BUGS 0.6 *
Bayesian inference Using Gibbs Sampling (Addendum to Manual)

David Spiegelhalter **Andrew Thomas** **Nicky Best**
Wally Gilks

MRC Biostatistics Unit, Institute of Public Health,
Robinson Way, Cambridge CB2 2SR

Tel: 44-1223-330300 *Fax: 44-1223-330388*
e-mail: bugs@mrc-bsu.cam.ac.uk *ftp: ftp.mrc-bsu.cam.ac.uk*
http://www.mrc-bsu.cam.ac.uk/bugs

September 26, 1997

This Addendum specifies additional features of BUGS 0.6, and should be read in conjunction with the current manual for BUGS 0.5 (Spiegelhalter *et al.*, 1996*a*) .

Contents

 The support of the Economic and Social Research Council (UK) is gratefully acknowledged. The work was funded in part by ESRC (UK) Award Number H519 25 5023.

1 Getting started

1.1 Getting the software

`BUGS` 0.6 may be obtained from World Wide Web page `http://www.mrc-bsu.cam.ac.uk/bugs`, or by anonymous ftp from `ftp.mrc-bsu.cam.ac.uk` in directory `pub/methodology/bugs`: login as `anonymous` and give your full e-mail address as the password. The message and `README` files will tell you how to obtain the program, examples and documentation.

1.2 The script file for 'bugs' (Sparc)

```
#!/bin/sh
# runs BUGS interactively
#
case $# in
     0) bugs06.sparc 32 bugs;;
     *) echo 'bugs6' ;;
esac
rm bugs.buf
```

"`32`" refers to the number of bins in the metropolis algorithm (Section 2.2).

"`bugs`" refers to the header for filenames.

An appropriate path name should be added before "`bugs06.sparc`".

1.3 The script file for 'backbugs' (Sparc)

```
#!/bin/sh
# runs BUGS taking commands from command file
#
case $# in
     0) echo 'usage backbugs command_file'
        exit 1 ;;
     1) bugs06.sparc 32 bugs $1;;
     *) echo 'usage backbugs command_file' ;;
esac
rm bugs.buf
```

An appropriate path name should be added before "`bugs06.sparc`".

So to submit a series of commands from a file `job.cmd`, use the command `backbugs job.cmd`.

2 New Facilities in 0.6

2.1 Checkpoint command

A command now exists to save people wasting a whole run if a crash occurs. If you type, say, `checkpoint(1000)`, then after every 1000 iterations `BUGS` will

- write to file `bugs.res` the current parameter values
- if anything is being monitored, write the simulated values sequentially to `bugs.sif` (an 'ind' file) and `bugs.sof` (an 'out' file). This format is readable by `CODA`.

Note that `bugs.res` can be used as an initial value file for restarting a run.

2.2 Metropolis sampling

A general Metropolis-within-Gibbs routine can now be used for non-log-concave sampling. This routine uses a simple histogram- based proposal distribution (Ritter and Tanner, 1992), and therefore any parameter that requires Metropolis sampling *must* have bounded range, which is best set up using the `I( , )` construct on the prior distribution, although this is not necessary if a uniform or beta prior distribution are assumed. This is best adapted to as narrow a range as possible to bound the posterior distribution. The default number of bins in the histogram is 32, but this can be changed within the `BUGS` script (see Section 1.2).

If a bounded range is *not* given, an error message will say that `BUGS` is `Unable to choose update method for node`.

Improved Metropolis routines are being written for future versions, and will be available in `BUGS` for Windows.

2.3 Minor changes

1. Non-integer r and n can be used in binomial sampling.
2. Nodes are updated in the reverse order to their specification.

3 Corrected bugs from Version 0.5

1. The adaptive rejection sampling routine has been changed in the hope of avoiding the crashes that have occurred. However, this may have made it slower.
2. Compilation has been speeded up.
3. The `stats` command no longer overwrites the contents of `bugs1.out`.
4. The `I(,)` construct works for all univariate distributions (previously it was ignored when using, for example, the logistic)

5. The `sd` function has been fixed.

6. Some of the more annoying error messages have been fixed: for example a missing data or initial value file no longer leads to a `memory fault` crash.

4 Known restrictions still existing in Version 0.6

1. It is still not possible to place any structure on a covariance matrix given an inverse Wishart distribution. For bivariate normal distributions, such as an intercept and slope of a growth curve model, we recommend modelling as two univariate normals with, say, intercept being a covariate of slope.

2. The COSMOS example does not fully work.

5 Examples

5.1 Dugongs with Metropolis

This problem is described in Spiegelhalter *et al.* (1996*b*)[page 4], and comprises a non-linear and non-conjugate model:

$$\begin{aligned} Y_i &\sim \text{Normal}(\mu_i, \tau), \qquad i = 1, ..., 27 \\ \mu_i &= \alpha - \beta\gamma^{X_i} \qquad \alpha, \beta > 1; 0 < \gamma < 1. \end{aligned}$$

This gives a non-log-concave distribution for γ. The problem was previously handled by discretizing γ, and specifying equal prior probabilities for each discrete value. The `BUGS` 0.6 code is shown below.

```
model dugongs;
const
   N = 27;  # number of observations
var
  x[N],Y[N],mu[N],alpha,beta,gamma,tau,sigma,U1,U2,U3;
data x, Y in "dugongs.dat";
inits in "dugongs.in";
{
  for (i in 1:N) {
     mu[i] <- alpha - beta*pow(gamma,x[i]);
     Y[i] ~ dnorm(mu[i],tau)
  }
  alpha ~ dnorm(0.0,1.0E-4);
  beta ~ dnorm(0.0,1.0E-4);
  tau ~ dgamma(1.0E-3,1.0E-3); sigma <- 1.0/sqrt(tau);
  gamma ~ dunif(0.5,1.0);

# Transform alpha, beta and gamma to scale used by Carlin and Gelfand
  U1 <- log(alpha);
  U2 <- log(beta);
```

```
    U3 <- logit(gamma);
  }
```

We note that γ has been given a bounded domain by using a uniform prior distribution. BUGS 0.6 detects that the Metropolis sampler is required and reports during compilation: `Metropolis method choosen for node gamma`.

Analysis

After a 500 iteration burn-in, 1000 iterations took 31 seconds using the default 32 bins for the Metropolis sampler.

	U1 (log α)	**U2 (log β)**	**U3 (logit γ)**	σ
C & G posterior mode	0.975	−0.014	1.902	–
Ratkowsky least squares estimate	0.981	−0.028	1.932	–
BUGS posterior mode (95% interval)				
Discretisation	0.979 (0.933, 1.032)	−0.029 (−0.180, 0.117)	1.896 (1.366, 2.364)	0.098 (0.074, 0.129)
Metropolis	0.978 (0.931, 1.032)	−0.031 (−0.185, 0.114)	1.893 (1.319, 2.398)	0.099 (0.075, 0.132)

(The **C & G posterior** refers to that in Carlin and Gelfand (1991)).

We note that the Metropolis sampler is considerably faster than using discretisation, and the results are virtually indistinguishable.

5.2 Epilepsy with hierarchical centering

This example is described in Spiegelhalter *et al.* (1996*c*)[page 30], in which convergence problems are noted. Gelfand *et al.* (1995) and Gelfand *et al.* (1996) discuss the method of *hierarchical centering* for such models, in which each stochastic variable is, as far as possible, considered as arising from a stochastic mean. In effect, covariates are entered as 'high' in the model as possible. They argue this procedure should often improve convergence, and further evidence is provided by Roberts and Sahu (1997).

For Model III in the epilepsy example, rather than having both random effects entering into a single regression for the Poisson mean, we may separate out the random effects to create an additional level on the model. The model is thus given by:

$$
\begin{aligned}
y_{jk} &\sim \text{Poisson}(m_{jk}) \\
\log m_{jk} &= b_{jk} \\
b_{jk} &\sim \text{Normal}(\mu_{jk}, \tau_b) \\
\log \mu_{jk} &= \alpha_0 + \alpha_{Base} \log(\text{Base}_j/4) + \alpha_{Trt}\text{Trt}_j \\
&\quad \alpha_{BT}\text{Trt}_j \log(\text{Base}_j/4) + \alpha_{Age}\text{Age}_j + \alpha_{V4}\text{V4}_k + b1_j \\
b1_j &\sim \text{Normal}(0, \tau_{b1})
\end{aligned}
$$

Coefficients and precisions are given independent "noninformative" priors. The appropriate graph is shown in Figure 1.

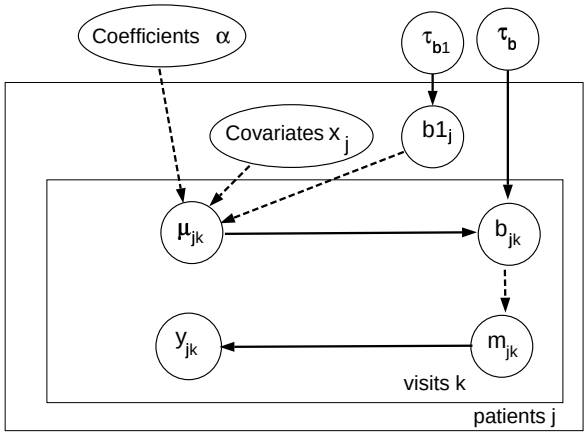

Figure 1: Graphical model for `epil` example, using a hierarchically centered parameterisation

Model specification for `epil` example with hierarchical centering. The part indicated `+++++` is identical to that given in Spiegelhalter *et al.* (1996*c*)[page 30].

```
   for(j in 1:N) {
     for(k in 1:T) {
        mu[j,k] <- a0 + alpha.Base * (log.Base4[j]-log.Base4.bar)
                              + alpha.Trt * (Trt[j]-Trt.bar)
                              + alpha.BT  * (BT[j] - BT.bar)
                              + alpha.Age * (log.Age[j]-log.Age.bar)
                              + alpha.V4  * (V4[k] - V4.bar)
                              + b1[j];
        y[j,k] ~ dpois(m[j,k]);
log(m[j,k]) <- b[j,k];
        b[j,k] ~ dnorm(mu[j,k],tau.b);       # subject*visit random effects
     }
     b1[j]         ~ dnorm(0.0,tau.b1);        # subject random effects
+++++
```

Analysis

A burn-in of 3000 iterations was followed by a further 7000 iterations. This took approximately 30 minutes.

	PQL	BUGS
variable	coeff ± SE	coeff ± SE
Fixed effects		
constant	-1.27 ± 1.2	-1.43 ± 1.25
Base	.86 ± .13	.87 ± .14
Trt	-.93 ± .40	-1.02 ± .42
Base x Trt	.34 ± .21	.38 ± .21
Age	.47 ± .35	.50 ± .36
V4	-.10 ± .09	-.11 ± .09
Subject level random effects		
σ_{b1}	.48 ± .06	.50 ± .07
Unit level random effects		
σ_b	.36 ± .04	.36 ± .04

We have generally found that hierarchical centering leads to both quicker sampling and earlier convergence.

5.3 Orange trees - non-linear hierarchical models

This example is analysed in Lindstrom and Bates (1990) as an example of a mixed non-linear growth curve model. The data describe the growth of each of five orange trees, with measurements at seven common times:

Tree	y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	30	58	87	115	120	142	145
2	33	69	111	156	172	203	203
3	30	51	75	108	115	139	140
4	32	62	112	167	179	209	214
5	30	49	81	125	142	174	177
Time ($\underline{x}$)	18	484	664	1004	1231	1372	1582

A logistic growth curve model, with an unknown maximum, is assumed. We first standardise the covariate x_j to $z_j = (x_j - \overline{x})/\mathrm{sd}(\underline{x})$ in order to improve convergence and stability of estimates, and to make the random effects assumptions more reasonable.

$$\begin{aligned} y_{ij} &\sim \mathrm{Norm}(m_{ij}, \tau_c) \\ m_{ij} &= \frac{e^{\theta_{i1}}}{1 + e^{\theta_{i2} + \theta_{i3} z_j}} \\ \theta_{ik} &\sim \mathrm{Norm}(\mu_k, \tau_k) \end{aligned}$$

Lindstrom and Bates (1990) only take $\phi_{i1} = e^{\theta_{i1}}$ as a random effect with a Gaussian population distribution. We shall allow all three growth parameters to vary between trees; means and precisions are given independent "noninformative" priors.

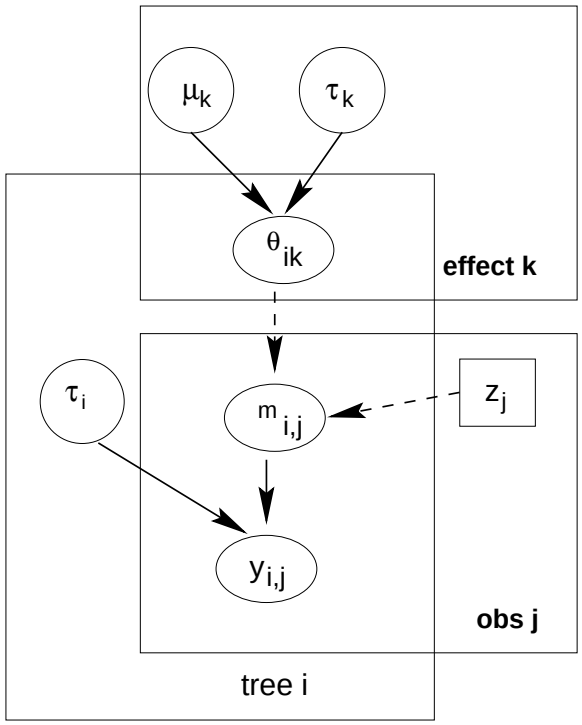

Figure 2: Graph of the orange tree example.

As mentioned in Section 2.2, we need to specify a range for each of the θ parameters to be sampled using the Metropolis algorithm. It is convenient to first run a fixed-effect model in which five independent growth curves are fitted to the data, specifying only that the θ's are all between -20 and +20. This gives rise to estimates for θ that suggest generous lower and upper bounds of (4,6) for θ_{i1}'s, (-2,0) for θ_{i2}'s and (-3,0) for θ_{i3}'s. These are then placed in the data file as the `lower` and `upper` vectors.

Model specification for `orange` example

```
model otree;
const
n = 7,
K = 5;
var
tauC,mu[3],tau[3], Y[K,n],m[K,n],phi[K,3],theta[K,3],
lower[3],upper[3],sigmaC,sigma[3],x[n],x.bar,x.sd;
data in "otree.dat";
inits in "otree.in";
{
x.bar <- mean(x[]);
x.sd <- sd(x[]);
for (i in 1:K) {
   for (j in 1:n) {
      Y[i, j] ~ dnorm(m[i, j], tauC)
```

```
      m[i, j] <- exp(theta[i,1]) /
                 (1 +  exp(theta[i,2] + theta[i, 3] * (x[j]-x.bar)/x.sd));
      }
   for (k in 1:3) {
   theta[i, k] ~ dnorm(mu[k], tau[k])I(lower[k], upper[k])
    }
}
tauC ~ dgamma(1.0E-3, 1.0E-3)
sigmaC <- 1 / sqrt(tauC)
for (k in 1:3) {
   mu[k] ~ dnorm(0, 1.0E-4)
   tau[k] ~ dgamma(1.0E-3, 1.0E-3)
   sigma[k] <- 1 / sqrt(tau[k])
   }
}
```

Analysis

A burn-in of 500 iterations followed by a further 1000 iterations took approximately 2.25 minutes.

variable ($\underline{\theta_k}$)	mean (μ_k) $\pm$ SE	sd (σ_k) $\pm$ SE
$\underline{\theta_1}$	5.23 $\pm$.11	.24 $\pm$ .12
$\underline{\theta_2}$	-.64 $\pm$ .15	.14 $\pm$ .12
$\underline{\theta_3}$	-1.44 $\pm$ .13	.12 $\pm$ .11

From the size and standard deviations of the random effects σ's, the assumption of Lindstrom and Bates (1990) that a random effect is only required for $\underline{\theta_1}$ appears reasonable.

5.4 `ddIddC`: a longitudinal Laird-Ware mixed model

Consider the Gaussian linear mixed model (Laird and Ware, 1982),

$$\begin{aligned} Y_i &= X_i\alpha + W_i\beta_i + \epsilon_i \\ \beta_i &\sim N_q(0, V^{-1}) \end{aligned} \tag{1}$$

where the Y_i are vectors of length n_i containing the observations on the i^{th} unit, and the ϵ_i are error vectors of the same length independently distributed as $N_{n_i}(0, \sigma^{-2}I_{n_i})$, $i = 1, \ldots, k$: note that all our Normal parameterisations are in terms of precisions. In this mixed model, X_i is an $n_i \times p$ design matrix of covariates and α is a corresponding $p \times 1$ vector of fixed effects. In contrast, W_i is a $n_i \times q$ design matrix (q typically less than p), and β_i is a $q \times 1$ vector of subject-specific random effects. The β_i model the subject-specific means, as well as enabling the model to capture marginal dependence among the observations on the i^{th} unit. The hierarchical specification of this model is completed by adding the prior distributions $\Omega \equiv V^{-1} \sim \text{Wishart}(R, \rho)$, $\tau \equiv \sigma^{-2} \sim \text{Gamma}(a, b)$, and $\alpha \sim N_p(c, D)$.

We apply this model to continuous longitudinal data from a clinical trial originally described by Abrams *et al.* (1994), which compared the effectiveness of two antiretroviral drugs (didanosine, *ddI*, and zalcitabine, *ddC*) in 467 persons with advanced HIV infection. The response variable Y_{ij} for

patient i at time j is the square root of the patient's CD4 count, a seriological measure of immune system health and prognostic factor for AIDS-related illness and mortality. The dataset records patient CD4 counts at study entry and again at 2, 6, 12, and 18 months after entry, though a great many of these observations are missing for many patients (the sample sizes at the five time points for the two drug groups are (230, 182, 153, 102, 22) and (236, 186, 157, 123, 14), respectively).

Following a Bayesian reanalysis of these data (Carlin, 1996; Carlin and Louis, 1996), we seek to fit model (1) where the j^{th} row of the patient i's design matrix W_i takes the form

$$w_{ij} = (1\ ,\ t_{ij}\ ,\ (t_{ij} - 2)^+)\ ,$$

where $t_{ij} \in \{0, 2, 6, 12, 18\}$ and $z^+ = \max(z, 0)$. Thus the three columns of W_i correspond to individual-level intercept, slope, and possible change in slope after the two month visit (by which time the drugs are expected to produce a detectable benefit). We further account for the effect of two covariates by including them in the fixed effect design matrix X_i. These covariates are d_i, a binary variable indicating whether patient i received ddI ($d_i = 1$) or ddC ($d_i = 0$), and a_i, a binary variable telling whether the patient was diagnosed as having AIDS at baseline ($a_i = 1$) or not ($a_i = 0$). Each of these covariates is allowed to influence the intercept, slope and change, and hence

$$X_i = (W_i \mid d_i W_i \mid a_i W_i)\ ,$$

so that $p = 3q = 9$.

We complete our model specification with minimally informative priors, taking care to ensure that they do not lead to improper posterior distributions for the variance components σ^2 and D. Following previous work, we set $\rho = 24$ and $R = 24 \times Diag(2^2, (.25)^2, (.25)^2)$, which should preserve identifiability while still allowing the random effects a reasonable amount of freedom. For the prior on τ^2 we take $a = 1, b = 100$ (a prior with both mean and standard deviation equal to $(1/10)^2$), while for α we set

$$\begin{aligned} c &= (10, 0, 0, 0, 0, 0, -3, 0, 0)\ , \quad \text{and} \\ D &= Diag(2^2, 1^2, 1^2, (.1)^2, 1^2, 1^2, 1^2, 1^2, 1^2)\ , \end{aligned}$$

a prior biased strongly away from 0 only for the baseline intercept, α_1, and the intercept adjustment for a positive AIDS diagnosis, α_7. This prior also forces the drug group intercept (i.e., the effect at baseline) α_4 to be very small, since patients were assigned to drug group at random.

Here is the `BUGS` code to fit this model, where `ind` indexes the individual in the study, and `i` and `j` index the rows and columns of the design matrices, respectively. By placing `NA`'s in the data file, the W matrix is common to all individuals, but the X matrix is still individual-specific.

```
model ddIddC;
const
   N = 467,  # number of patients
   s = 5,    # number of time points
   q = 3,    # number of random effects
   p = 9;    # number of fixed effects
var
   X[N,s,p],W[s,q],Y[N,s],alpha[p],beta[N,q],d[N],a[N],Omega[q,q],V[q,q],
   Sigma2[q,q],sigma,tau,R[q,q],rho,c[p],Omega.alpha[p,p],
   mu[N,s],mu.beta[q];
```

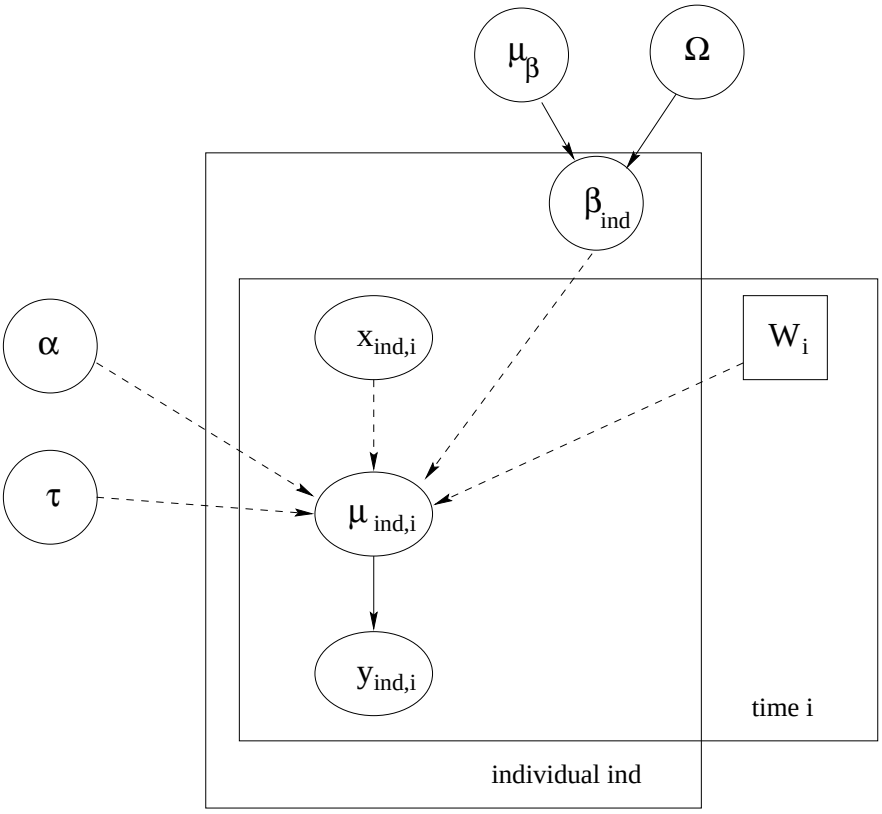

Figure 3: Graph of the ddIddc example.

```
data d,a in "drugaids.dat", Y in "Y.dat",
  W in "W.dat", c in "priormean.dat", Omega.alpha in "priorprec.dat";
inits in "ddIddC.in";
{
    for (ind in 1:N) {
      for (i in 1:s) {

        for (j in 1:q) {
          X[ind,i,j] <- W[i,j];
          X[ind,i,j+3] <- d[ind]*W[i,j];
          X[ind,i,j+6] <- a[ind]*W[i,j];
}
        Y[ind,i] ~ dnorm(mu[ind,i],tau);
        mu[ind,i] <- inprod(X[ind,i,],alpha[]) + inprod(W[i,],beta[ind,]);
}
     beta[ind,] ~ dmnorm(mu.beta[],Omega[,]);  # trivariate Normal
}
  tau  ~ dgamma(1, 100);  sigma <- 1.0/sqrt(tau);
  Omega[,]  ~ dwish(R[,],24);    # Wishart prior on precision matrix
  R[1,1] <- 96.0;  R[1,2] <- 0.0;  R[1,3] <- 0.0;
  R[2,1] <- 0.0;   R[2,2] <- 1.5;  R[2,3] <- 0.0;
  R[3,1] <- 0.0;   R[3,2] <- 0.0;  R[3,3] <- 1.5;
  V[,] <- inverse(Omega[,])

  mu.beta[1] <- 0.0; mu.beta[2] <- 0.0; mu.beta[3] <- 0.0;
#  for (j in 1:p){alpha[j] ~ dnorm(c[j],Omega.alpha[j,j]);}  #  univ normals
  alpha[] ~ dmnorm(c[],Omega.alpha[,]);  #  mv normal -- better convergence!
}
```

Running this BUGS code for 5000 iterations produces the summaries in Table 1. The results are quite comparable to those given by Carlin and Louis (1996)[pp.280-281]. Interestingly, this original work took several hundred lines of code in Fortran 77, augmented with IMSL subroutine calls for matrix manipulation and random variate generation – a stark contrast with the fewer than 40 lines of BUGS code above.

The single line that is commented out in the above BUGS code can be used to specify the (independence) prior for α componentwise using dnorm, instead of all at once using dmnorm. While mathematically equivalent, Table 2 shows that the univariate specification to be inferior in terms of convergence speed, since BUGS then updates the α_i one at a time, instead of as a vector. Laboring against the cross-correlations within this vector, overall performance deteriorates.

We remark that the hierarchically centered version of this model recommended for this dataset by Gelfand *et al.* (1995), namely

$$\eta_i = \alpha^{(0)} + d_i\alpha^{(d)} + a_i\alpha^{(a)} + \beta_i \ ,$$

where $\alpha^{(0)} = (\alpha_1, \alpha_2, \alpha_3)$, $\alpha^{(d)} = (\alpha_4, \alpha_5, \alpha_6)$, and $\alpha^{(a)} = (\alpha_7, \alpha_8, \alpha_9)$, is not possible within the current version of BUGS. This is because BUGS cannot calculate the proper multivariate normal mean and precision matrix when the "data" (in the centered version, the η_i) are not univariate, unless

parameter	mean	95% interval lower	upper
α_1	9.938E+0	9.338E+0	1.053E+1
α_2	-3.817E-2	-2.532E-1	1.782E-1
α_3	-1.437E-1	-3.890E-1	1.025E-1
α_4	9.326E-3	-1.767E-1	1.971E-1
α_5	3.280E-1	9.792E-2	5.530E-1
α_6	-3.596E-1	-6.270E-1	-9.829E-2
α_7	-4.279E+0	-4.996E+0	-3.506E+0
α_8	-3.297E-1	-5.607E-1	-9.252E-2
α_9	3.835E-1	1.139E-1	6.474E-1
$\beta_{8,1}$	-7.618E+0	-9.738E+0	-5.611E+0
$\beta_{8,2}$	-3.349E-1	-8.263E-1	1.270E-1
$\beta_{8,3}$	3.071E-1	-2.040E-1	8.550E-1
σ	1.681E+0	1.587E+0	1.774E+0
$\Omega_{1,1}$	1.048E-1	6.838E-2	1.711E-1
$\Omega_{1,2}$	-1.591E-1	-6.515E-1	2.222E-1

Table 1: Posterior summaries, ddI/ddC data model

the data mean is identical to the multivariate normal prior. `BUGS` can however accommodate some simpler, univariate centering forms, as in the revised `epil` example.

5.5 `PK`: a nonlinear population pharmacokinetic model

Wakefield et al. (1994) consider the data in Table 3, which record the plasma concentration Y_{ij} of the drug Cadralazine at various time lags x_{ij} following the administration of a single dose of 30 mg in 10 cardiac failure patients. Here, $i = 1, \ldots, 10$ indexes the patient, while $j = 1, \ldots, n_i$ indexes the observations, $5 \leq n_i \leq 8$. These authors suggest a "one-compartment" nonlinear pharmacokinetic model wherein the mean plasma concentration $\eta_{ij}(x_{ij})$ is given by

$$\eta_{ij}(x_{ij}) = 30\alpha_i^{-1} \exp(-\beta_i x_{ij}/\alpha_i) \; .$$

Subsequent unpublished work by these same authors suggests this model is best fit on the log scale. That is, we suppose

$$Z_{ij} \equiv \log Y_{ij} = \log \eta_{ij}(x_{ij}) + \epsilon_{ij} \; ,$$

where $\epsilon_{ij} \overset{ind}{\sim} N(0, \tau_i)$. The mean structure for the Z_{ij}'s thus emerges as

$$\begin{aligned} \log \eta_{ij}(x_{ij}) &= \log \left[30\alpha_i^{-1} \exp(-\beta_i x_{ij}/\alpha_i) \right] \\ &= \log 30 - \log \alpha_i - \beta_i x_{ij}/\alpha_i \\ &= \log 30 - a_i - \exp(b_i - a_i) x_{ij} \; , \end{aligned}$$

where $a_i = \log \alpha_i$ and $b_i = \log \beta_i$.

Following the analysis by Wakefield et al. (1994), we assume the subject-specific random effects $\boldsymbol{\theta}_i \equiv (a_i, b_i)'$ are i.i.d. from a $N_2(\boldsymbol{\mu}, \Omega)$ distribution, where $\boldsymbol{\mu} = (\mu_a, \mu_b)$. These authors also

parameter	α updating method componentwise	vector
α_1	0.979	0.798
α_2	0.995	0.194
α_3	0.994	0.207
α_4	0.686	0.204
α_5	0.993	0.436
α_6	0.993	0.408
α_7	0.972	0.811
α_8	0.991	0.134
α_9	0.988	0.154
σ	0.529	0.530
D_{11}	0.385	0.388
D_{21}	0.942	0.942
D_{22}	0.891	0.891
D_{31}	0.934	0.934
D_{32}	0.967	0.967
D_{33}	0.917	0.918

Table 2: Lag 1 sample autocorrelations, algorithms for ddI/ddC data model

recommend the usual conjugate prior specification, namely $\boldsymbol{\mu} \sim N_2(\boldsymbol{\lambda}, C)$, $\tau_i \overset{iid}{\sim} G(\nu_0/2\ ,\ \nu_0\tau_0/2)$, and $\Omega \sim \text{Wishart}(R, \rho)$. Since the full conditional distributions of the random effects $\boldsymbol{\theta}_i$ are neither simple conjugate forms nor guaranteed to be log-concave, the new Metropolis capability of `BUGS` 0.6 is required. This Metropolis routine requires bounds to be placed on variables using the `I(,)` construction, so unfortunately, the model for $\boldsymbol{\theta}_i$ cannot be specifed bivariately as above, since `BUGS` currently cannot handle multivariate range restrictions. However, the model may still be specified in `BUGS` using the product formulation of the bivariate normal, namely

$$\begin{aligned} a_i &\sim N(\mu_a, \tau_a) I(L_a, U_a) \\ b_i|a_i &\sim N\left(k_0 + k_1(a_i - c)\ ,\ \tau_b\right) I(L_b, U_b)\ , \end{aligned}$$

where (L_a, U_a) and (L_b, U_b) are broad truncation regions to enable the grid-based Metropolis algorithm, and c is a constant used to roughly center the a_i's (hence reduce correlation between the intercept k_0 and slope k_1). Under this formulation, we replace the normal prior for $\boldsymbol{\mu}$ and the Wishart prior for Ω with gamma priors for τ_a and τ_b and normal priors for μ_a, k_0 and k_1.

The `BUGS` code to fit this model follows. As can be seen, we adopt the tuning constants $c = 3$, $L_a = L_b = -5$, and $U_a = U_b = 10$. The latter values comfortably contain all the posterior mass for the a_i and b_i; significantly more widely dispersed values (say, $L_a = L_b = -50$ and $U_a = U_b = 100$) do in fact lead to sharp drops in the Metropolis acceptance rate, hence reductions in efficiency.

```
model PK;
const
   N = 10,    # number of patients
   T = 8;     # number of time points
var
```

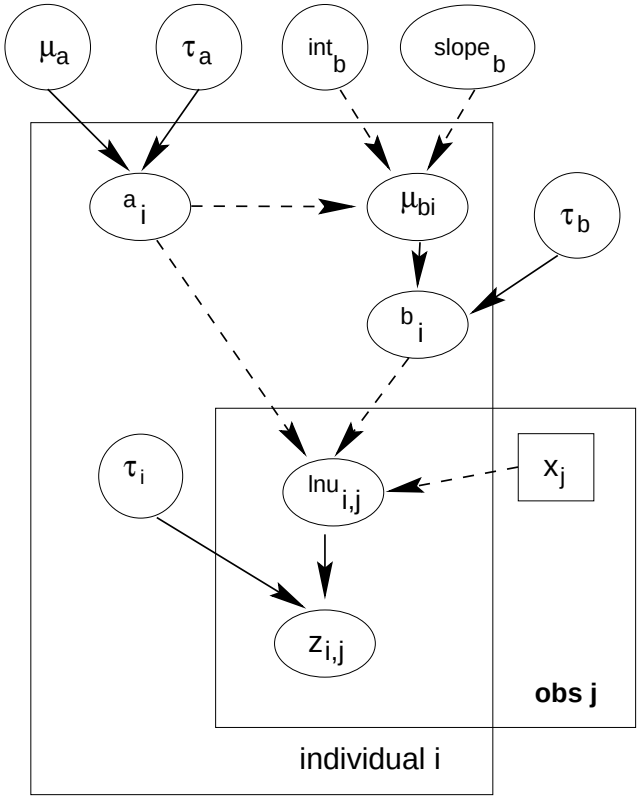

Figure 4: Graph of the PK example.

patient	no. of hours following drug administration							
	2	4	6	8	10	24	28	32
1	1.09	0.75	0.53	0.34	0.23	0.02	–	–
2	2.03	1.28	1.20	1.02	0.83	0.28	–	–
3	1.44	1.30	0.95	0.68	0.52	0.06	–	–
4	1.55	0.96	0.80	0.62	0.46	0.08	–	–
5	1.35	0.78	0.50	0.33	0.18	0.02	–	–
6	1.08	0.59	0.37	0.23	0.17	–	–	–
7	1.32	0.74	0.46	0.28	0.27	0.03	0.02	–
8	1.63	1.01	0.73	0.55	0.41	0.01	0.06	0.02
9	1.26	0.73	0.40	0.30	0.21	–	–	–
10	1.30	0.70	0.40	0.25	0.14	–	–	–

Table 3: Cadralazine concentration data

```
   X[T],Z[N,T],theta[N,2],a[N],b[N],lnu[N,T],tau[N],sigma[N],Y[N,T],
   mu.a,tau.a,mub[N],tau.b,int.b,slope.b,mu.b;
data Z in "PKZ.dat", X in "PKX.dat";
inits in "PK.in";
{
    for (i in 1:N) {
      for (j in 1:T) {
        Z[i,j] ~ dnorm(lnu[i,j],tau[i]);
        Y[i,j] <- exp(Z[i,j]);
        lnu[i,j] <- log(30) - a[i] - exp(b[i]-a[i])*X[j];
} # end of j loop

    a[i] ~ dnorm(mu.a,tau.a) I(-5, 10);
    b[i] ~ dnorm(mub[i],tau.b) I(-5, 10);
    mub[i] <- int.b + slope.b * (a[i] - 3.0);  #  center the a_i's

    tau[i]  ~ dgamma(.0001, .0001);  sigma[i] <- 1.0/sqrt(tau[i]);
} # end of i loop

  mu.a ~ dnorm(0.0, 0.0001);
  int.b ~ dnorm(0.0, 0.0001);  slope.b ~ dnorm(0.0, 0.0001);
  mu.b <- int.b + slope.b * (mu.a - 3.0);
  tau.a ~ dgamma(1, 0.04); tau.b ~ dgamma(1, 0.04); # vague Wakefield prior

} # end of PK.bug program
```

Using a sequence of univariate Metropolis (Gaussian proposals) and Gibbs steps, Sargent *et al.* (1997) fit the original Wishart formulation of this model using the priors recommended by Wakefield et al. (1994), namely $\nu_0 = 0$, $\boldsymbol{\lambda} = \mathbf{0}$, $C^{-1} = Diag(0.01, 0.01)$, $\rho = 2$, and $R = \rho * Diag(0.04, 0.04)$. We attempt a comparable prior in our formulation by taking $G(0.0001, 0.0001)$ priors for the τ_i, $N(0, 0.0001)$ priors for μ_a, k_0 and k_1, and $G(1, 0.04)$ priors for τ_a and τ_b.

Running this `BUGS` code for 5000 iterations (following a 250 iteration burn-in period) produces

	BUGS V0.6			Sargent et al. (1997)		
parameter	mean	sd	lag 1 acf	mean	sd	lag 1 acf
a_1	2.956	0.0479	0.969	2.969	0.0460	0.947
a_2	2.692	0.0772	0.769	2.708	0.0910	0.808
a_7	2.970	0.1106	0.925	2.985	0.1360	0.938
a_8	2.828	0.1417	0.828	2.838	0.1863	0.934
b_1	1.259	0.0335	0.972	1.268	0.0322	0.951
b_2	0.234	0.0648	0.661	0.239	0.0798	0.832
b_7	1.157	0.0879	0.899	1.163	0.1055	0.925
b_8	0.936	0.1458	0.759	0.941	0.1838	0.932
τ_1	362.4	260.4	0.313	380.8	268.8	0.220
τ_2	84.04	57.60	0.225	81.40	58.41	0.255
τ_7	18.87	12.07	0.260	15.82	11.12	0.237
τ_8	2.119	1.139	0.085	1.499	0.931	0.143
μ_a	2.838	0.0715	0.421	2.829	0.0740	0.870
μ_b	1.051	0.1472	0.0457	1.049	0.1371	0.350
k_0	1.324	0.2087	0.769	–	–	–
k_1	1.773	1.154	0.826	–	–	–
τ_a	44.58	23.24	0.350	46.25	23.57	0.400
τ_b	14.75	12.72	0.673	–	–	–
$Y_{2,8}$	0.1338	0.0339	0.288	0.1347	0.0264	–
$Y_{7,8}$	0.00891	0.00443	0.178	0.00884	0.00255	–

Table 4: Posterior summaries and lag 1 sample autocorrelations, PK data model

the posterior summaries and lag 1 sample autocorrelations given in Table 4. Also shown are the results produced by Sargent *et al.* (1997), to which the `BUGS` results are quite comparable, given the slight differences in model and prior formulation. (Results for $\Omega_{1,1}$ in the Wishart model are given in the τ_a row of the table; however, results for $\Omega_{2,2}$ are not shown in the τ_b row since τ_b is a *conditional* precision, given the a_i.) The relatively large posterior means for τ_7 and τ_8 (and correspondingly large posterior variances for a_7, a_8, b_7, and b_8) at first seem counter-intuitive, since these two patients had the most data available for study. However, their final 2 to 3 observations fit the overall model poorly (with those for patient 8 not even being monotone), explaining this oddity. Finally (and relatedly), note the predicted values of the final observations for patients 2 (whose clearance rate is the slowest) and 7 (whose rate is amongst the fastest). The former has mean somewhat larger than that suggested by the posterior predictive distribution under the "power model" fit on the original (unlogged) scale by Wakefield *et al.* (1994).

Acknowledgements

We are grateful to all those people who have proposed ways of improving the `BUGS` software, and apologise for implementing so few of their suggestions. We are particularly grateful for Brad Carlin for providing the `ddIddC` and the `PK` examples.

References

Abrams, D., Goldman, A., Launer, C., Korvick, J., Neaton, J., Crane, L., Grodesky, M., Wakefield, S., Muth, K., Kornegay, S., Cohn, D., Harris, A., R., L.-H., Markowitz, N., Sampson, J., Thompson, M., Deyton, L., and the Terry Beirn Community Programs for Clinical Research on AIDS (1994). Comparative trial of Didanosine and Zalcitabine in patients with human immunodeficiency virus infection who are intolerant of or have failed Zidovudine therapy. *New England J Medicine*, **330**, 657–62.

Carlin, B. P. (1996). Hierarchical longitudinal modelling. In *Markov Chain Monte Carlo Methods in Practice*, (ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter), pp. 303–19. Chapman and Hall, New York.

Carlin, B. P. and Gelfand, A. E. (1991). An iterative Monte Carlo method for nonconjugate Bayesian analysis. *Statistics and Computing*, **1**, 119–28.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London, U.K.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parameterization for normal linear mixed models. *Biometrika*, **82**, 479–88.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1996). Efficient parameterizations for generalised linear models (with discussion). In *Bayesian Statistics 5*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 165–80. Clarendon Press, Oxford, UK.

Laird, N. M. and Ware, J. H. (1982). Randon effects models for longitudinal data. *Journal of the American Statistical Association*, **38**, 963–74.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673–87.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, **87**, 861–8.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structures, blocking and parameterization of the Gibbs sampler. *Journal of the Royal Statistical Society, Series B*, **59**, 291–317.

Sargent, D., Hodges, J., and Carlin, B. (1997). Structured Markov chain Monte Carlo. Technical report, Research Report 97–009, Division of Biostatistics, University of Minnesota.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996*a*). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5, (version ii)* . MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996*c*). *BUGS Examples Volume 1, Version 0.5, (version ii)*. MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1996*b*). *BUGS Examples Volume 2, Version 0.5, (version ii)*. MRC Biostatistics Unit, Cambridge.

Wakefield, J., Smith, A., Racine-Poon, A., and Gelfand, A. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics*, **43**, 201–21.