.



(a) PSRF: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_{c} (b) Max. e.v.: $\mathbf{\Sigma}$, $\widehat{V}$ vs. W_{c}



(c) PSRF: $\mathbf{\Sigma}$, W_{m} vs. $W_{\mathrm{m}}W_{\mathrm{c}}$ (d) Max. e.v.: $\mathbf{\Sigma}$, W_{m} vs. $W_{\mathrm{m}}W_{\mathrm{c}}$

Figure 17: Potential scale reduction factor and maximum eigenvalue plots for $(\log\sigma_{11}, \log\sigma_{22}, z(\rho_{12}))$, additional chains with hex pattern.