

**BAYESIAN CALIBRATION OF
MULTISTATE STOCHASTIC SIMULATORS**

Department of Statistics, The Ohio State University

Supplementary Material

S1 Selecting n_c via Cross-Validation

We employ a leave-one-out cross-validation approach for selecting n_c . For an ensemble of m simulator outputs, remove the j th output from the ensemble and take the observation \mathbf{y} to be the mean (over the N stochastic simulator samples) of the j th output. Next, run the calibration model with the remaining $m - 1$ simulator outputs to predict the mean of the held-out (j th) output and the corresponding calibration parameters settings of the j th output. Repeat for $j = 1, \dots, m$, and then calculate appropriate criteria of interest. Repeat this entire process for a judicious range of n_c values, and then compare the criteria to select the number of bases to use in the approximation for calibrating the real data.

Two simple criteria we use to perform this cross-validation is the Mean Squared Prediction Error (MSPE) for the held-out mean simulator, and the Mean Squared

Error (MSE) for the held-out calibration parameter setting using the posterior mean calibration parameter estimates from the cross-validation runs. That is,

$$\text{MPSE} = \sum_{j=1}^m \sum_{i=1}^n (y_i - \bar{\mathbf{x}}_i(\boldsymbol{\theta}_j))^2$$

where $\bar{\mathbf{x}}_i(\boldsymbol{\theta}_j)$ is the posterior mean state(s) from running the calibration model at the j th step of the cross-validation, and

$$\text{MSE}(\boldsymbol{\theta}) = \sum_{j=1}^m \sum_{l=1}^q (\bar{\theta}_{jl} - \theta_{jl})^2$$

where $\bar{\theta}_{jl}$ is the posterior mean of the l th calibration parameter from running the calibration model at the j th step of the cross-validation (in actuality we re-scale these so that the squared errors are comparable for the q different calibration parameters).

S2 Specifying the Additional Prior Distributions

Prior on weight-space precision, λ_{v_l}

We specify independent gamma priors on the inverse variance of the GP model for each latent weight space,

$$\lambda_{v_l} \sim \text{Gamma}(\alpha_{v_l}, \beta_{v_l}),$$

by choosing a shape, α_{v_l} , and rate, β_{v_l} . Usually a shape $\alpha_{v_l} \geq 1$ is chosen and then

the rate is selected so that the mean of the prior, $\frac{\alpha_{v_l}}{\beta_{v_l}}$, is on the order of the empirical state variability.

Prior on weight-space correlations, $\rho_{l,t}$

The correlations are specified independent beta prior distributions,

$$\rho_{l,t} \sim \text{Beta}(\alpha_{\rho_{l,t}}, \beta_{\rho_{l,t}}),$$

where $\alpha_{\rho_{l,t}}$ and $\beta_{\rho_{l,t}}$ are usually chosen to favour a smooth response, which places more weight towards a correlation of 1. Our default choice for this prior, which generally works well, is $\alpha_{\rho_{l,t}} = 5, \beta_{\rho_{l,t}} = 1$.

Prior on observation precision, λ_f

The prior for λ_f is

$$\lambda_f \sim \text{Gamma}(\alpha_f, \beta_f),$$

where the shape parameter is again usually selected as $\alpha_f \geq 1$. If prior information on the observational error is known, this can be used to calibrate the prior. Otherwise, selecting β_f so that the inverse of the mean, $\left(\frac{\alpha_f}{\beta_f}\right)^{-1}$, is on the order of the expected observational error variance is reasonable. In some cases, we have observed that calibration can be sensitive to this parameter, so a careful consideration of the

interplay between additive discrepancy, multiplicative discrepancy and observational error variance may be warranted.

Prior on discrepancy correlations, $\psi_{k,t}$

The correlations are specified independent beta prior distributions,

$$\psi_{k,t} \sim \text{Beta}(\alpha_{\psi_{k,t}}, \beta_{\psi_{k,t}}),$$

where $\alpha_{\psi_{k,t}}$ and $\beta_{\psi_{k,t}}$ are usually chosen to favour a smooth response, but also recognizing that the discrepancy is often modeling smaller-scale variability in the unobserved state unaccounted for by the emulated simulator. Our default choice for this prior is $\alpha_{\psi_{k,t}} = 2, \beta_{\psi_{k,t}} = 10$.

S3 MCMC Algorithm

The MCMC algorithm for sampling the posterior distribution (3.3) proceeds according to the following steps:

1. Draw $\rho_{l,t}|\cdot$ for $l = 1, \dots, n_c$ and $t = 1, \dots, q$ (MH step)
2. Draw $\lambda_{v_l}|\cdot$ for $l = 1, \dots, n_c$ (Gibbs step)
3. Draw $\theta_t, v_1(\theta_t), \dots, v_{n_c}(\theta_t)$ by proposing a new θ'_t and
 - (a) Draw $v'_l(\theta'_t, \boldsymbol{\theta}_{-t})$ from $v_l(\theta'_t, \boldsymbol{\theta}_{-t})|\mathbf{V}_l, \theta'_t, \boldsymbol{\theta}_{-t}$ for $l = 1, \dots, n_c$ (Gibbs step)

(b) Calculate the acceptance probability

$$\alpha = \frac{\pi(\mathbf{y}|\mathbf{U}_u, \mathbf{v}'(\theta'_t, \boldsymbol{\theta}_{-t}), \boldsymbol{\mu}_\delta, \boldsymbol{\lambda}_f, \boldsymbol{\lambda}_\delta, \boldsymbol{\psi}, \boldsymbol{\mu}_\kappa, \boldsymbol{\lambda}_\kappa)\pi(\theta'_t)}{\pi(\mathbf{y}|\mathbf{U}_u, \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t}), \boldsymbol{\mu}_\delta, \boldsymbol{\lambda}_f, \boldsymbol{\lambda}_\delta, \boldsymbol{\psi}, \boldsymbol{\mu}_\kappa, \boldsymbol{\lambda}_\kappa)\pi(\theta_t)}$$

where $\pi(\mathbf{y}|\mathbf{U}_u, \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t}), \boldsymbol{\mu}_\delta, \boldsymbol{\lambda}_f, \boldsymbol{\lambda}_\delta, \boldsymbol{\psi}, \boldsymbol{\mu}_\kappa, \boldsymbol{\lambda}_\kappa) = \int_{\boldsymbol{\kappa}} \int_{\boldsymbol{\delta}} \pi(\mathbf{y}|\mathbf{U}_u, \mathbf{v}(\boldsymbol{\theta}), \boldsymbol{\delta}, \boldsymbol{\kappa})d\pi(\boldsymbol{\delta})d\pi(\boldsymbol{\kappa})$

(c) Accept $\theta'_t, v_1(\theta'_t, \boldsymbol{\theta}_{-t}), \dots, v_{n_c}(\theta'_t, \boldsymbol{\theta}_{-t})$ with probability α .

Repeat 3(a)-3(c) for $t = 1, \dots, q$ (MH steps).

4. Draw $\boldsymbol{\delta}_k|\mathbf{U}_{u,k}, \mathbf{v}(\boldsymbol{\theta}), \mathbf{y}_k, \lambda_{f,k}, \lambda_{\delta_k}, \boldsymbol{\psi}_k$ for $k = 1, \dots, n_s$ (Gibbs step)
5. Draw $\psi_{k,t}|\cdot$ for $t = 1, \dots, p$ and $k = 1, \dots, n_s$ (MH step)
6. Draw $\lambda_{\delta_k}|\cdot$ for $k = 1, \dots, n_s$ (Gibbs step)
7. Draw $\kappa_k|\cdot$ for $k = 1, \dots, n_s$ (Gibbs step)
8. Draw $\lambda_{f,k}|\cdot$ for $k = 1, \dots, n_s$ (Gibbs step).

The MCMC algorithm's steps can be implemented as follows.

In step 1,

- Draw a proposed $\rho'_{l,t}$ from $q(\rho'_{l,t}|\rho_{l,t})$
- Calculate $\alpha = \frac{\pi(\mathbf{V}_l|\rho'_{l,t}, \cdot)\pi(\rho'_{l,t})q(\rho_{l,t}|\rho'_{l,t})}{\pi(\mathbf{V}_l|\rho_{l,t}, \cdot)\pi(\rho_{l,t})q(\rho'_{l,t}|\rho_{l,t})}$
- Accept $\rho'_{l,t}$ with probability α .

In step 2, draw λ_{v_l} from $\text{Gamma}(\alpha_{v_l} + \frac{m}{2}, \beta_{v_l} + \frac{1}{2}\mathbf{V}^T\mathbf{R}_{v_l}^{-1}\mathbf{V})$.

In step 3,

$$\mathbf{y} | \mathbf{U}_{u,k}, \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t}), \mu_{\delta_k}, \lambda_{f,k}, \lambda_{\delta_k}, \psi_k, \mu_{\kappa_k}, \lambda_{\kappa_k} \sim N(\mu_{\delta_k} + \mu_{\kappa_k} \mathbf{U}_{u,k} \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t}), \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \frac{1}{\lambda_{f,k}} \mathbf{I}_n + \frac{1}{\lambda_{\delta_k}} \mathbf{R}_{\delta_k}(\psi_k) + \frac{1}{\lambda_{\kappa_k}} (\mathbf{U}_{u,k} \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t})) (\mathbf{U}_{u,k} \mathbf{v}(\theta_t, \boldsymbol{\theta}_{-t}))^T$ for $k = 1, \dots, n_s$.

In step 4,

$$\boldsymbol{\delta}_k | \mathbf{U}_{u,k}, \mathbf{v}(\boldsymbol{\theta}), \mathbf{y}_k, \lambda_{f,k}, \lambda_{\delta_k}, \psi_k \sim N(\boldsymbol{\Sigma}_{\delta_k} (\lambda_{f,k} \mathbf{I}_n (\mathbf{y}_k - \kappa_k \mathbf{U}_{u,k} \mathbf{v}(\boldsymbol{\theta})) + \lambda_{\delta_k} \mathbf{I}_n \mathbf{R}_{\delta_k}(\psi_k)^{-1} \boldsymbol{\mu}_{\delta_k}), \boldsymbol{\Sigma}_{\delta_k})$$

where $\boldsymbol{\Sigma}_{\delta_k}^{-1} = \lambda_{f,k} \mathbf{I}_n + \lambda_{\delta_k} \mathbf{R}_{\delta_k}(\psi_k)^{-1}$ for $k = 1, \dots, n_s$.

In step 5,

- Draw a proposed $\psi'_{k,t}$ from $q(\psi'_{k,t} | \psi_{k,t})$
- Calculate $\alpha = \frac{\pi(\boldsymbol{\delta}_k | \psi'_{k,t}, \cdot) \pi(\psi'_{k,t}) q(\psi_{k,t} | \psi'_{k,t})}{\pi(\boldsymbol{\delta}_k | \psi_{k,t}, \cdot) \pi(\psi_{k,t}) q(\psi'_{k,t} | \psi_{k,t})}$
- Accept $\psi'_{k,t}$ with probability α .

In step 6, draw λ_{δ_k} from

$$\text{Gamma}\left(\alpha_{\delta_k} + \frac{n}{2}, \beta_{\delta_k} + \frac{1}{2} (\boldsymbol{\delta}_k - \boldsymbol{\mu}_{\delta_k})^T \mathbf{R}_{\delta_k}^{-1} (\boldsymbol{\delta}_k - \boldsymbol{\mu}_{\delta_k})\right)$$

for $k = 1, \dots, n_s$.

In step 7, draw κ_k from

$$N(\sigma_{\kappa_k} (\lambda_{f,k} (\mathbf{y}_k - \boldsymbol{\delta}_k))^T (\mathbf{U}_{u,k} \mathbf{v}(\boldsymbol{\theta})) + \lambda_{\kappa_k} \mu_{\kappa_k}, \sigma_{\kappa_k})$$

where $\sigma_{\kappa_k}^{-1} = \lambda_{f,k}(\mathbf{U}_{u,k}\mathbf{v}(\boldsymbol{\theta}))^T(\mathbf{U}_{u,k}\mathbf{v}(\boldsymbol{\theta})) + \lambda_{\kappa_k}$ for $k = 1, \dots, n_s$.

In step 8, draw $\lambda_{f,k}$ from

$$\text{Gamma}\left(\alpha_{f,k} + \frac{n}{2}, \beta_{f,k} + \frac{1}{2}(\mathbf{y}_k - \kappa_k \mathbf{U}_{u,k}\mathbf{v}(\boldsymbol{\theta}) - \boldsymbol{\delta}_k)^T(\mathbf{y}_k - \kappa_k \mathbf{U}_{u,k}\mathbf{v}(\boldsymbol{\theta}) - \boldsymbol{\delta}_k)\right)$$

for $k = 1, \dots, n_s$.

S4 Additional Figures for the JAK-STAT Example

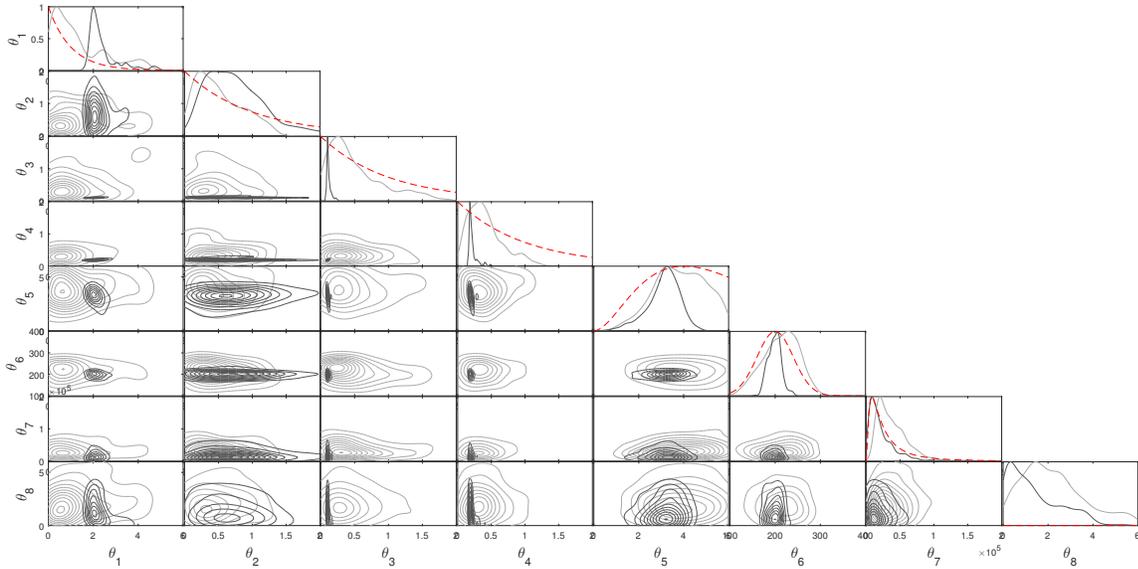


Figure 5: Kernel density estimates of the marginal calibrated posterior (gray) with $m = 50$ model runs, and exact posterior (black, Chkrebtii et al. (2016)) for the JAK-STAT system. Marginal prior densities are shown as dotted lines.

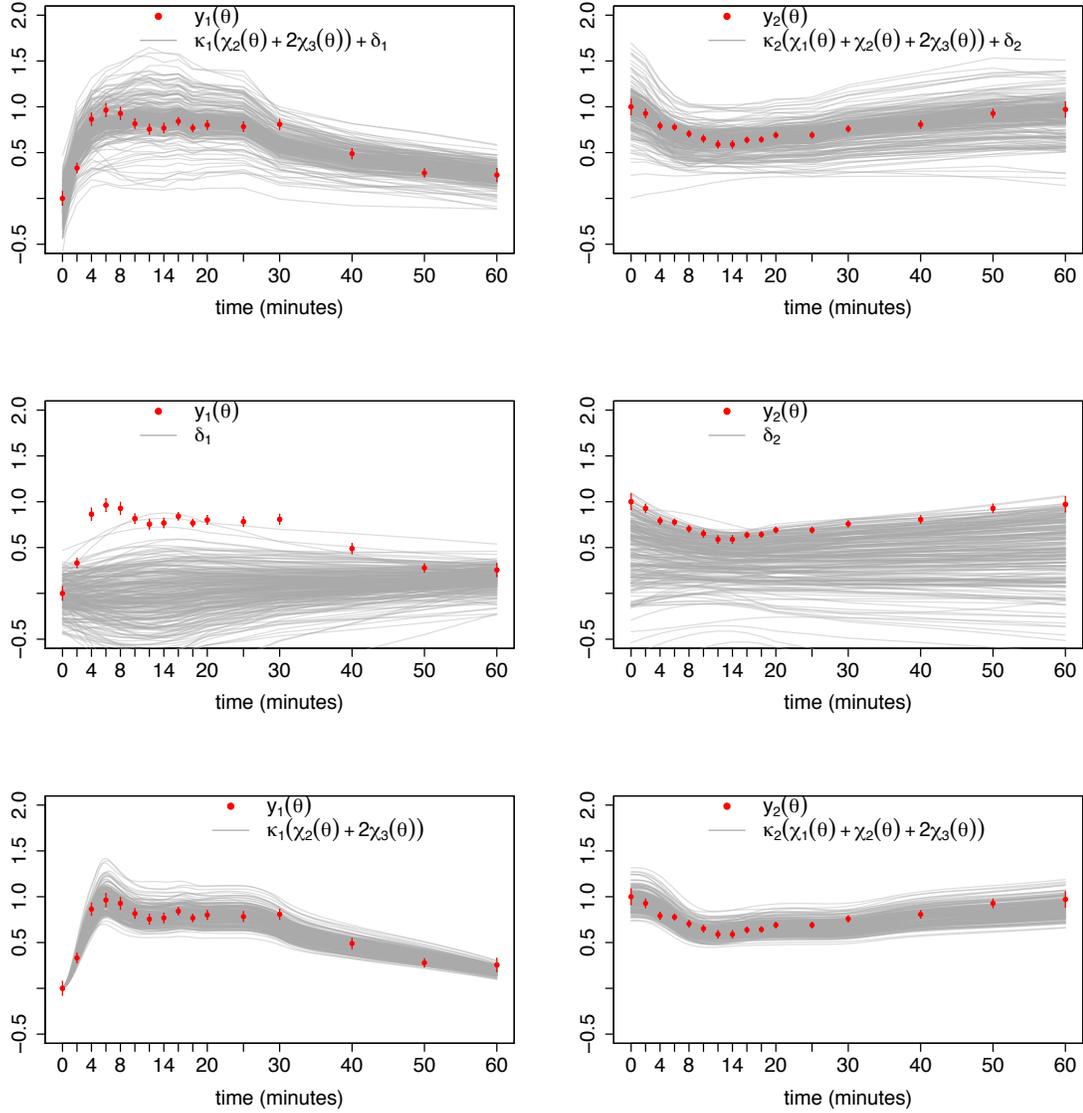


Figure 6: 200 samples from the marginal calibrated model posterior with $m = 50$ model runs (top row), discrepancies δ_1 and δ_2 (middle row), and exact posterior for comparison (bottom row, Chkrebtii et al. (2016)) over the first two observation processes of the JAK-STAT system, for which experimental data is available. Experimental data locations are shown as red circles with error bars representing twice the experimental error standard deviation.

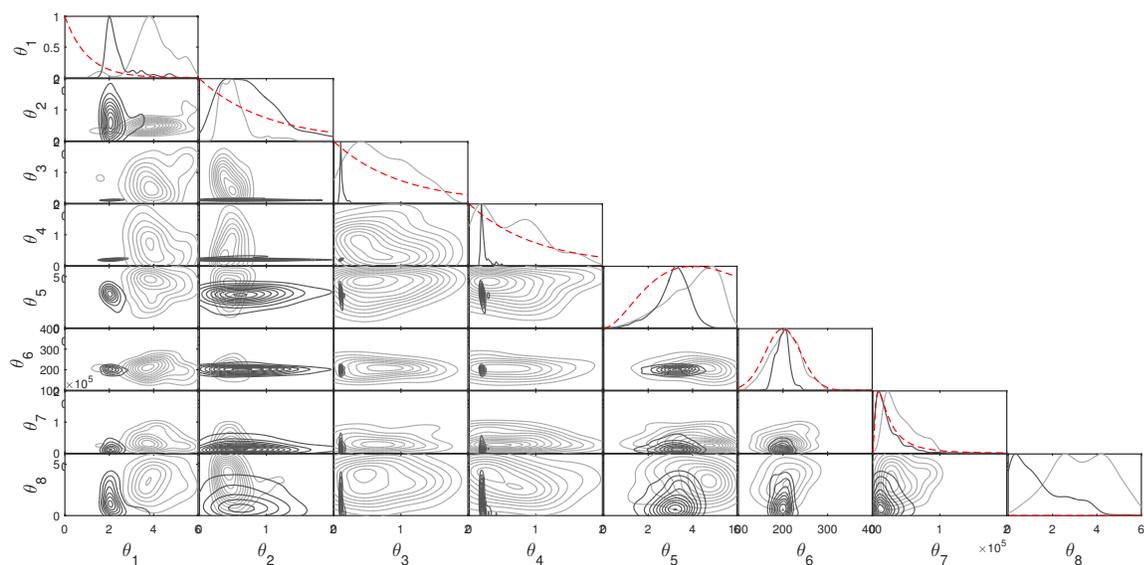


Figure 7: Kernel density estimates of the marginal stochastically calibrated posterior (gray) with $m = 20$ model runs, and exact posterior (black, Chkrebtii et al. (2016)) for the JAK-STAT system. Marginal prior densities are shown as dotted lines.

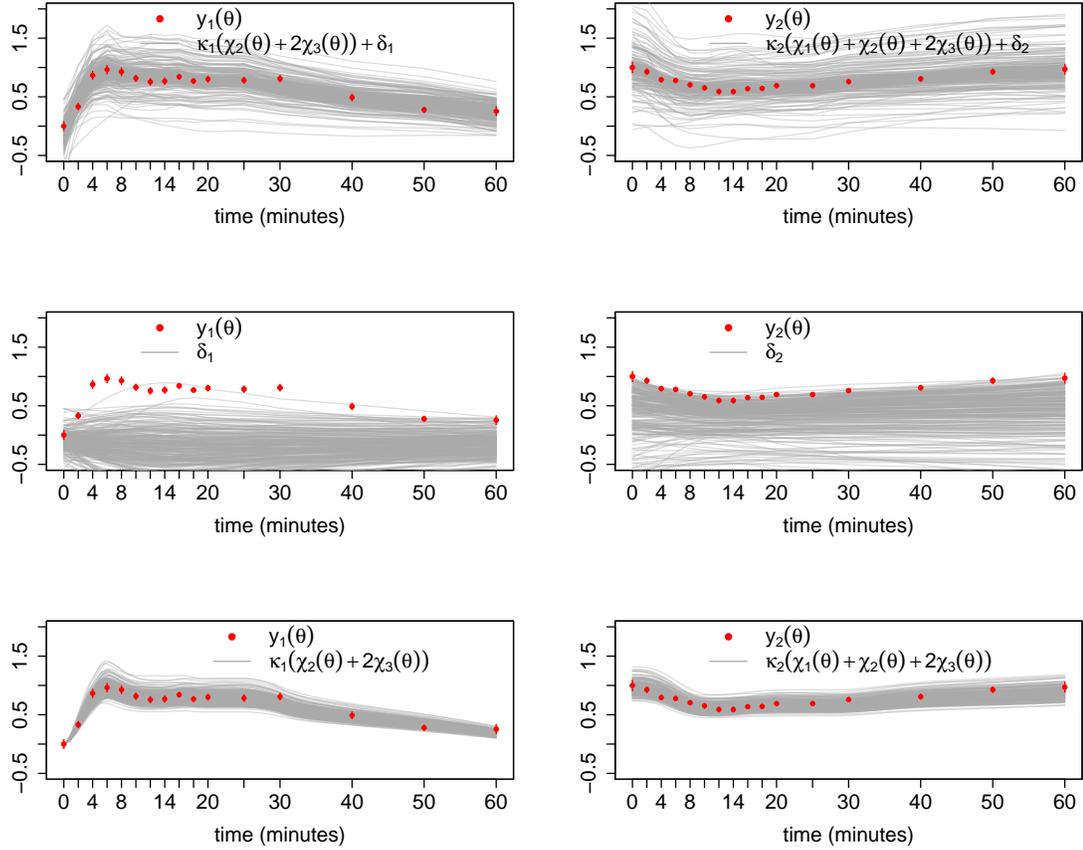


Figure 8: 200 samples from the marginal calibrated model posterior with $m = 20$ model runs (top row), discrepancies δ_1 and δ_2 (middle row), and exact posterior for comparison (bottom row, Chkrebtii et al. (2016)) over the first two observation processes of the JAK-STAT system, for which experimental data is available. Experimental data locations are shown as red circles with error bars representing twice the experimental error standard deviation.

References

- Chkrebtii, O., D. A. Campbell, B. Calderhead, and M. Girolami (2016). Bayesian solution uncertainty quantification for differential equations. *Bayesian Analysis* 11, 1239–1267.