A weakly informative prior for Bayesian dynamic model selection with applications in fMRI

Jairo A. Fúquene Patiño[a], Brenda Betancourt[b], João B. M. Pereira[c]

[a] Department of Statistics, University of Warwick, UK.
[b] Department of Statistical Science, Duke University, USA.
[c] Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil.

May 31, 2017

Appendix A: MCMC scheme

We present the Markov Chain Monte Carlo (MCMC) algorithm for sampling from the joint posterior distribution of the parameters of the proposed model. We need to obtain posterior inference for the set of parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_t, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_i\}$. In particular, for sampling from the full conditional distributions of these parameters, we consider Gibbs sampling and Metropolis-Hastings algorithms. The MCMC scheme can be summarized in the following steps:

- Sample the states $(\boldsymbol{\theta}_t|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$. In order to obtain posterior inference on the state parameters $\boldsymbol{\theta}_t = \{\theta_{0,t}, \theta_{1,t}, \theta_{2,t}, \ldots, \theta_{p,t}\}$, we use the forward-filtering backward-sampling (FFBS) method proposed by Fruwirth-Schnatter (1994), which is practically a simulation of the smoothing recursions.

- Sample $(\lambda_y|\lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_i, \boldsymbol{y}_{1:t})$ in a Gibbs step using a Gamma distribution,

$$\lambda_y|... \sim \text{Gamma}\left(T+2, \frac{1}{2}SSy^* + \frac{1}{2}SS^*_{\theta,i}\right),$$

  where $SSy^* = \sum_{t=1}^T \omega_{y,t}(y_t - F_t\theta_t)^2$ and $SS^*_{\theta,i} = \sum_{t=1}^T \lambda_{\theta,i}\omega_{\theta,t_i}(\theta_{t_i} - (G_t\theta_{t-1})_i)^2$.

- Sample $(\lambda_{\theta,i}|\boldsymbol{\alpha}, \lambda_y, \omega_{\theta,t_i}, \rho_{t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution,

$$\lambda_{\theta,i}|\cdot \sim \text{Gamma}\left(q + \frac{T+\nu_{t_i}-1}{2}, \frac{1}{2}SS^{**}_{\theta,i} + \beta_{t_i}\rho_{t_i}\right),$$

  where $SS^{**}_{\theta,i} = \sum_{t=1}^T \lambda_y\omega_{\theta,t_i}(\theta_{t_i} - (G_t\theta_{t-1})_i)^2$.

- Sample $(\omega_{\theta,t_i}|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \rho_{t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution,

$$\omega_{\theta,t_i}|\cdot \sim \text{Gamma}\left(\frac{\nu_{t_i}+1}{2}, \frac{\nu_{t_i}+\lambda_y\lambda_{\theta,i}(\theta_{t_i}-(G_t\theta_{t-1})_i)^2}{2}\right).$$

- Sample $(\rho_{t_i}|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \nu_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution,

$$\rho_{t_i}|\cdot \sim \text{Gamma}\left(\frac{\nu_{t_i}+1}{2}, \beta_{t_i}\lambda_{\theta,i}+1\right).$$

- Sample $(\beta_{t_i}|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}\nu_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution,

$$\beta_{t_i}|\cdot \sim \text{Gamma}\left(\frac{\nu_{t_i}+1}{2}, \rho_{t_i}\lambda_y\lambda_{\theta,i}+\xi_{t_i}\right).$$

- Sample $(\xi_{t_i}|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}\nu_{t_i}, \beta_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution,

$$\xi_{t_i}|\cdot \sim \text{Gamma}\left(2, \beta_{t_i}+1\right).$$

- Sample $(\nu_{t_i}|\boldsymbol{\alpha}, \lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ in a Metropolis-Hastings step using ARMS

$$p(\nu_{t_i}=k) \propto \text{Gamma}(\omega_{\theta,t_i}|(\upsilon_{t_i}=k+1)/2, \rho_{t_i}/\beta_{t_i}) \times \text{Gamma}(\lambda_{\theta,i}|\nu_{\theta,t_i=k}, \nu_{\theta,t_i=k})\varphi_i,$$

on the set $\{n_1, n_2, \ldots, n_k\}$ for $i = 1, \ldots, p$ and $t = 1, \ldots, T$;

$$\varphi_i|... \sim \text{Dirichlet}(\alpha_i, N_I), \tag{0.1}$$

where $N_I = \{N_{i,1}, ...N_{i,k}\}$ with, for each k, $N_{i,k} = \sum_{i=1}^{T}(\nu_{t_i}=k)$, for $i = 1, ..., p$.

- Sample $(\phi_{i,j}|\boldsymbol{\alpha}, \nu_{t_i}\lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \beta_{t_i}, \xi_{t_i}, \zeta_{i,j}, \boldsymbol{y}_{1:t})$ from

$$\phi_{i,j} = \begin{cases} 0 & \text{with probability } \ ,1-\pi^*_{ij} \\ \\ N(\phi_{i,j}|\mu_{i,j}, \beta^2_{i,j}) & \text{with probability } \ \pi^*_{ij}. \end{cases} \tag{0.2}$$

where

$$\pi^*_{ij} = \frac{\pi p_1(\boldsymbol{\theta^*_{i,j,t}})}{\pi p_1(\boldsymbol{\theta^*_{i,t}}) + (1-\pi) p_2(\boldsymbol{\theta^*_{i,j,t}})}, \qquad \beta_{i,j} = \left[\frac{1}{\eta^2_{ij}} + \boldsymbol{\theta^{*\prime}_{j,t-1}} \boldsymbol{\lambda_{\theta,i}} \boldsymbol{\theta^*_{j,t-1}}\right]^{-1},$$

and

$$\mu_{i,j} = \beta_{i,j}\left[\boldsymbol{\theta}^{*'}_{j,t-1}\lambda_{\theta,i}\boldsymbol{\theta}^*_{i,j,t}\right], \qquad p_1(\boldsymbol{\theta}^*_{i,j,t}) = N_T(\boldsymbol{\theta}^*_{i,j,t}|0, \lambda^{-1}_{\theta,i} I + \zeta_{i,j}\boldsymbol{\theta}^*_{j,t-1}\boldsymbol{\theta}^{*'}_{j,t-1}),$$

$$p_2(\boldsymbol{\theta}^*_{i,j,t}) = N_T(\boldsymbol{\theta}^*_{i,j,t}|0, \lambda^{-1}_{\theta,i} I) \tag{0.3}$$

where $p_1(\boldsymbol{\theta}^*_{i,j,t})$ and $p_2(\boldsymbol{\theta}^*_{i,j,t})$ are the predictive multivariate normal densities. Here

$$\theta^*_{i,j,t} = \theta_{i,t} - \sum_{k\neq j} \phi_{i,k}(x_{k,t-1})(\theta_{k,t-1}) = \phi_{i,j}(x_{j,t-1})(\theta_{j,t-1}) + \omega_{i,t}. \tag{0.4}$$

Using the usual notation for linear regression:

$$\begin{pmatrix} \theta^*_{i,j,1} \\ \theta^*_{i,j,2} \\ . \\ . \\ . \\ \theta^*_{i,j,T} \end{pmatrix} = \begin{pmatrix} (x_{j,0})(\theta_{j,0}) \\ (x_{j,1})(\theta_{j,1}) \\ . \\ . \\ . \\ (x_{j,T-1})(\theta_{j,T-1}) \end{pmatrix} \phi_{i,j} + \begin{pmatrix} \omega_{i,1} \\ \omega_{i,2} \\ . \\ . \\ . \\ \omega_{i,T} \end{pmatrix}, \tag{0.5}$$

which can be expressed in matrix notation as follows:

$$\begin{aligned} \boldsymbol{\theta}^*_{i,j,t} &= \boldsymbol{\theta}^*_{j,t-1}\phi_{i,j} + \boldsymbol{\omega}_{i,t} & t &= 1, 2, ..., T \\ \boldsymbol{\theta}^*_{j,t-1} &= \boldsymbol{x}_{j,t-1}\boldsymbol{\theta}_{j,t-1} & \boldsymbol{\omega}_{i,t} &\sim N(0, \lambda^{-1}_{\theta,i} I). \end{aligned}$$

- Sample $(\zeta_{i,j}|\boldsymbol{\alpha}, \nu_{t_i}\lambda_y, \lambda_{\theta,i}, \omega_{\theta,t_i}, \rho_{t_i}, \beta_{t_i}, \xi_{t_i}, \phi_{i,j}, \boldsymbol{y}_{1:t})$ in a Gibbs step from a Gamma distribution, Gamma$(\zeta_{i,j}|a_{\zeta_{i,j}}, b_{\zeta_{i,j}})$, with:

$$a_{\zeta_{i,j}} = c + \frac{3}{2}, \qquad b_{\zeta_{i,j}} = d + \frac{\phi^2_{i,j}}{2}. \tag{0.6}$$

Appendix B: The marginal prior for states

Proof. Integrating out $\omega_{\theta,t}$, we have $\theta_t \sim$ Student's $\mathrm{t}(G_t\theta_{t-1}, \sigma\tau_t, \upsilon_t)$ where υ_t corresponds to the degrees of freedom, $G_t\theta_{t-1}$ corresponds to the location and $\sigma\tau_t$ corresponds to the scale of the Student's t density:

$$\pi(\theta_t|G_t, \theta_{t-1}, \sigma, \tau_t, \upsilon_t) = \frac{k_1}{\sigma\tau_t}\left(1 + \frac{1}{\upsilon_t}\left(\frac{\theta_t - G_t\theta_{t-1}}{\sigma\tau_t}\right)^2\right)^{-(\upsilon_t+1)/2}$$

where $k_1 = \dfrac{\Gamma((\upsilon_t+1)/2)}{\Gamma(\upsilon_t/2)\sqrt{\upsilon_t\pi}}$. Using the fact that $\tau_t^2 \sim \mathrm{Gamma}(1, \beta_t/\rho_t)$ and $\rho_t \sim \mathrm{Gamma}((\upsilon_t - 1)/2, 1)$, we have that

$$\pi(\theta_t|G_t\theta_{t-1}) = \begin{cases} k(\sigma\beta_t)^{(\nu_t-1)/2}\nu_t^{\nu_t/2}/(\theta_t - G_t\theta_{t-1})^{\nu_t/2}2F1((\upsilon_t+1)/2, \upsilon_t/2, \upsilon_t/2+1, 1-\sigma\beta_t\nu_t/(\theta_t - G_t\theta_{t-1})^2), \\ \text{if} \quad \theta_t \neq G_t\theta_{t-1}, \\ \\ k_1\mathrm{Be}(1/2, \upsilon_t/2)/(\beta_t^{1/2}\mathrm{Be}(1, (\upsilon_t-1)/2)), \quad \text{if} \quad \theta_t = G_t\theta_{t-1}, \end{cases}$$

where $k = k_1\mathrm{Be}(\upsilon_t/2, 1+\upsilon_t/2)/\mathrm{Be}(1, (\upsilon_t-1)/2)$. Here, $Be(a, b)$ denotes the beta function and $2F1(a, b, c, z)$ denotes the hypergeometric function. Finally, we use the identities (15.3.5) and (15.1.13) for $\theta_t \neq G_t\theta_{t-1}$ from the book of Abramowitz & Stegun (1970).

□

Appendix C: Plots from simulation study

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 2$ and $\lambda_{\theta,i} = 1$ known

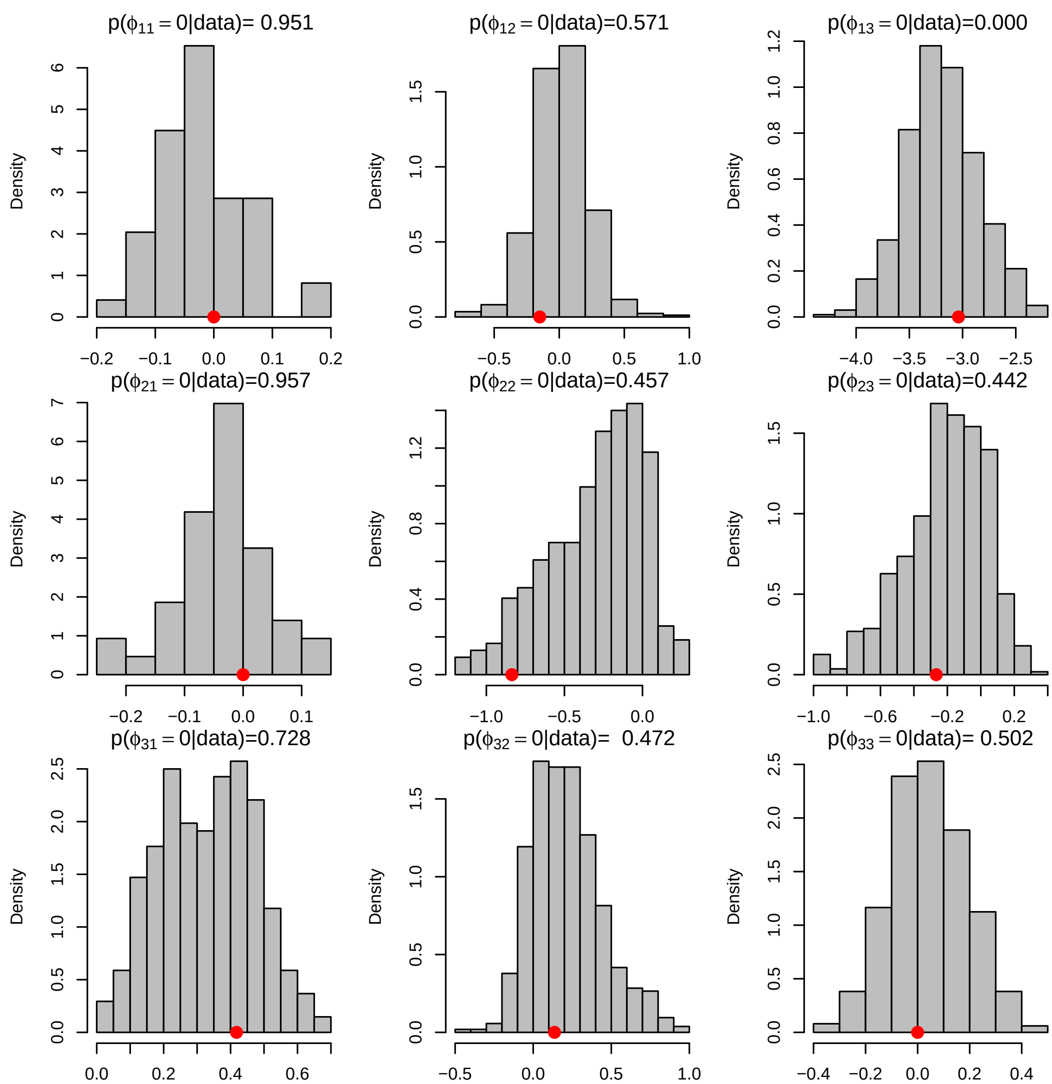

Figure 1: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

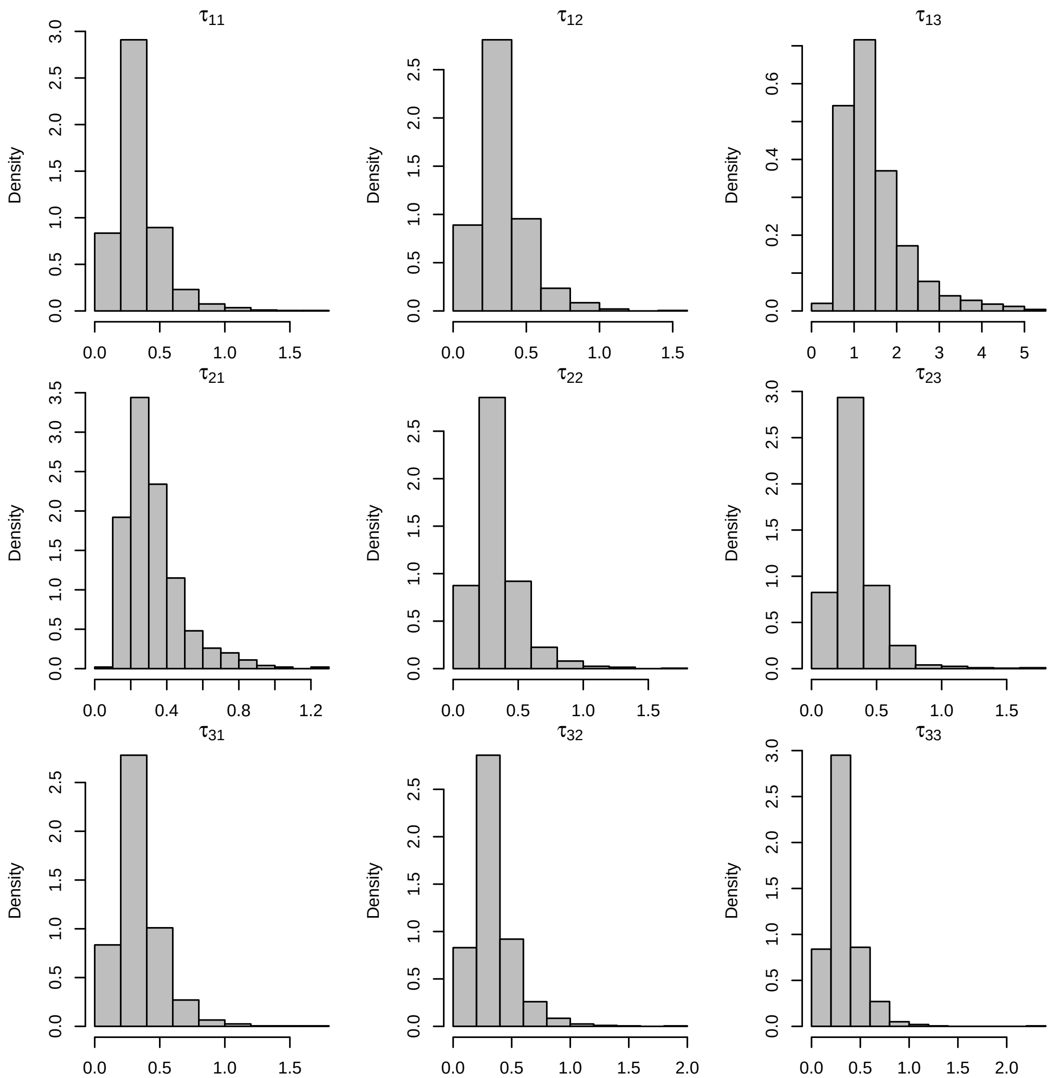

Figure 2: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

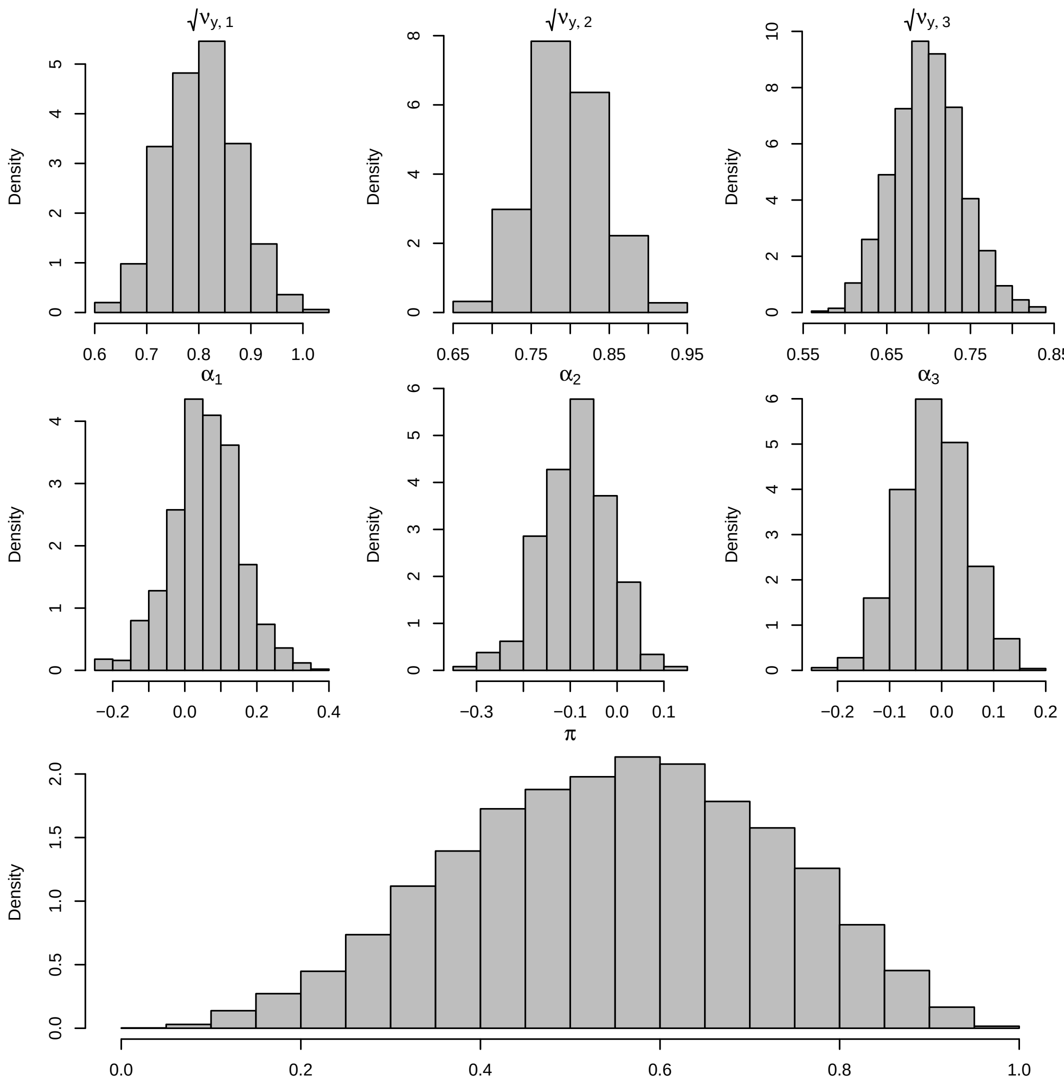

Figure 3: Posterior distribution: observational variances, trends and weights.

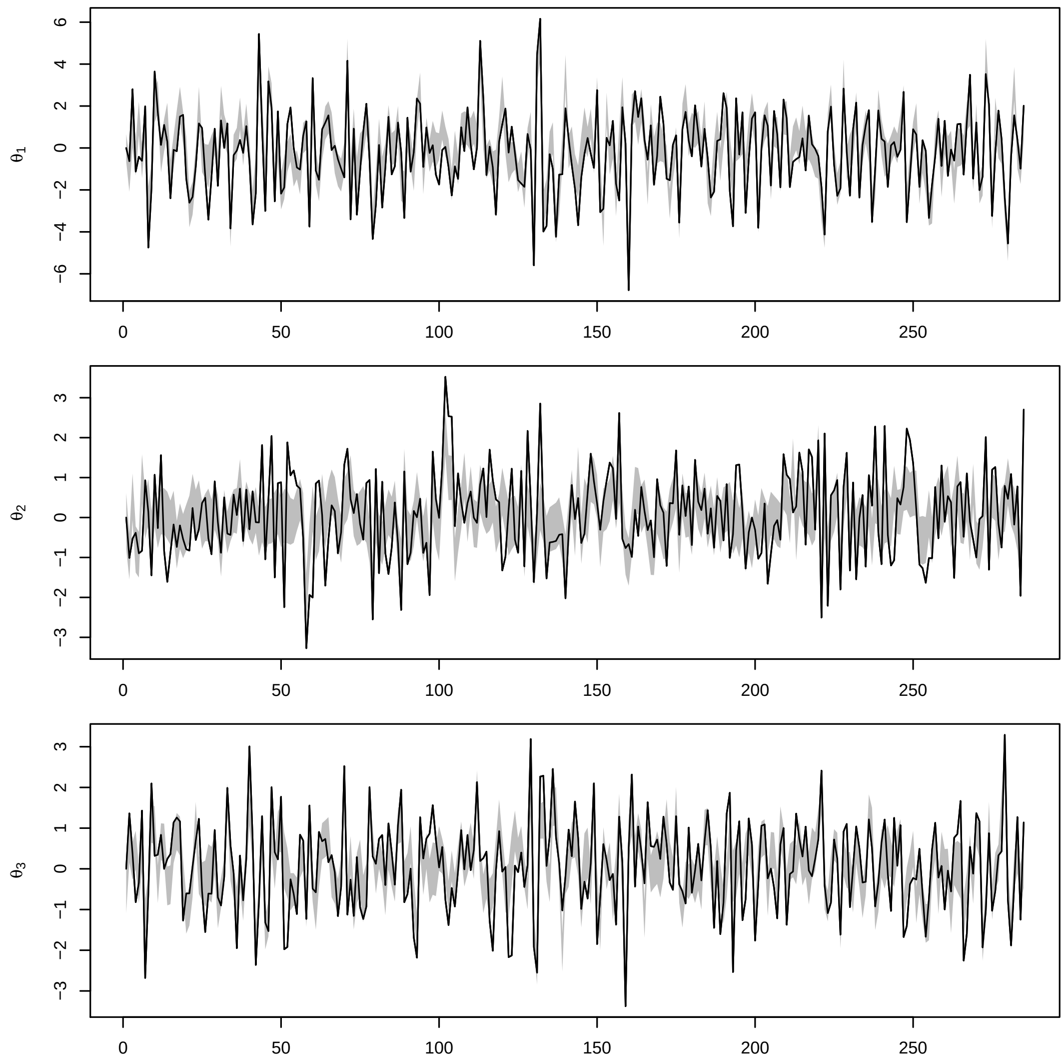

Figure 4: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 2$ and $\lambda_{\theta,i} = 1$ unknown

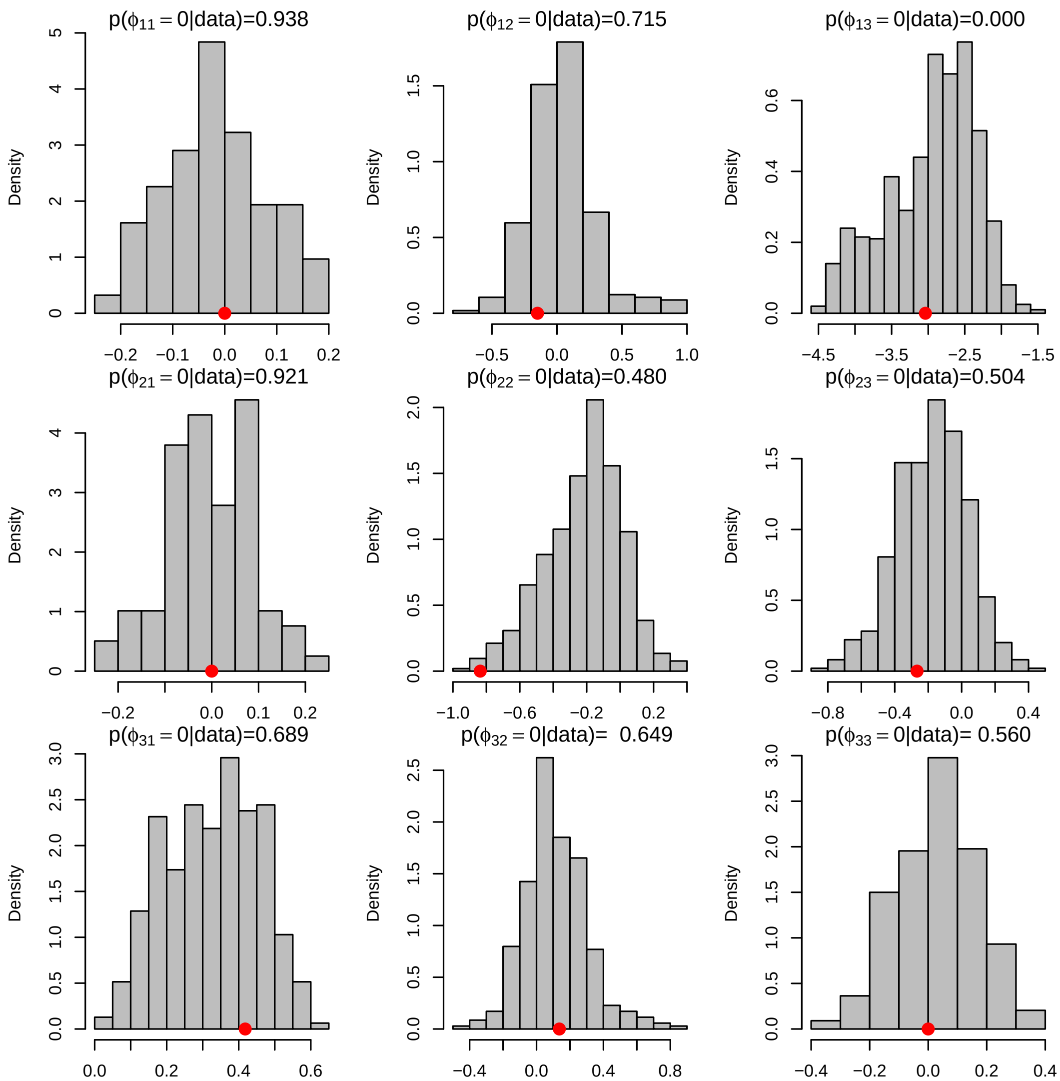

Figure 5: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

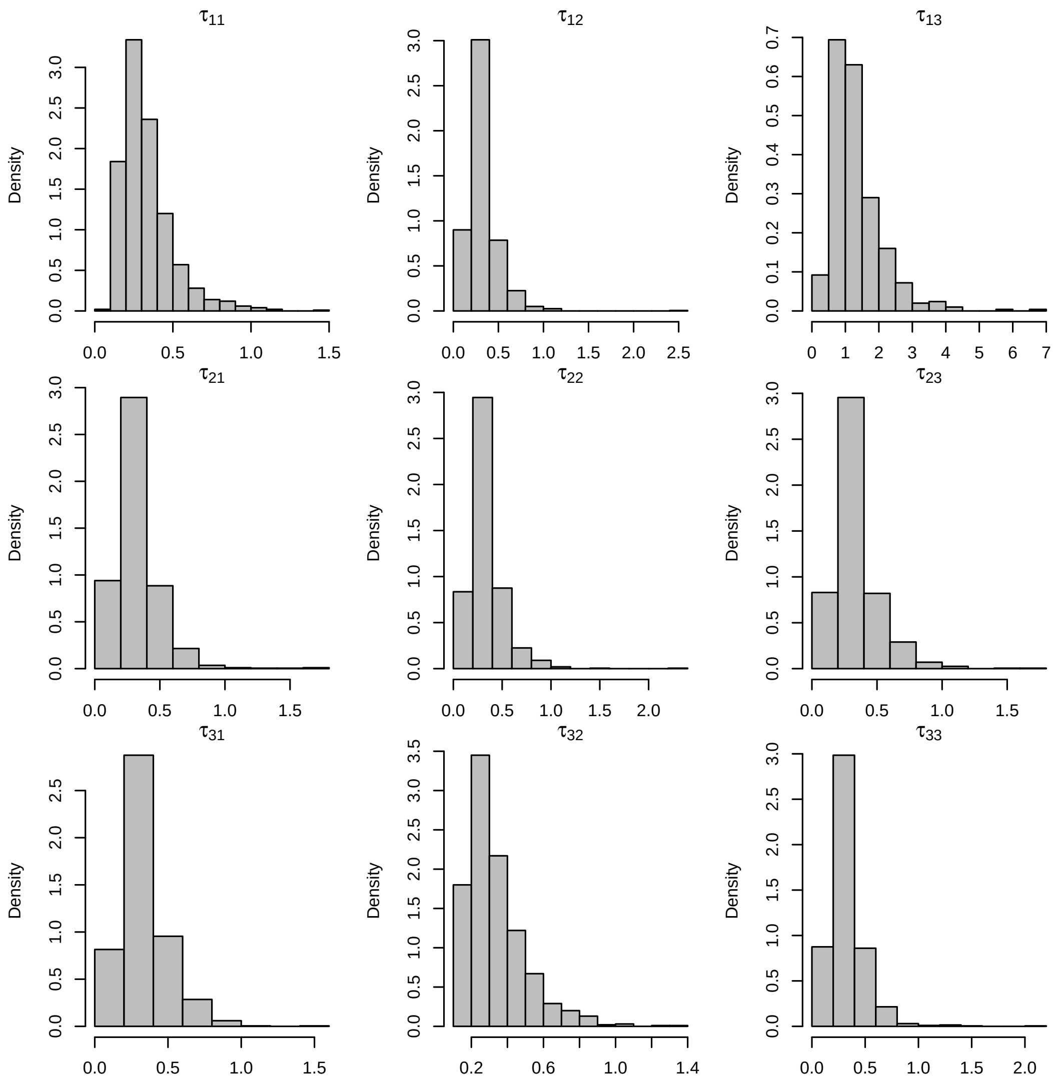

Figure 6: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

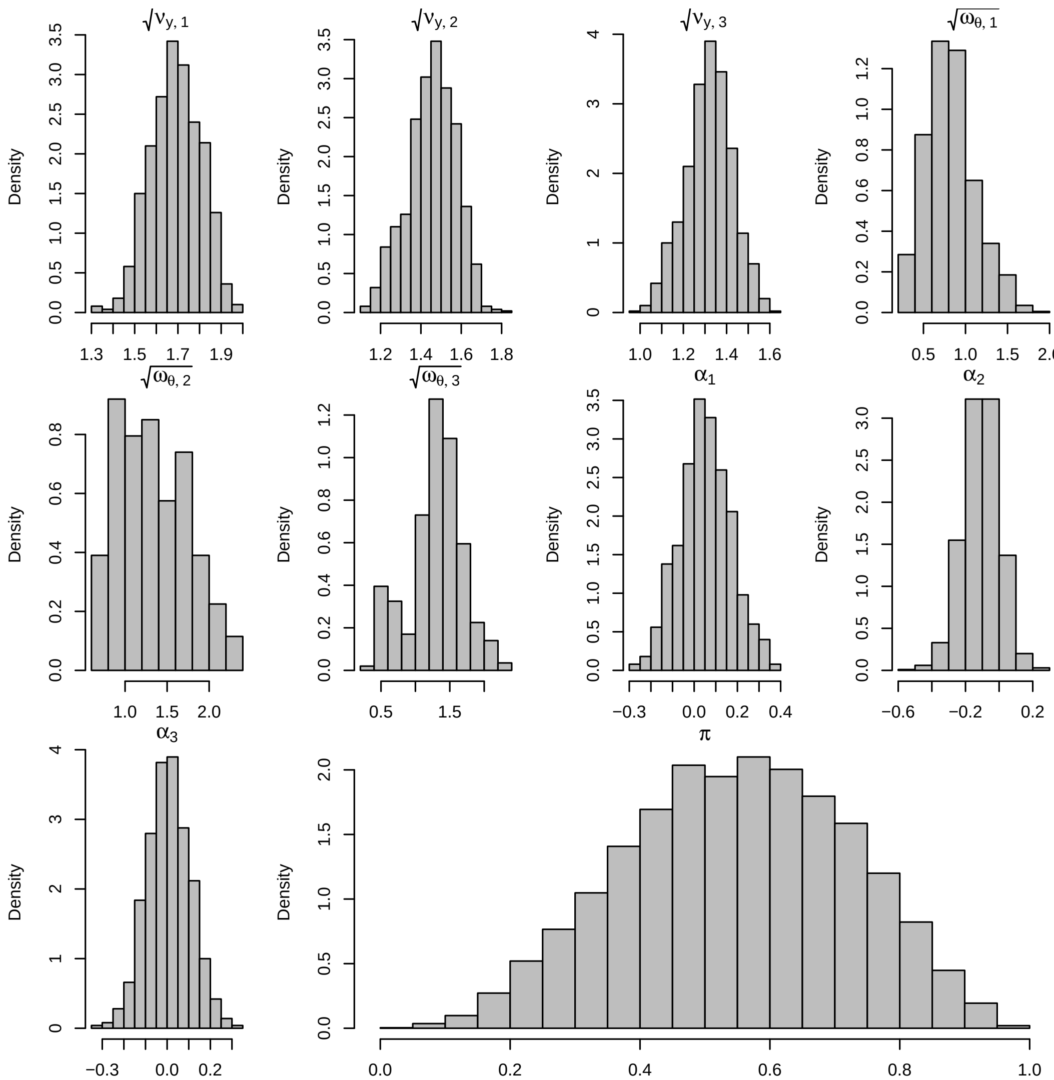

Figure 7: Posterior distribution: observational and state variances, trends and weights.

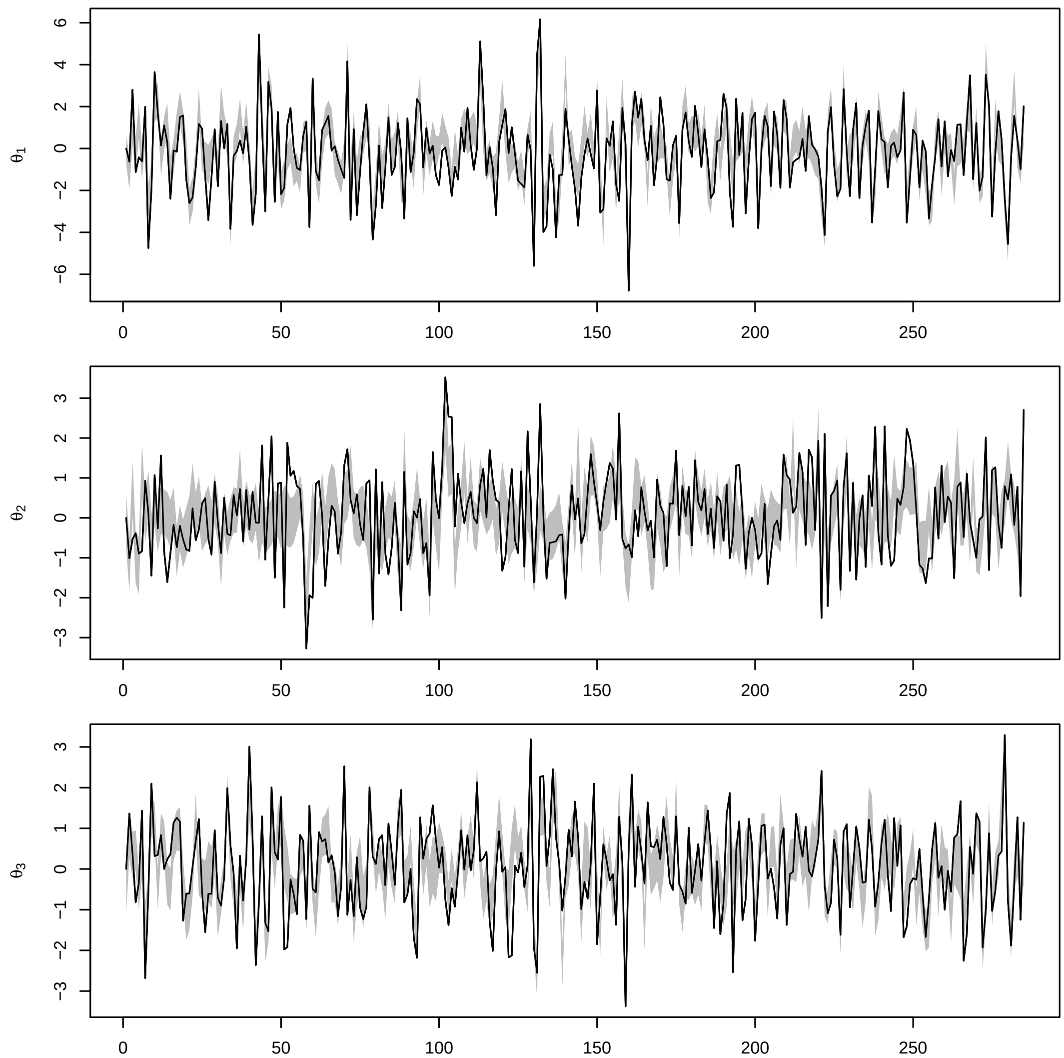

Figure 8: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 1$ and $\lambda_{\theta,i} = 1$ known

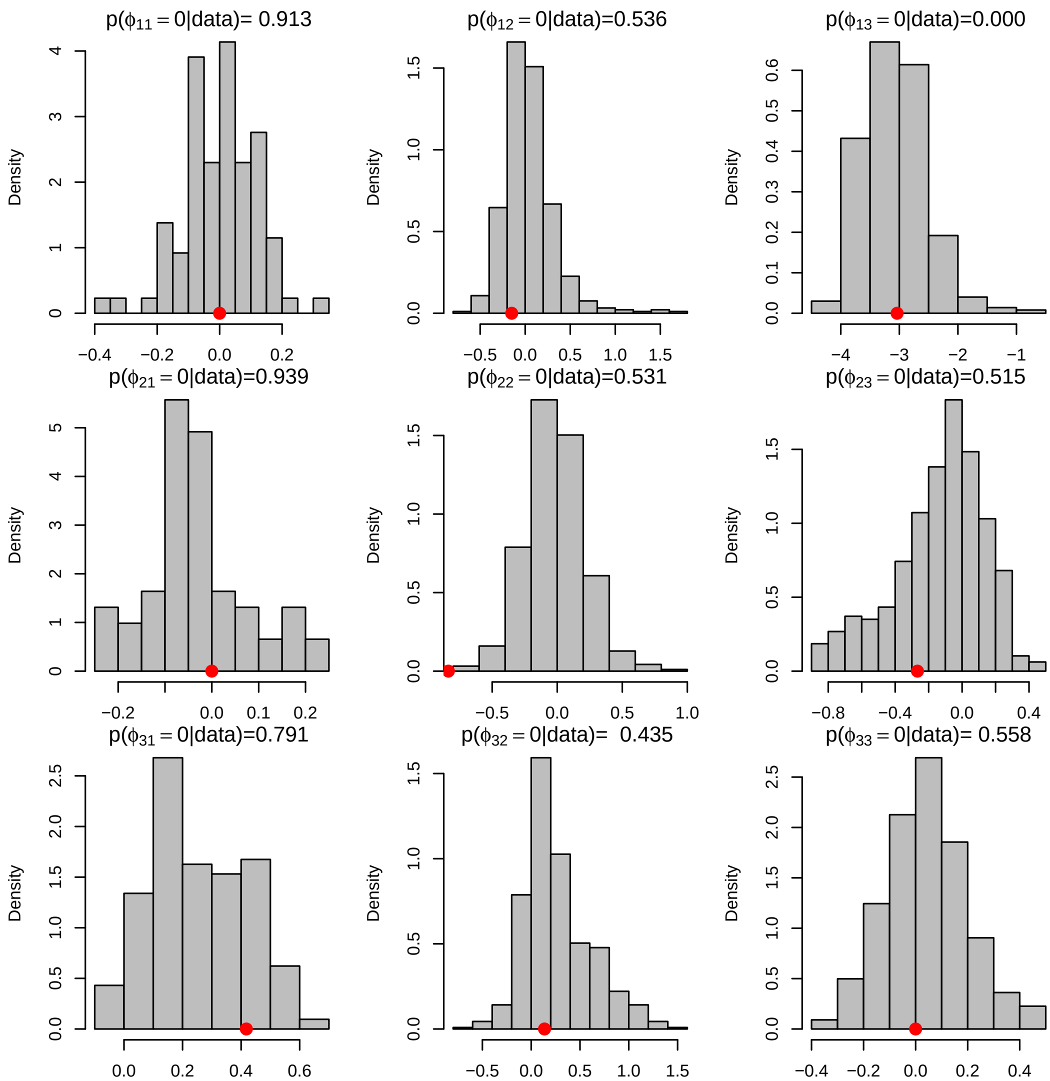

Figure 9: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

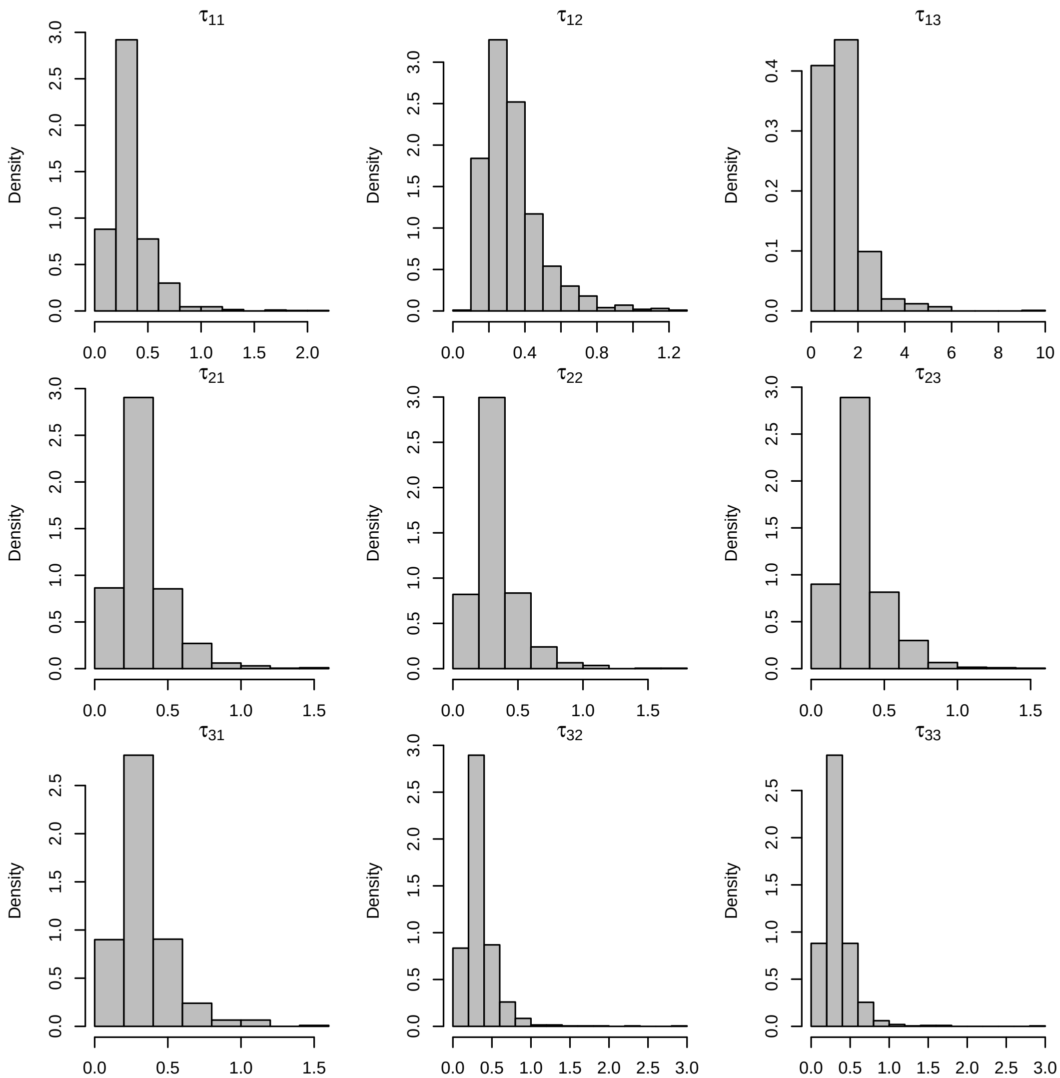

Figure 10: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

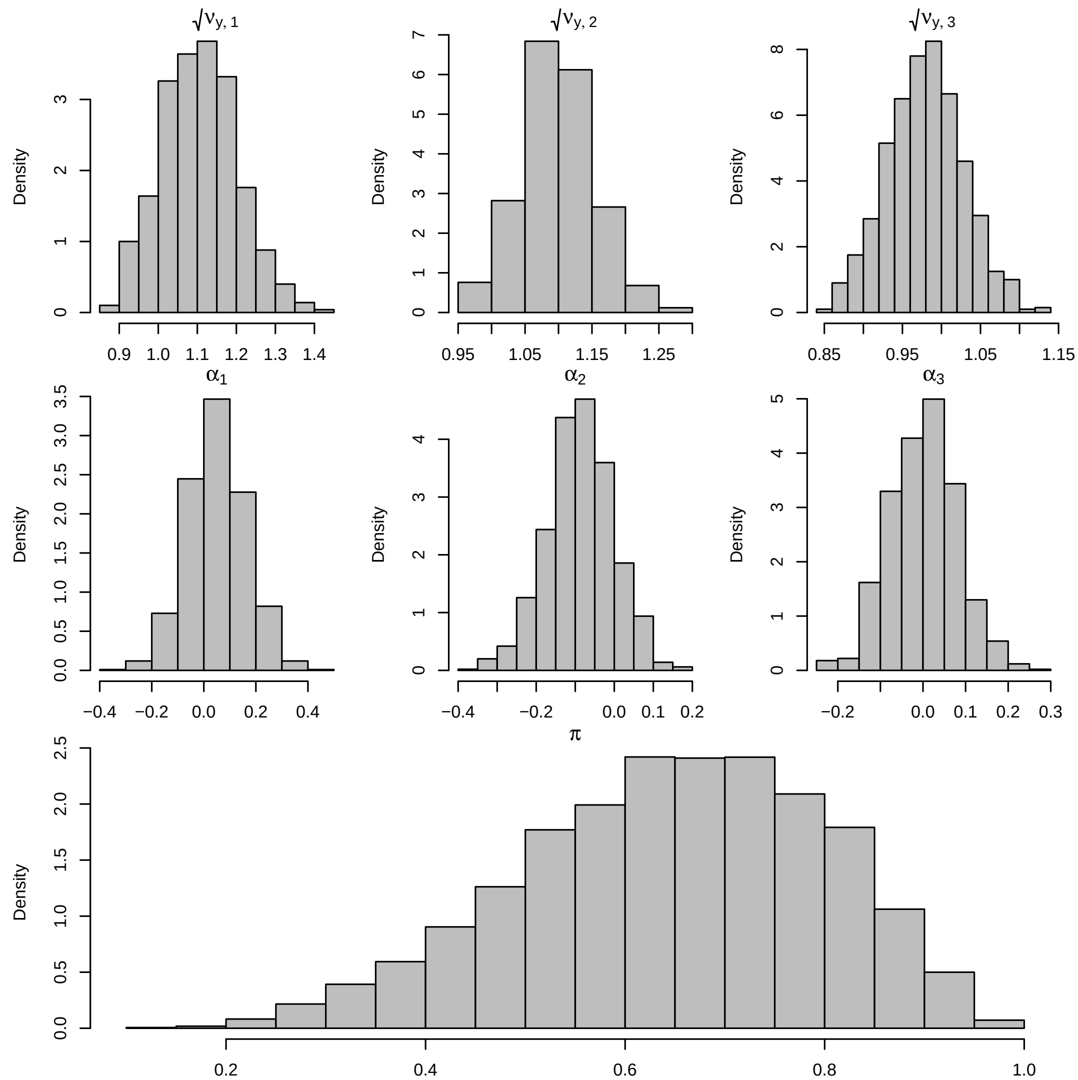

Figure 11: Posterior distribution: observational variances, trends and weights.

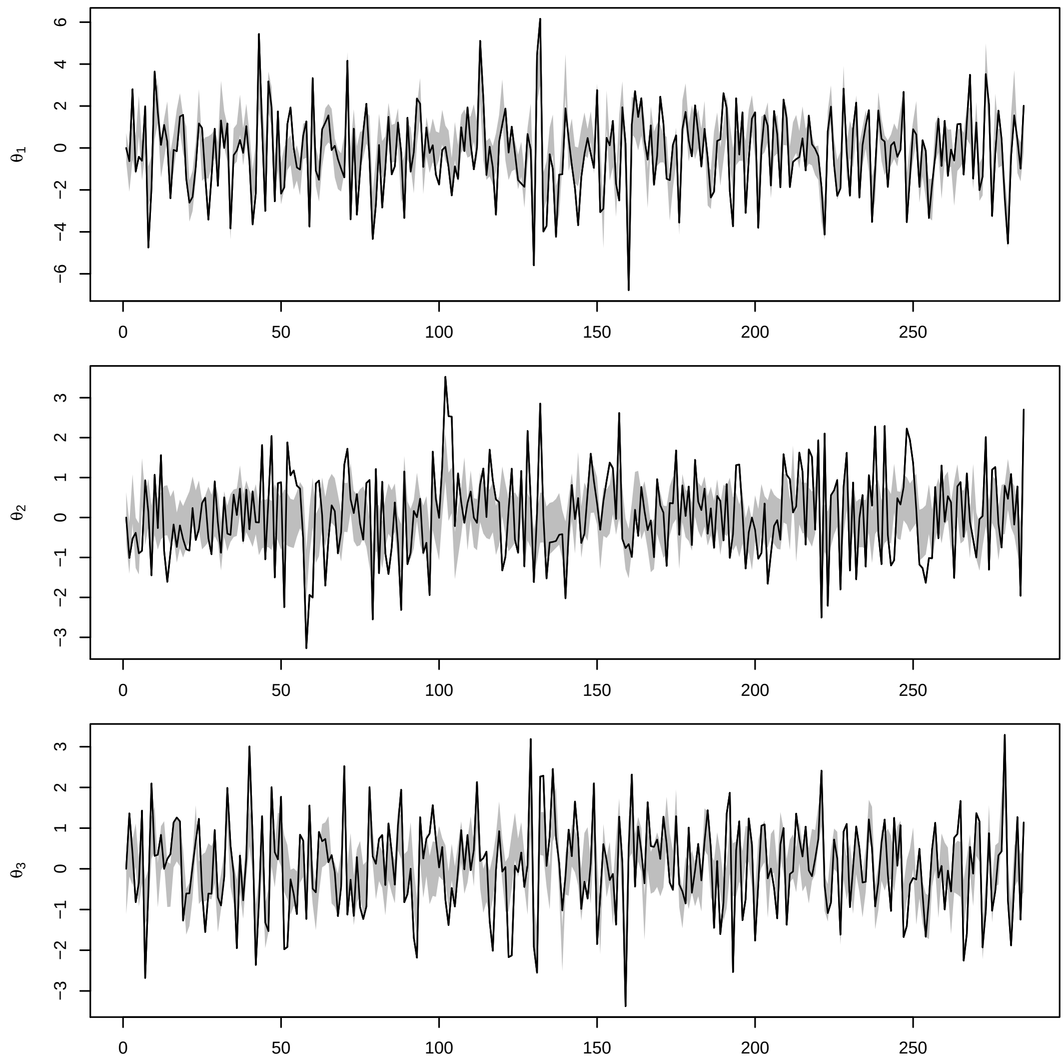

Figure 12: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 1$ and $\lambda_{\theta,i} = 1$ unknown

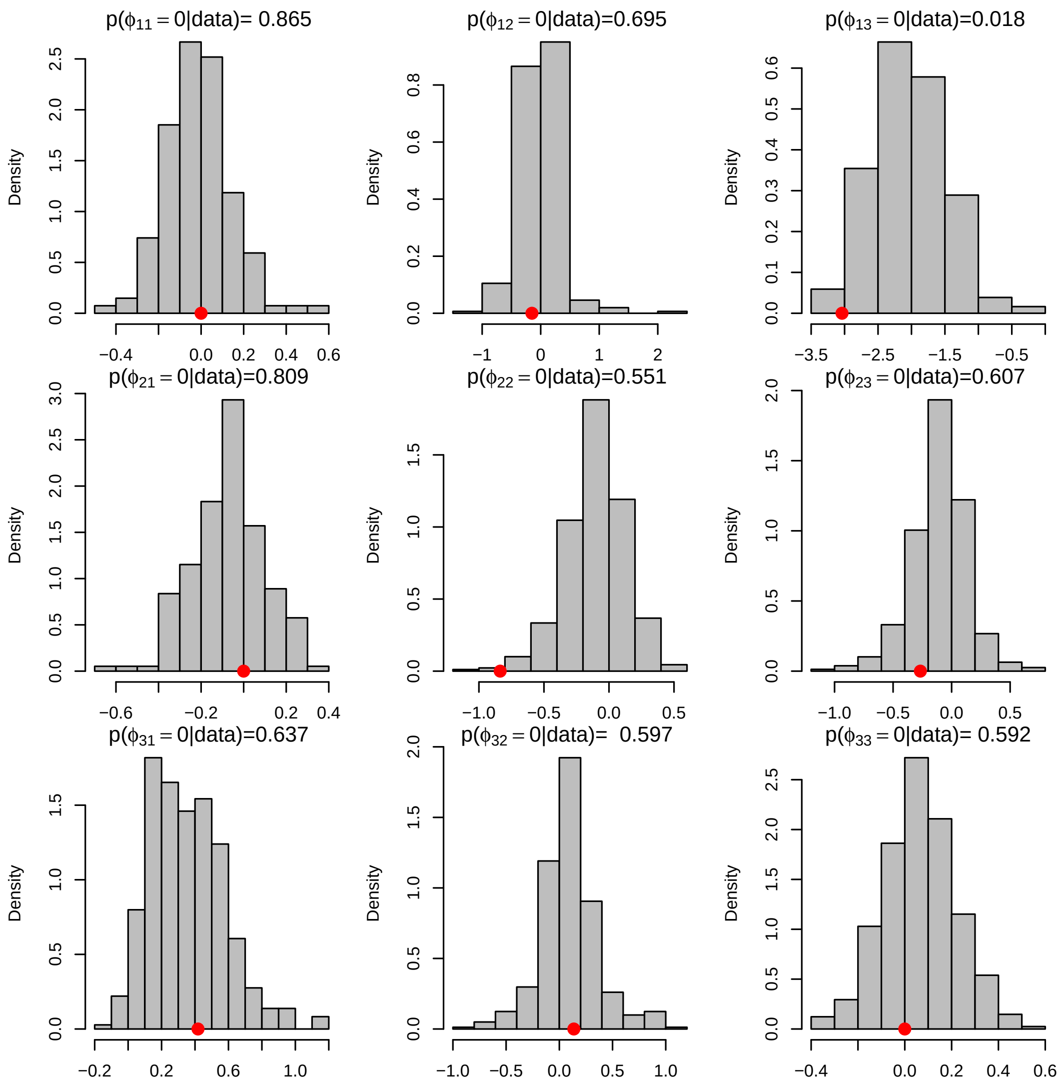

Figure 13: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

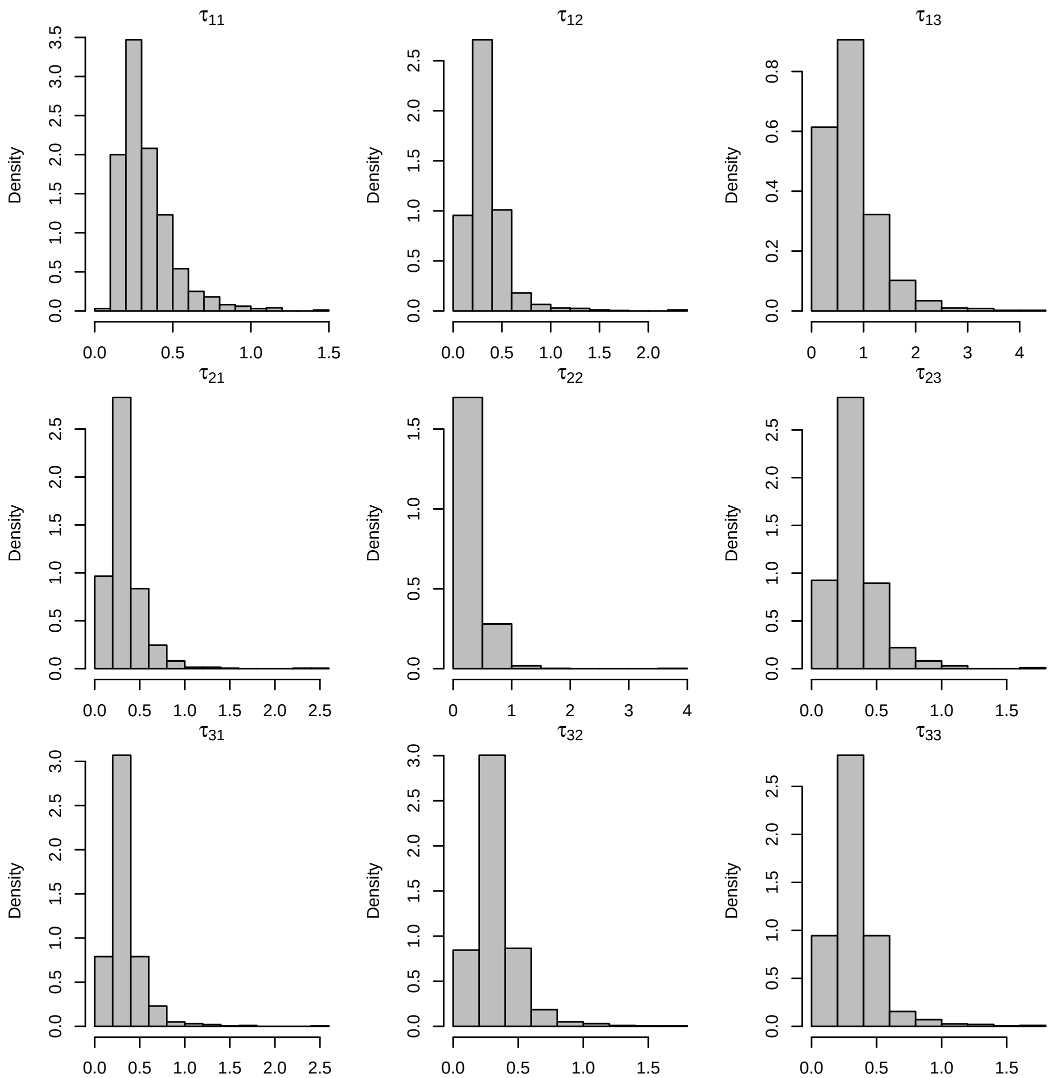

Figure 14: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

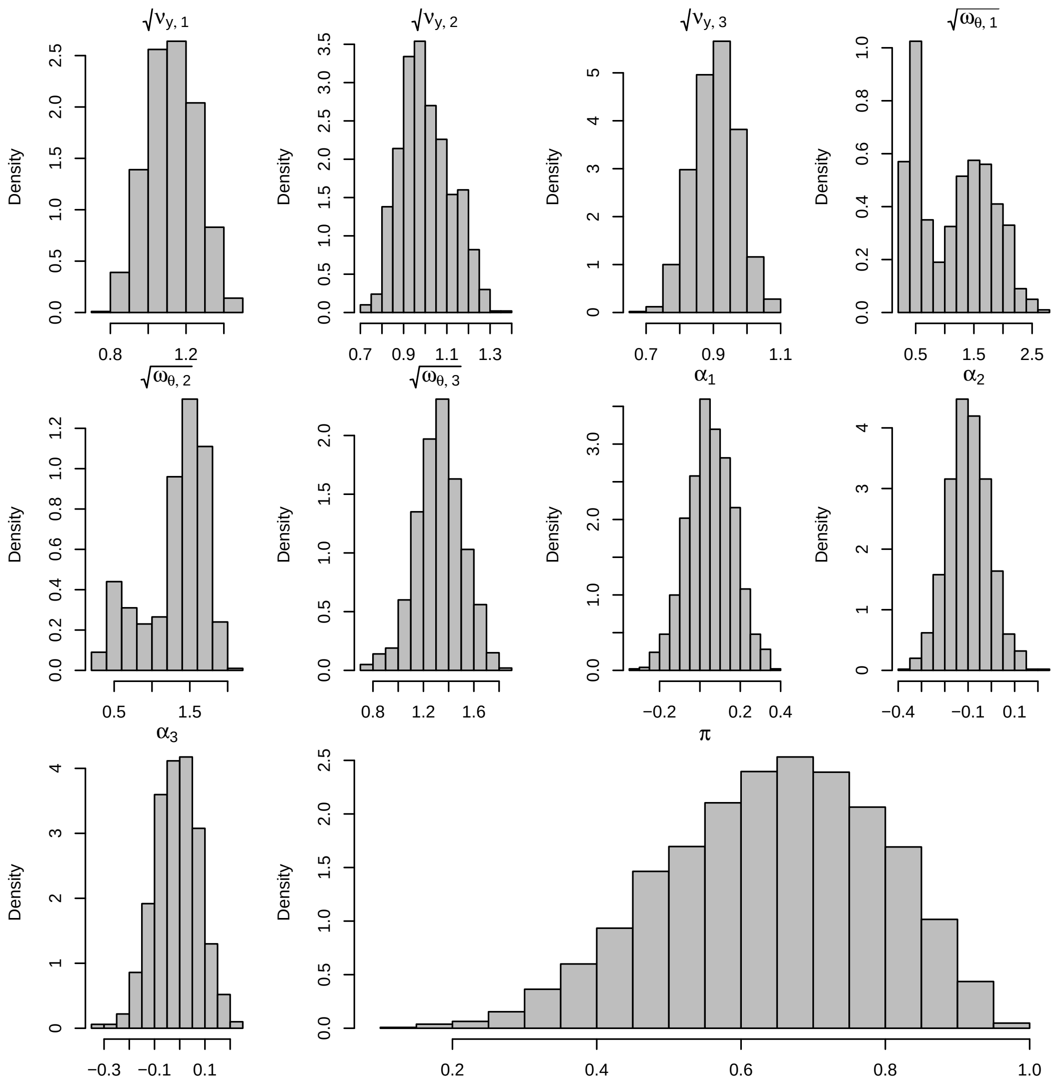

Figure 15: Posterior distribution: observational and state variances, trends and weights.

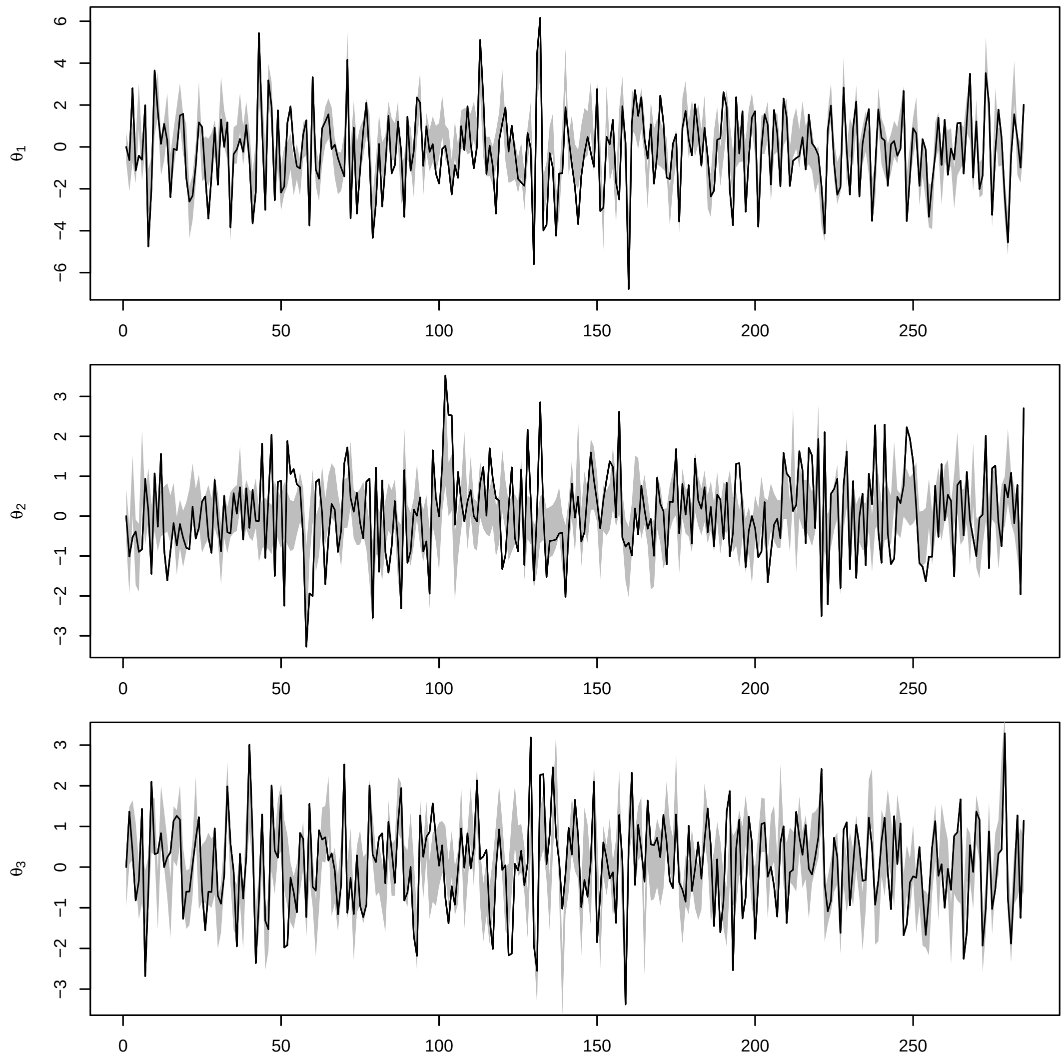

Figure 16: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 0.5$ and $\lambda_{\theta,i}$ known

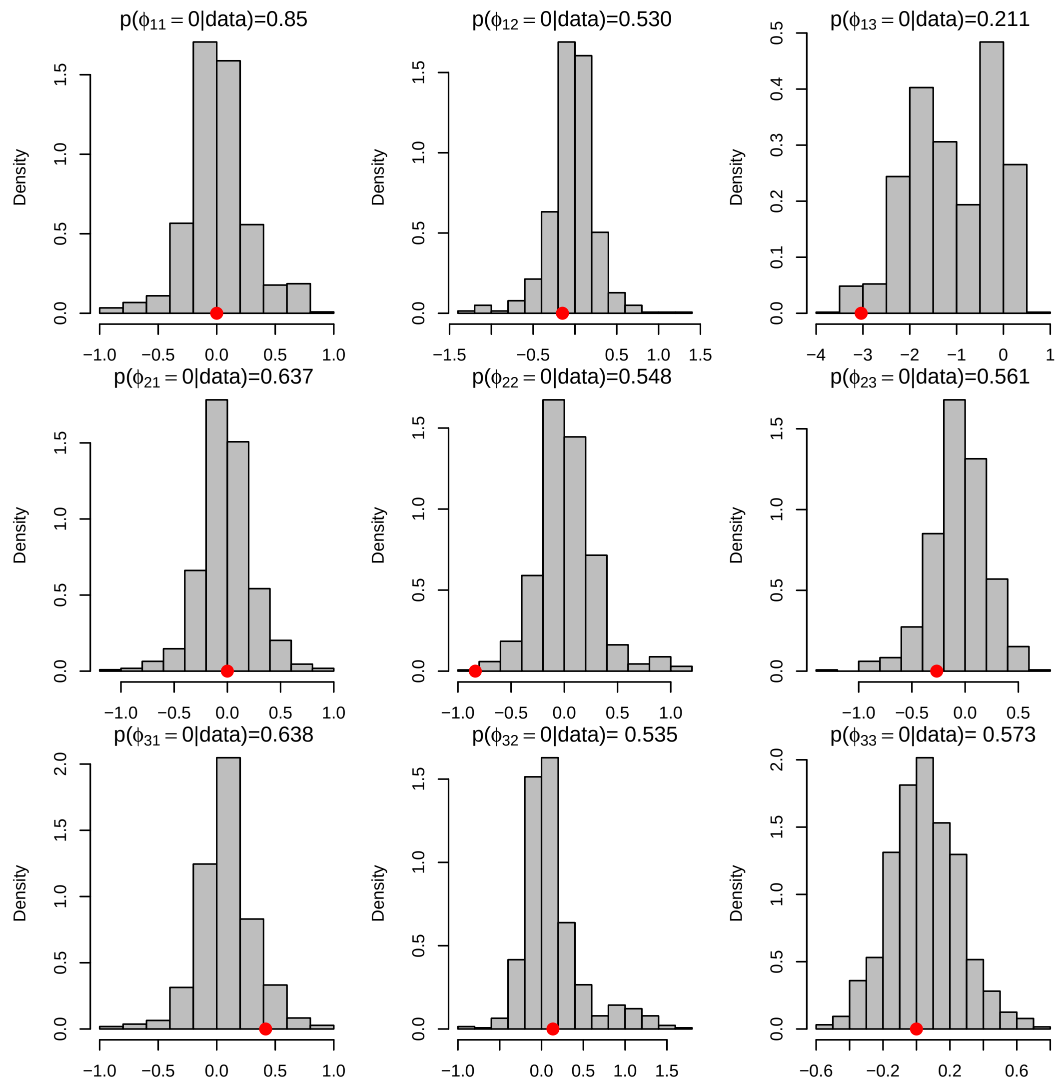

Figure 17: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

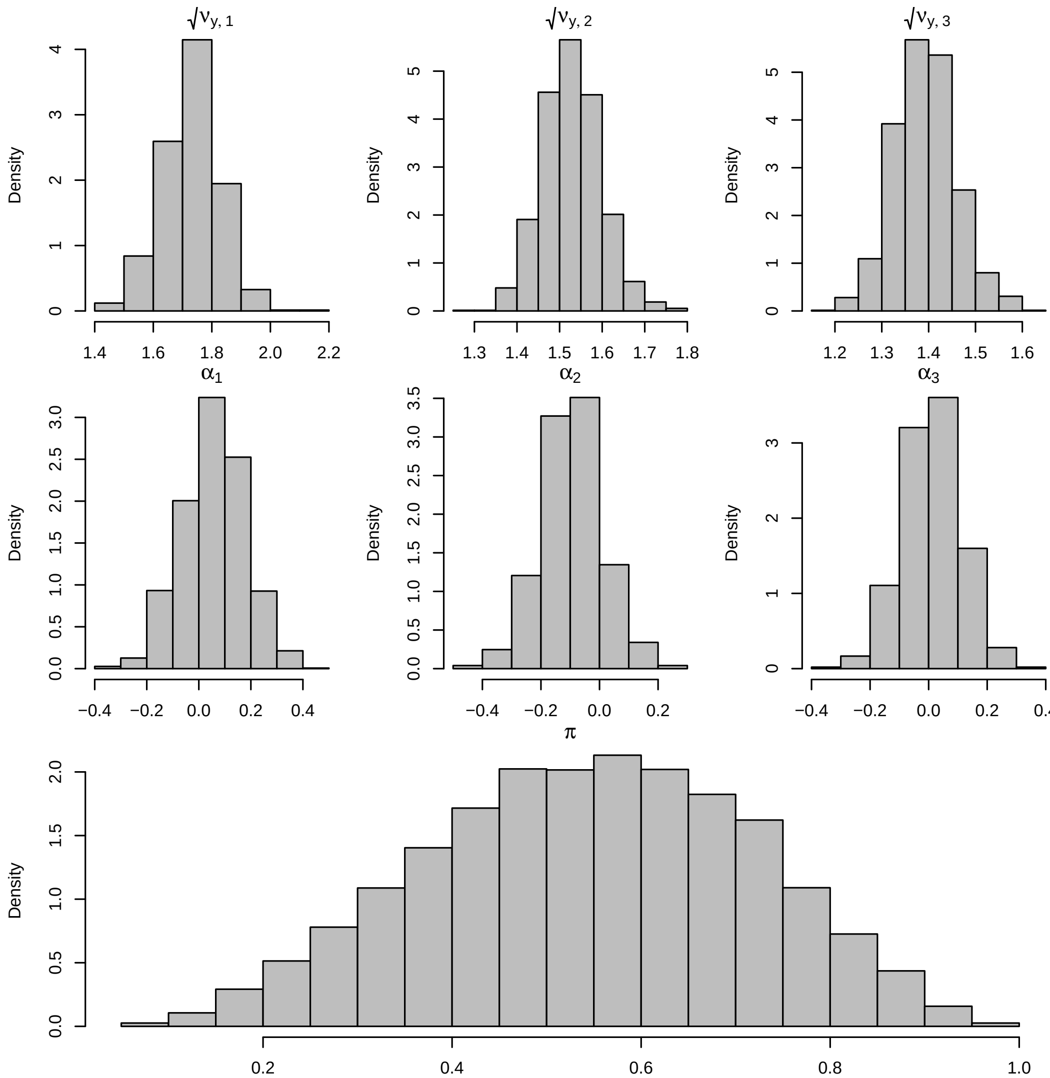

Figure 18: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

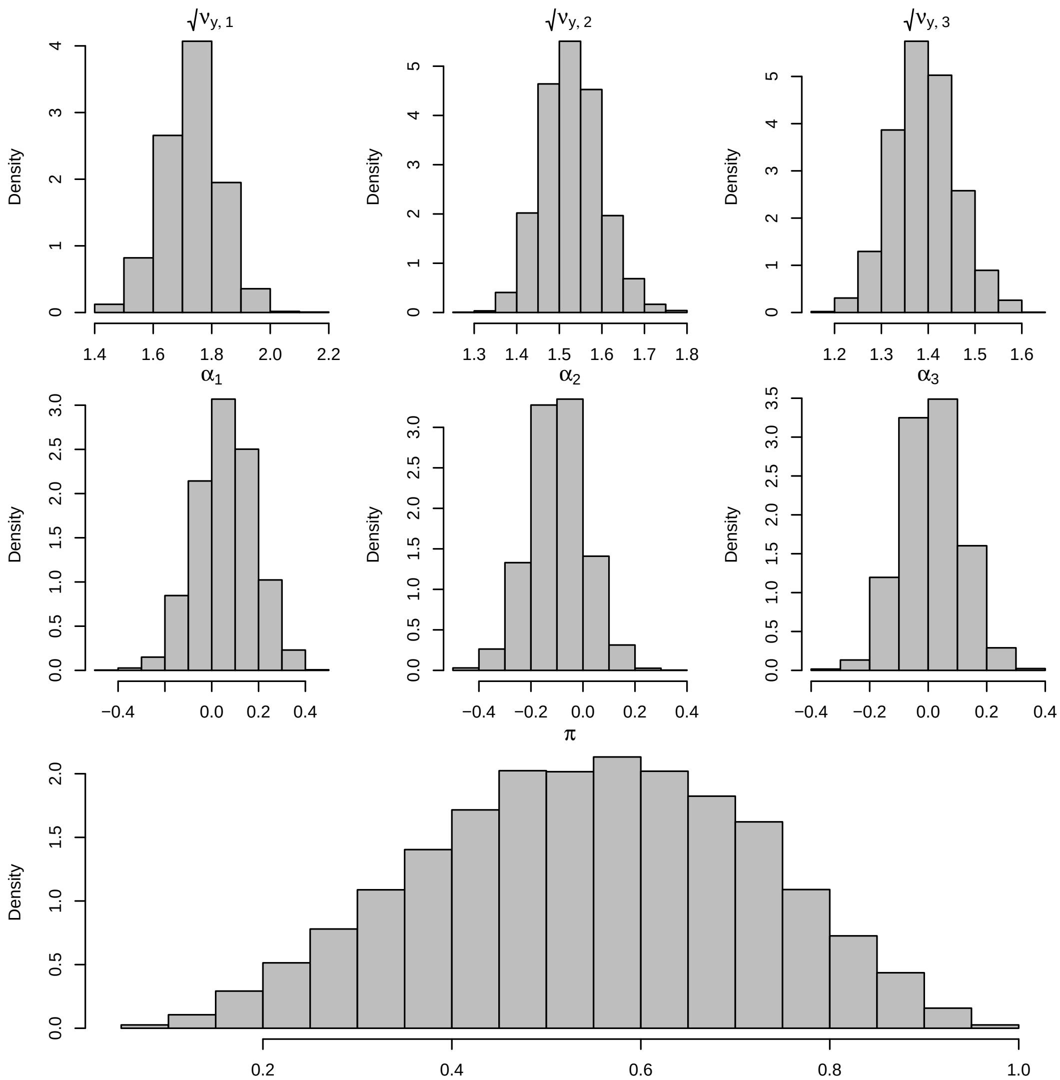

Figure 19: Posterior distribution: observational variances, trends and weights.

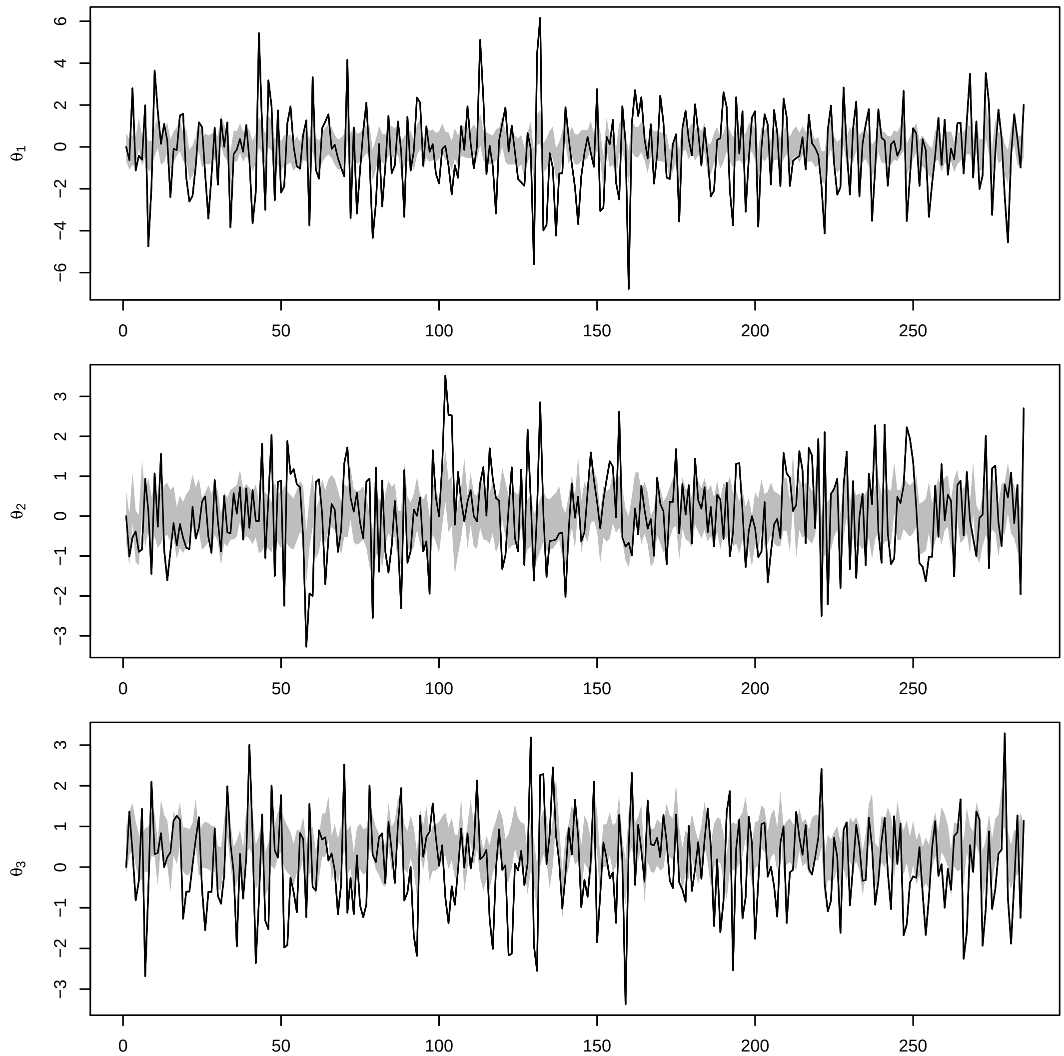

Figure 20: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

Case $\lambda_{\theta,i}^{-1}/\lambda_{y,i}^{-1} = 0.5$ and $\lambda_{\theta,i}$ unknown

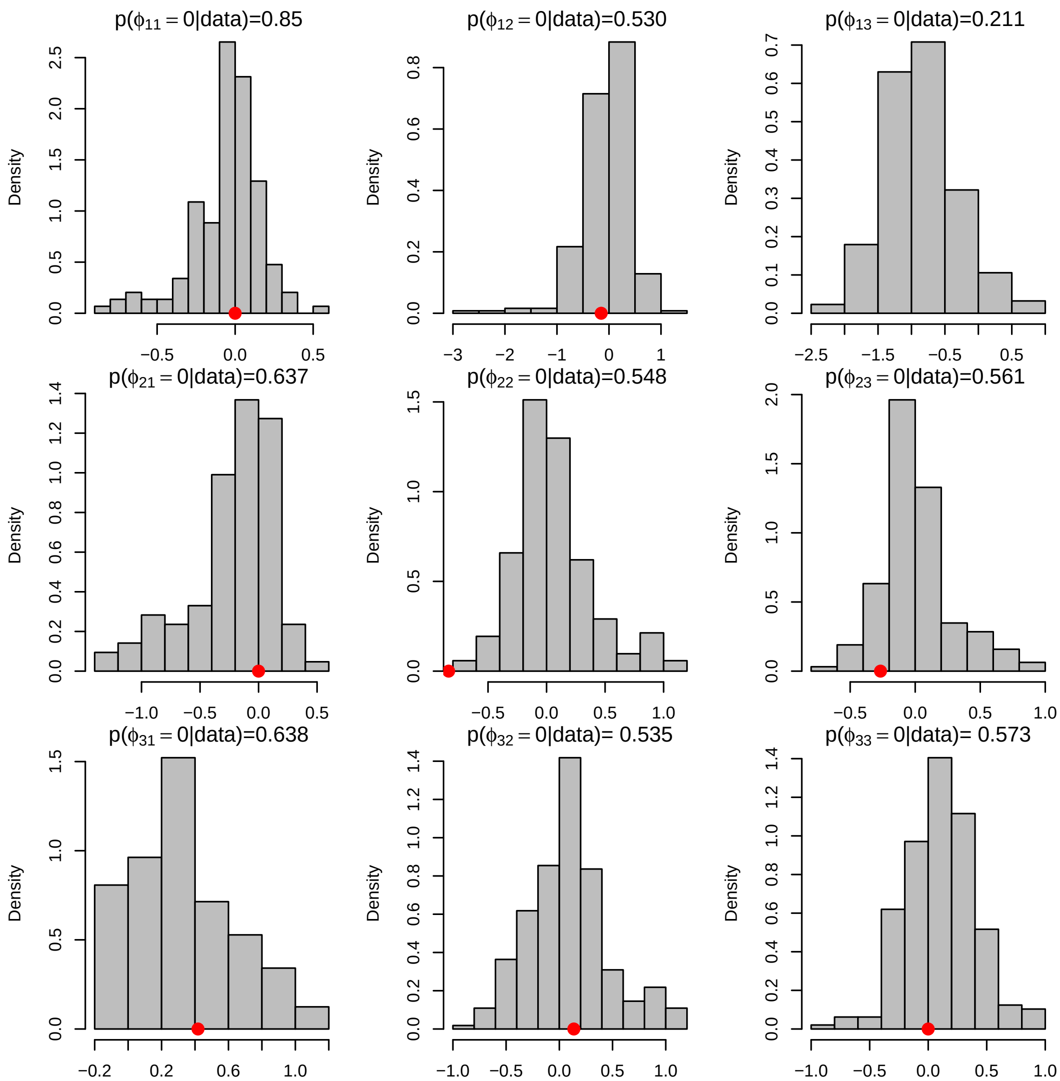

Figure 21: Posterior distributions of the mean of the connectivity regions ϕ_{ij}.

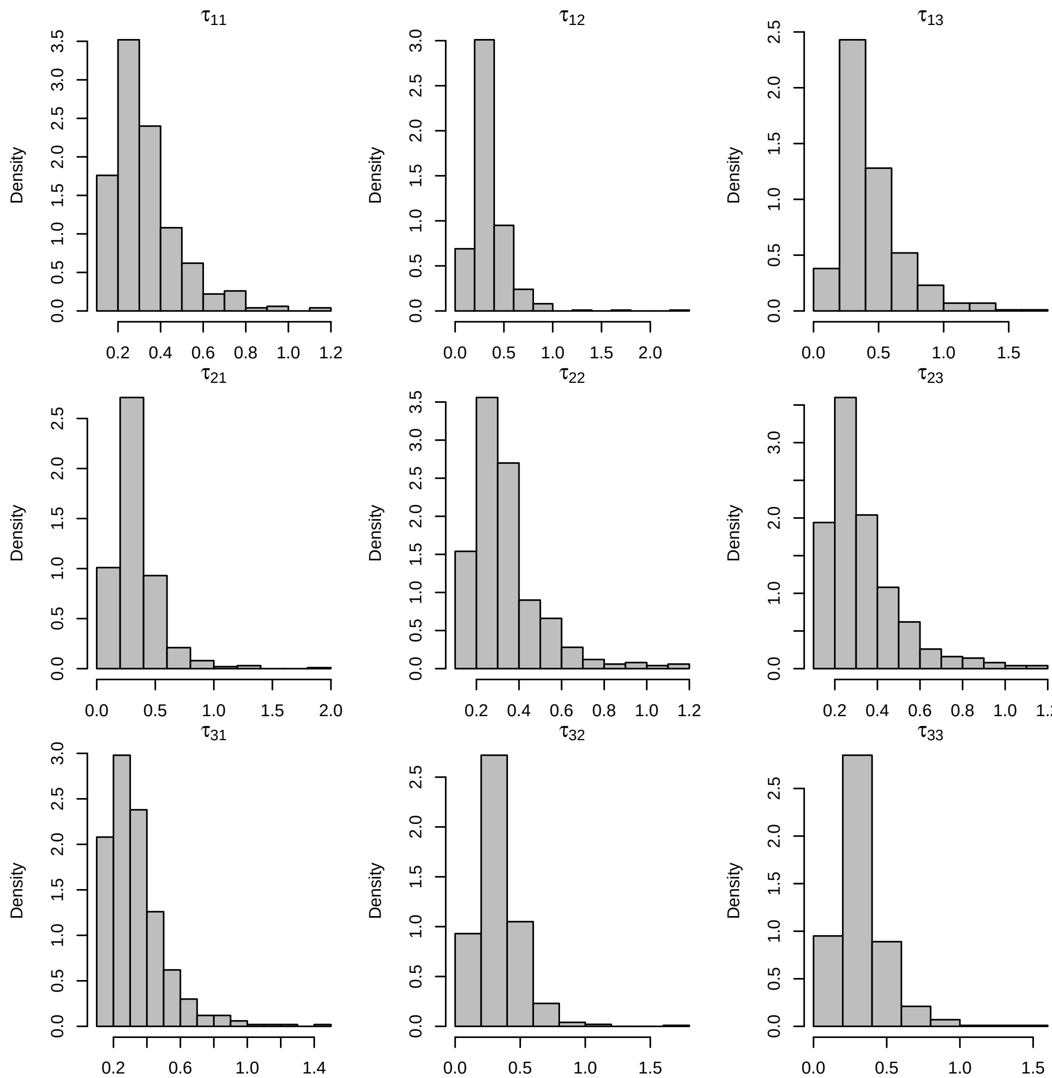

Figure 22: Posterior distributions of the precision of the connectivity regions ϕ_{ij}.

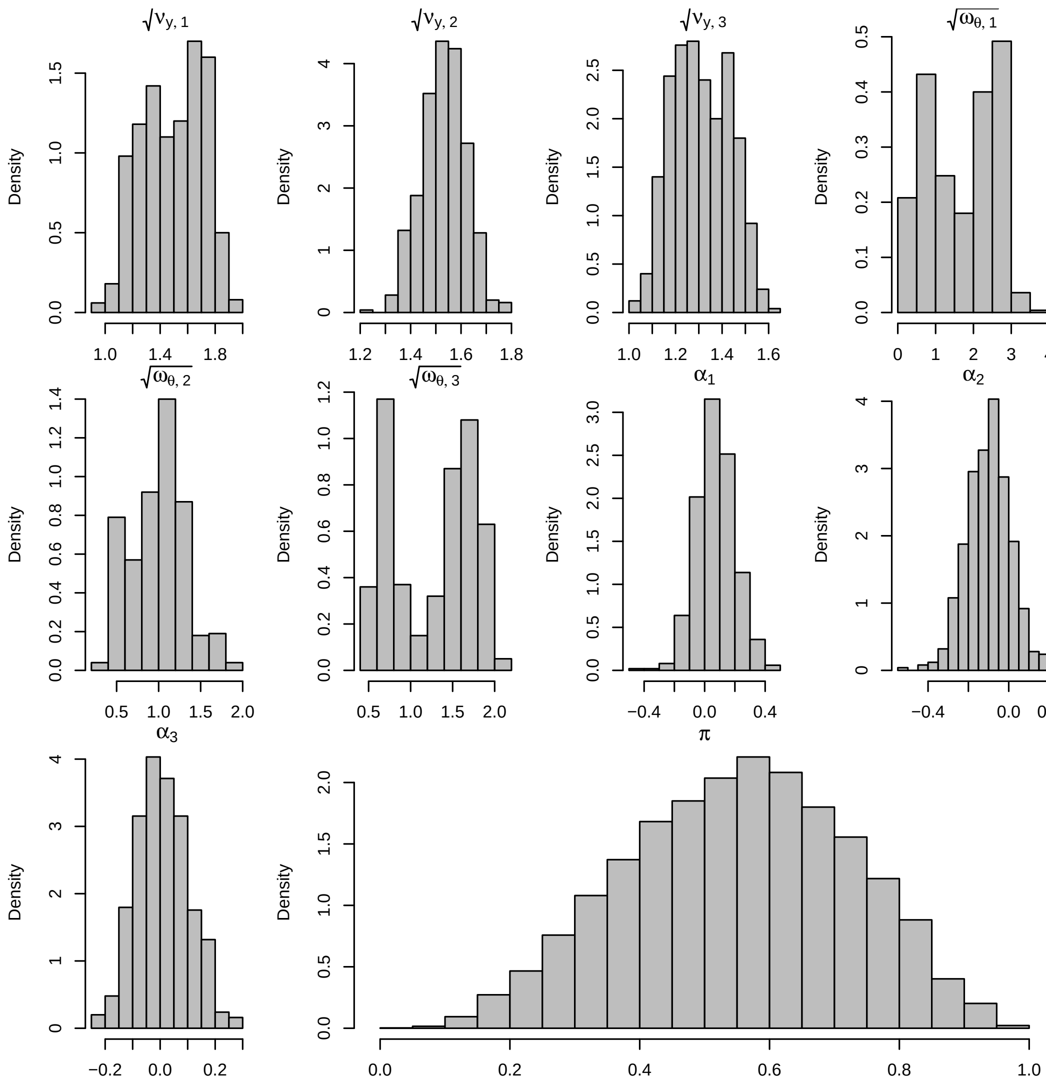

Figure 23: Posterior distribution: observational and state variances, trends and weights.

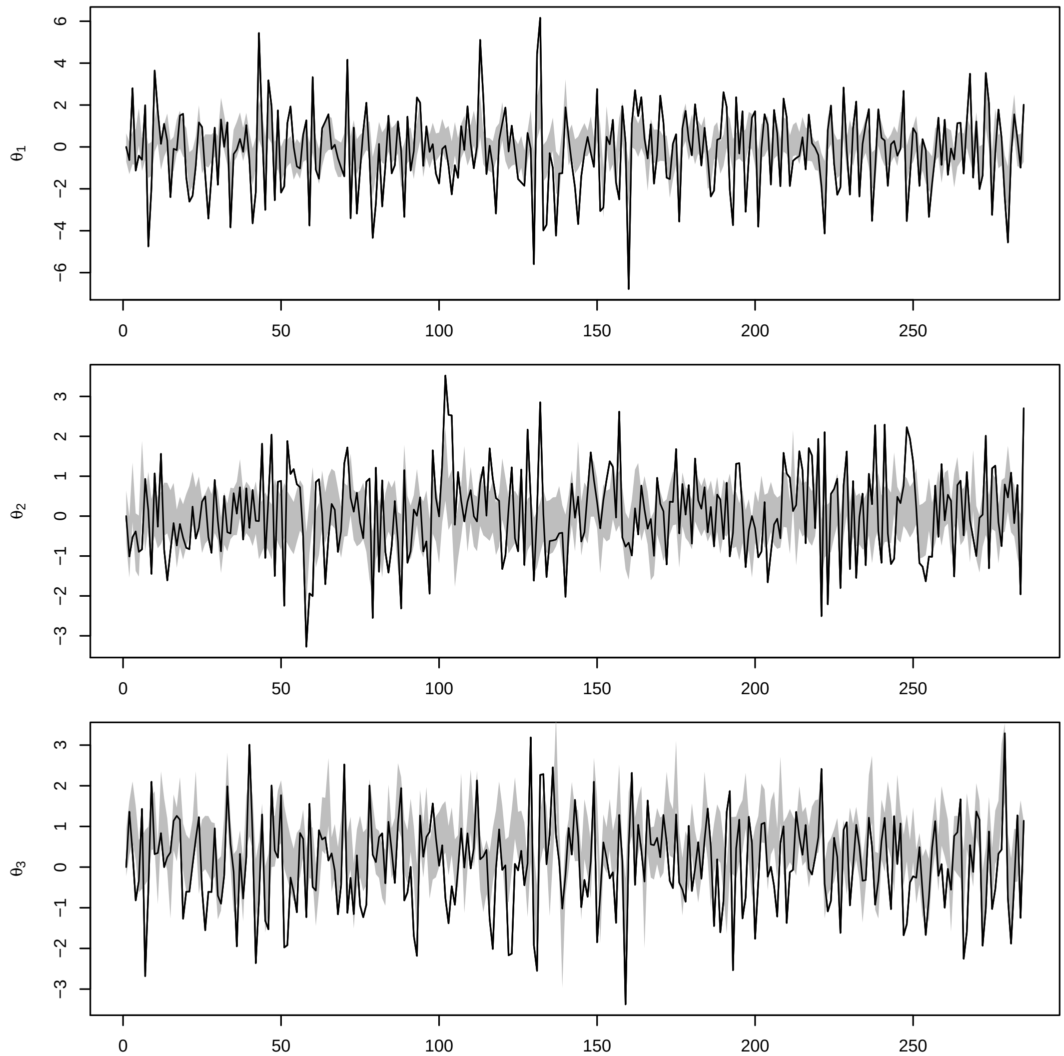

Figure 24: 95% posterior credible interval (shadow) of the state parameters using the FFBS algorithm and true simulated state parameters (black line).

References

Abramowitz, M. & Stegun, I. (1970), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.

Fruwirth-Schnatter, S. (1994), 'Data augmentation and dynamic linear models', *Journal of Time Series Analysis* **15**, 183–202.