Calibration and Validation of Computer Models: a Bayesian Approach
Lecture 3: Computer models: calibration and validation

Rui Paulo

ISEG/CEMAPRE Technical University of Lisboa, Portugal

June 30 2011

- Output of computer model for input vector $\mathbf{z}$ is $y^M(\mathbf{z})$
- $\mathbf{z} = (\mathbf{x}, \mathbf{u})$ where

  $\mathbf{x}$ are *controllable* inputs

  $\mathbf{u}$ are *uncertain* inputs or parameters, which can be *tuning* or *calibration* parameters
- Computer model aims at reproducing some real phenomenon which we denote by $y^R(\mathbf{x})$

Validation of computer models

- Question of interest: *Does the computer model adequately represent reality?*
- The answer to the question "Is the model correct?" is almost always "No"
- In general, people are interested in whether the model produces results that are accurate enough for the intended use

Tolerance bounds

- We respond to that question by producing statements like

$$P\{|\text{reality} - \text{model}| < \tau\} > \gamma$$

  for some tolerance τ and probability γ
- Example: $5.76 \pm .44$ — there's a specified chance (say 90%) that the true underlying process at certain input value lies within this specified range

Difficulties

- Uncertainty in $\mathbf{u}$ (or $\mathbf{x}$) arising from multiple sources
- Limited model runs
- Field data limited and/or noisy
- Model runs and field data at different $\mathbf{x}$
- Simultaneously "tune" $\mathbf{u}$ and validate model, based on the same set of field data
- y^M highly non-linear and biased
- Validation as dynamic process

The Framework

- Specify model inputs and parameters with associated uncertainties
- Determine evaluation criteria
- Data collection and design of experiments

$\Rightarrow$ Construct fast approximation to the computer model

$\Rightarrow$ Analysis of model output; comparing model output with field data

- Feedback and feed-forward

- A computer model is a computer implementation of a scientific model which purports to represent a real physical system
- A statistical analysis of this problem should take into account the potential mismatch between the real physical system and the computer model
- Goldstein (2010) labels this additional source of uncertainty *external* uncertainty
- Scientific model:

$$(\mathbf{x}, \mathbf{u}) \mapsto y^M(\mathbf{x}, \mathbf{u})$$

Reasons for the potential mismatch:

- $(\mathbf{x}, \mathbf{u})$ describe the process we are observing only approximately
- Given $(\mathbf{x}, \mathbf{u})$, the mathematical formulae involved in $y^M(\mathbf{x}, \mathbf{u})$ are an approximate description of the rules that lead to the properties of the real process
  - formulae need to be tractable
  - scientific knowledge is limited

These simplifications do not invalidate the use of the model, on the contrary! We just need to acknowledge this fact and account for the additional uncertainty.

The Statistical Model

- Reality and computer model: $\mathbf{u}_\star$ true value of $\mathbf{u}$, (Kennedy and O'Hagan 2001)

$$y^R(\mathbf{x}) = y^M(\mathbf{x}, \mathbf{u}_\star) + b(\mathbf{x})$$

  $b(\mathbf{x})$ is the so-called bias function

- Field measurements:

$$y_j^F(\mathbf{x}_i) = y^R(\mathbf{x}_i) + \varepsilon_{ij}$$
$$\varepsilon_{ij} \overset{iid}{\sim} \mathrm{N}(0, 1/\lambda^F)$$

Separating all sources of uncertainty in practice is a very challenging task due to potential confounding

$$y_j^F(\mathbf{x}_i) = y^M(\mathbf{x}_i, \mathbf{u}_\star) + b(\mathbf{x}_i) + \varepsilon_{ij}$$

- $\mathbf{u}$ in the model may not even represent the physical quantity for which we have scientific knowledge
- b is non-observable and large variance for ε can replace external uncertainty; replicates are not always easy to come by
- the function $y^M(\cdot)$ is very complex; model uncertainty can replace b

Prior on b

- The bias function is modeled *a priori* as a GP independent of $y^M(\cdot)$ and ε
- *A priori*, with $\boldsymbol{\theta}^b = (\mu^b, \boldsymbol{\beta}^b, \lambda^b)$,

$$b(\cdot) \mid \boldsymbol{\theta}^b \sim \text{GP}(\, \mu^b, c^b(\cdot,\cdot)/\lambda^b \,)$$
$$c^b(\mathbf{x}, \mathbf{x}^\star) = \prod_{j=1}^{p-p_\star} \exp[-\beta_j^b (x_j - x_j^\star)^2] \; .$$

- Note that $\alpha_j = 2$ meaning that we expect $b(\cdot)$ to be smooth
- Typically we set $\mu^b = 0$
- Missing: priors on $\boldsymbol{\theta}^b$, λ^F and $\mathbf{u}$.

Sufficient statistics

- The field design set is $D^F = \{\mathbf{x}_i,\ i = 1, \ldots, \ell\}$; the number of replicates observed at $\mathbf{x}_i$ is n_i
- $y_k^F(\mathbf{x}_i) \mid b(\mathbf{x}_i), y^M(\mathbf{x}_i, \mathbf{u}_\star), \lambda^F \sim \mathrm{N}(y^M(\mathbf{x}_i, \mathbf{u}_\star) + b(\mathbf{x}_i), 1/\lambda^F)$ independent

The sufficient statistics are then

$$\bar{\mathbf{y}}^F = (\bar{y}^F(\mathbf{x}_i), i = 1, \ldots, \ell)$$

$$s_F^2 = \sum_{i=1}^{\ell} \sum_{k=1}^{n_i} (y_k^F(\mathbf{x}_i) - \bar{y}^F(\mathbf{x}_i))^2$$

Notation

- $\mathbf{y}^M$ is the model data, i.e., computer model evaluated at D^M
- $\mathbf{u}_\star$ "true" value of the calibration parameter
- $\mathbf{y}^M_\star$ is model evaluated at $(\mathbf{x}_i, \mathbf{u}_\star)$, that is, at the field design set augmented with $\mathbf{u}_\star$, $D^F_\star$
- $\mathbf{b} = (b(\mathbf{x}_i),\ i = 1, \ldots, \ell)$

So, with $\boldsymbol{\Sigma}^F = \mathbf{diag}\ \mathbf{n}^{-1}/\lambda^F$

$$\bar{\mathbf{y}}^F \mid \mathbf{y}^M_\star, \mathbf{b}, \lambda^F \sim \mathrm{N}(\mathbf{y}^M_\star + \mathbf{b}, \boldsymbol{\Sigma}^F)$$

$$s^2_F \mid \lambda^F \sim \tfrac{1}{\lambda^F}\chi^2(\sum_{i=1}^{\ell}(n_i - 1)) \ .$$

Also, it is clear that

$$
\begin{aligned}
\mathbf{b} \mid \boldsymbol{\theta}^b, \mu^b &\sim \mathrm{N}(\mathbf{1}\mu^b, c^b(D^F, D^F)/\lambda^b) \equiv \mathrm{N}(\boldsymbol{\mu}^b, \boldsymbol{\Sigma}^b) \\
\mathbf{y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M &\sim \mathrm{N}(\mathbf{X}\boldsymbol{\theta}^L, c^M(D^M, D^M)/\lambda^M) \equiv \mathrm{N}(\boldsymbol{\mu}^M, \boldsymbol{\Sigma}^M) \\
\mathbf{y}^M_\star \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \mathbf{u}_\star &\sim \mathrm{N}(\mathbf{X}_\star\boldsymbol{\theta}^L, c^M(D^F_\star, D^F_\star)/\lambda^M) \equiv \mathrm{N}(\boldsymbol{\mu}^M_\star, \boldsymbol{\Sigma}^M_\star) \\
\mathbf{y}^M_\star \mid \mathbf{y}^M, \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \mathbf{u}_\star &\sim \mathrm{N}(\boldsymbol{\mu}_{\star|\bullet}, \boldsymbol{\Sigma}_{\star|\bullet}) \ .
\end{aligned}
$$

Above,

$$
\begin{aligned}
\boldsymbol{\mu}_{\star|\bullet} &= \boldsymbol{\mu}^M_\star + \boldsymbol{\Sigma}_{\star\bullet}\,[\boldsymbol{\Sigma}^M]^{-1}\,(\mathbf{y}^M - \boldsymbol{\mu}^M) \\
\boldsymbol{\Sigma}_{\star|\bullet} &= \boldsymbol{\Sigma}^M_\star \ - \ \boldsymbol{\Sigma}_{\star\bullet}\,[\boldsymbol{\Sigma}^M]^{-1}\,\boldsymbol{\Sigma}_{\bullet\star}
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\star\bullet} &= c^M(D^F_\star, D^M)/\lambda^M \\
\boldsymbol{\Sigma}_{\bullet\star} &= [\boldsymbol{\Sigma}_{\star\bullet}]' \ .
\end{aligned}
$$

(Augmented) Likelihood

$$
\begin{aligned}
f(\bar{\mathbf{y}}^F, s_F^2, \mathbf{b}, \mathbf{y}_\star^M, \mathbf{y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \mu^b, \boldsymbol{\theta}^b, \lambda^F, \mathbf{u}_\star) = \\
= f(s_F^2 \mid \lambda^F) \times \\
f(\bar{\mathbf{y}}^F \mid \mathbf{b}, \mathbf{y}_\star^M, \lambda^F) \times \\
f(\mathbf{b} \mid \boldsymbol{\theta}^b, \mu^b) \times \\
f(\mathbf{y}_\star^M \mid \mathbf{y}^M, \boldsymbol{\theta}^L, \boldsymbol{\theta}^M, \mathbf{u}_\star) \times \\
f(\mathbf{y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M)
\end{aligned}
$$

We know all these densities, and all except the first are Gaussian.

Observations

- Strictly speaking, the estimation of the parameters governing the GP prior of y^M should not be done by considering $f(\mathbf{y}^M \mid \boldsymbol{\theta}^L, \boldsymbol{\theta}^M)$ only
- if we try to estimate $(\boldsymbol{\theta}^L, \boldsymbol{\theta}^M)$ jointly with $\mathbf{u}_\star$ what may happen is that model uncertainty and calibration parameter uncertainty can get confounded
- We call $(\boldsymbol{\theta}^L, \boldsymbol{\theta}^M)$ the stage I parameters, and these are estimated in the first stage of the analysis
- Practical reasons
- First instance of *modularization*

Observations

- The structure of the likelihood makes it possible to integrate out $\mathbf{y}_\star^M$, or $\mathbf{b}$ or both
- This presents us with various MCMC possibilities
- Note that the priors on the so-called stage II parameters $\mu^b, \boldsymbol{\theta}^b, \lambda^F, \mathbf{u}_\star$ are still unspecified

Stage II parameters priors

- The prior on $\mathbf{u}$ has to be specified using expert knowledge
- The process $b(\cdot)$ is not observable, hence we have little information in the likelihood regarding μ^b and $\boldsymbol{\theta}^b$
- our choice of priors aims at producing a statistical estimation method that is stable; it's a mix of Bayesian and likelihood-based inference
- μ^b is fixed at zero
- $\boldsymbol{\beta}^b$ is fixed at an estimate
- λ^F and λ^b get exponential priors centered at (multiples of) estimates

Details

- Select a reasonable guess for $\mathbf{u}$, $\tilde{\mathbf{u}}$
- Using the emulator, predict $(\mathbf{y}^M(\mathbf{x}_i, \tilde{\mathbf{u}}),\ i = 1, \ldots, \ell) \equiv \tilde{\mathbf{y}}^M$
- Treat the vector $\mathbf{y}^F - \tilde{\mathbf{y}}^M$ as a realization from a multivariate normal with mean $\mu^b = 0$ and covariance $c^b(D^F)/\lambda^b + I/\lambda^F$
- Use specific software to compute estimates of β^b, λ^b and λ^F

- The only full conditional which is not standard form is that of $\mathbf{u}$ — Metroplis-Hastings step which we run for M steps
- $\mathbf{u}$ is likely to be correlated with λ^b, so it's a sensible idea to sample them as a block
- proposal: $q(\lambda, \mathbf{u} \mid \lambda_0, \mathbf{u}_0) = q(\lambda \mid \lambda_0)\ q(\mathbf{u} \mid \mathbf{u}_0)$ where

$$q(\lambda \mid \lambda_0) = f(\lambda \mid -)$$
$$q(\mathbf{u} \mid \mathbf{u}_0) = P\ \pi(\mathbf{u}) + (1 - P)\ U(\mathbf{u} \mid \mathbf{u}_0 - \boldsymbol{\delta}, \mathbf{u}_0 - \boldsymbol{\delta})$$

  where $P \in (0, 1)$ must be specified
- calibration parameters are notoriously difficult to estimate

Method 1

Direct Metropolis-Hastings within Gibbs:

- $[\mathbf{y}_\star^M, \mathbf{b} \mid \lambda^b, \lambda^F, \mathbf{u}_\star, D] \sim \text{Kalman filter}$
- $[\lambda^F \mid \lambda^b, \mathbf{u}_\star, \mathbf{y}_\star^M, \mathbf{b}, D] = \Gamma(\lambda^F \mid a_1^F, a_2^F)$
- $[\lambda^b, \mathbf{u}_\star \mid \lambda^F, \mathbf{y}_\star^M, \mathbf{b}, D] \propto \pi(\lambda^b)\ f(\mathbf{b} \mid \boldsymbol{\theta}^b)\ f(\mathbf{y}_\star^M \mid \mathbf{y}^M, \mathbf{u}_\star)$

The full conditional is $[\lambda^b \mid \lambda^F, \mathbf{u}_\star, \mathbf{y}_\star^M, \mathbf{b}, D] = \Gamma(\lambda^b \mid a_1^b, a_2^b)$

Method 2

Gibbs for $\pi(\mathbf{y}_\star^M, \lambda^b, \mathbf{u}_\star, \lambda^F, \mathbf{b} \mid D)$ can sample between

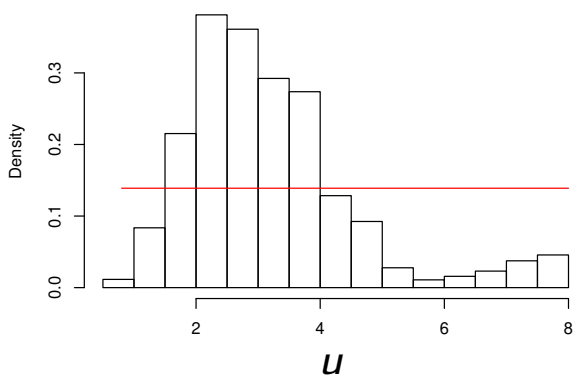
$$\pi(\mathbf{y}_\star^M, \lambda^b, \mathbf{u}_\star \mid \lambda^F, \mathbf{b}, D)$$
$$\pi(\lambda^F, \mathbf{b} \mid \mathbf{y}_\star^M, \lambda^b, \mathbf{u}_\star, D)$$

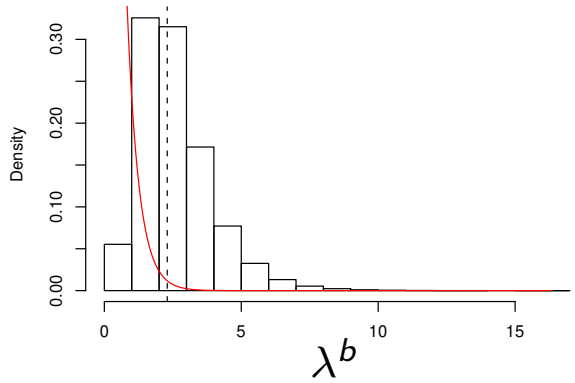
and

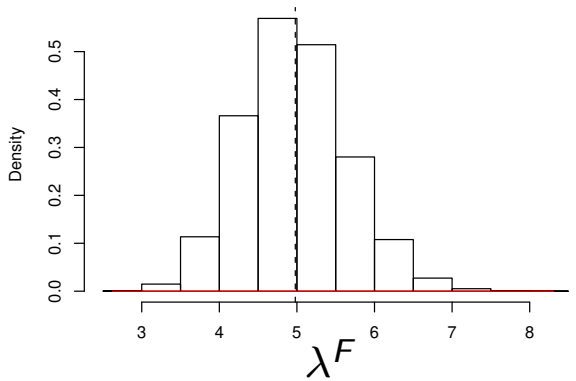
$$\pi(\mathbf{y}_\star^M, \lambda^b, \mathbf{u}_\star \mid \lambda^F, \mathbf{b}, D) = \pi(\mathbf{y}_\star^M \mid \lambda^b, \mathbf{u}_\star, \lambda^F, \mathbf{b}, D)\, \pi(\lambda^b, \mathbf{u}_\star \mid \lambda^F, \mathbf{b}, D)$$

This justifies a Gibbs sampler as before but with full conditional of $\lambda^b, \mathbf{u}_\star$ obtained by integrating out $\mathbf{y}_\star^M$ from the augmented likelihood.

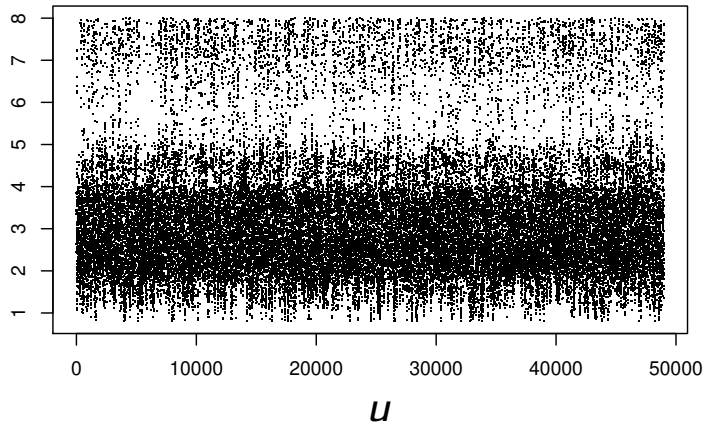
Spotweld results

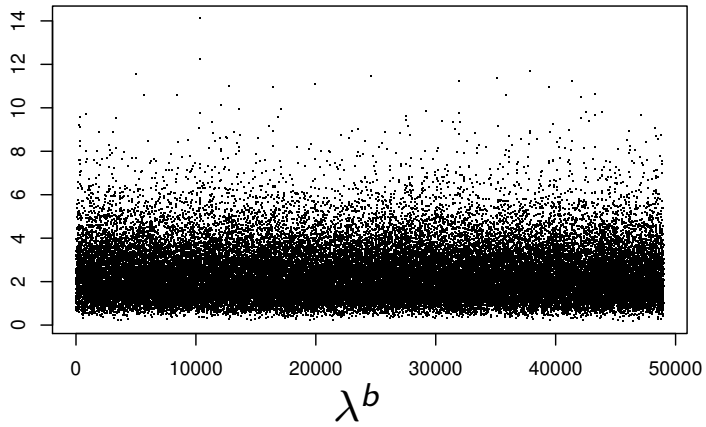


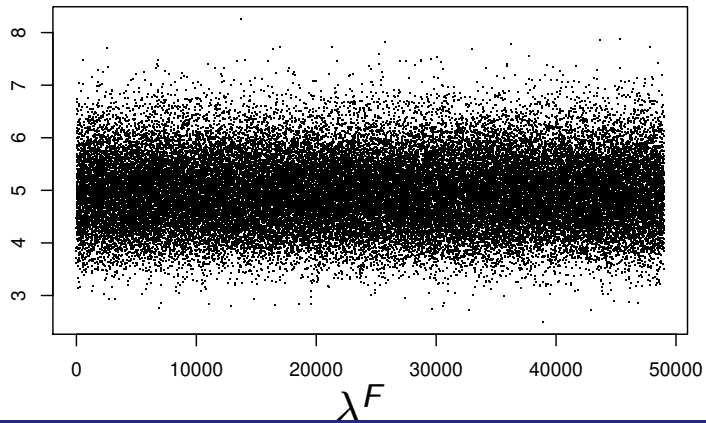




Spotweld results







Predict Reality

- Given a design set $D^F_{\text{new}} = \{\mathbf{x}_j\}$, let $D^M_{\mathbf{u},\text{new}} = \{(\mathbf{x}_j, \mathbf{u})\}$
- let $\mathbf{y}^M_{\mathbf{u},\text{new}} = y^M(D^M_{\mathbf{u},\text{new}})$, $\mathbf{b}_{\text{new}} = b(D^F_{\text{new}})$
- Draw samples from

$$\pi(\mathbf{y}^M_{\text{new}}, \mathbf{b}_{\text{new}} \mid \mathbf{y}^F, \mathbf{y}^M) =$$
$$= \int f(\mathbf{y}^M_{\mathbf{u},\text{new}}, \mathbf{b}_{\text{new}} \mid \mathbf{y}^F, \mathbf{y}^M, \mathbf{u}, \lambda^F, \lambda^b) \times$$
$$\times \pi(\mathbf{u}, \lambda^F, \lambda^b \mid \mathbf{y}^F, \mathbf{y}^M)\ d\mathbf{u}\ d\lambda^F\ d\lambda^b$$

- Denote them by $(\mathbf{y}^{M(t)}_{\text{new}}, \mathbf{b}^{(t)}_{\text{new}})$, $t = 1, \ldots, T$

Bias-corrected prediction of reality

- Then,

$$\mathbf{y}_{\text{new}}^{M(t)} + \mathbf{b}_{\text{new}}^{(t)}, \quad t = 1, \ldots, T$$

  is a sample from the posterior predictive distribution of y^R evaluated at D_{new}^F

- and

$$\hat{\mathbf{y}}_{\text{new}}^R = \frac{1}{T} \sum_{t=1}^{T} [\mathbf{y}_{\text{new}}^{M(t)} + \mathbf{b}_{\text{new}}^{(t)}]$$

  is a simulation-based estimate of the *bias-corrected prediction of reality*

Pure-model prediction of Reality

- If there are no calibration parameters, the pure-model prediction of $y^R(\mathbf{x})$, $\mathbf{x} \in D^F_{\text{new}}$ is simply $y^M(\mathbf{x})$, computed by actually running the computer model or using the emulator
- If there are calibration parameters, we then need an estimate of $\mathbf{u}$, e.g. the posterior mean or mode
- In that case, the pure-model prediction of $y^R(\mathbf{x})$, $\mathbf{x} \in D^F_{\text{new}}$ is $y^M(\mathbf{x}, \hat{\mathbf{u}}) \equiv \mathbf{y}^M_{\text{new}}(\hat{\mathbf{u}})$
- Samples form the posterior predictive distribution of the bias of this predictor are

$$\mathbf{b}^{(t)}_{\hat{\mathbf{u}}} = \mathbf{y}^{M(t)}_{\text{new}} + \mathbf{b}^{(t)}_{\text{new}} - \mathbf{y}^M_{\text{new}}(\hat{\mathbf{u}}), \quad t = 1, \ldots, T$$

Tolerance Bounds

Predictive accuracy assessement: with probability γ, the prediction is within τ of the true $y^R(\mathbf{x})$

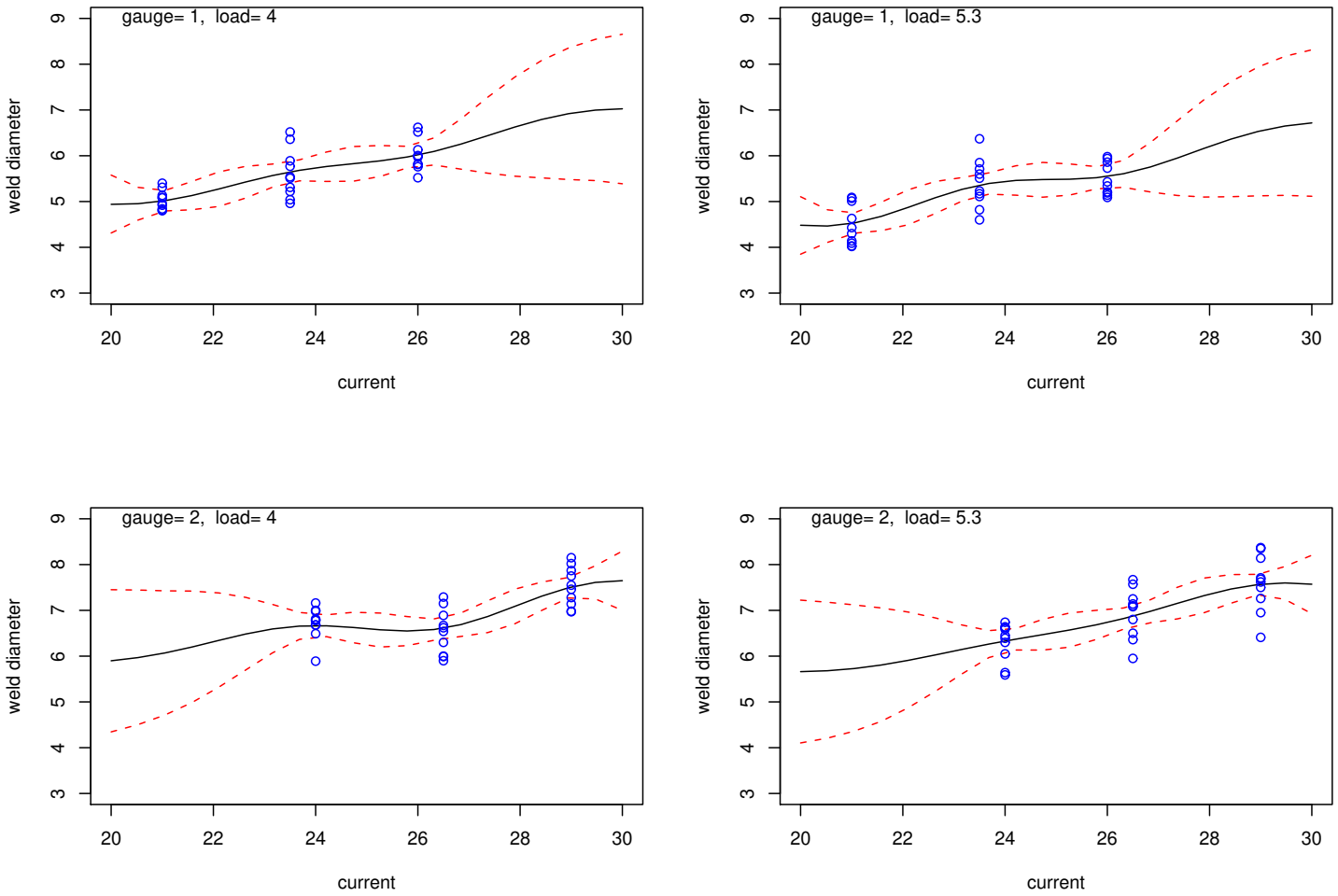
- Pure-model prediction: determine $\boldsymbol{\tau}$ such that $\gamma \times 100\%$ satisfy (component-wise statement)

$$|\mathbf{y}^M_{\text{new}}(\hat{\mathbf{u}}) - (\mathbf{y}^{M(t)}_{\text{new}} + \mathbf{b}^{(t)}_{\text{new}})| < \boldsymbol{\tau}$$

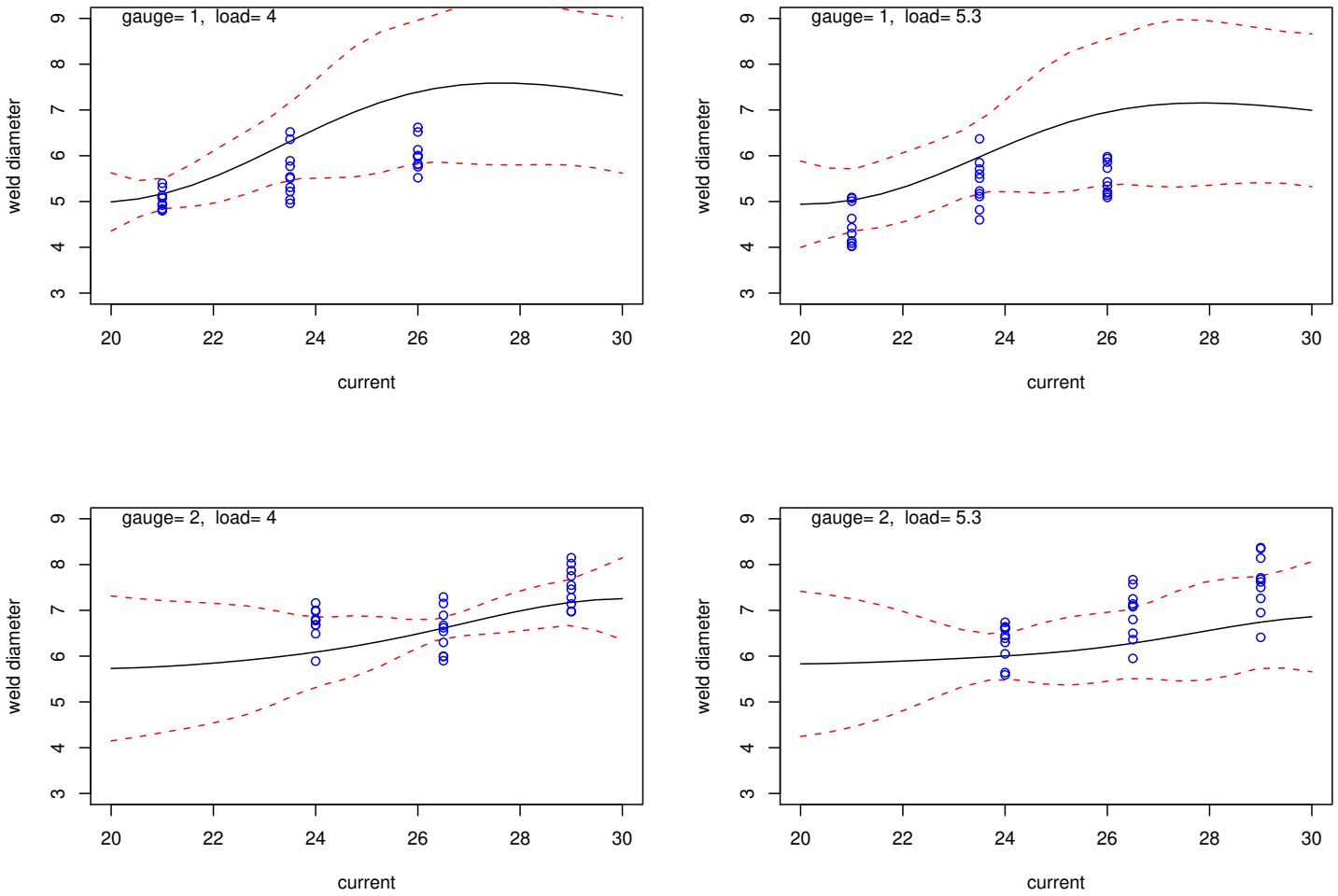
- Bias-corrected prediction: determine $\boldsymbol{\tau}$ such that $\gamma \times 100\%$ satisfy (component-wise statement)

$$|\mathbf{y}^R_{\text{new}} - (\mathbf{y}^{M(t)}_{\text{new}} + \mathbf{b}^{(t)}_{\text{new}})| < \boldsymbol{\tau}$$

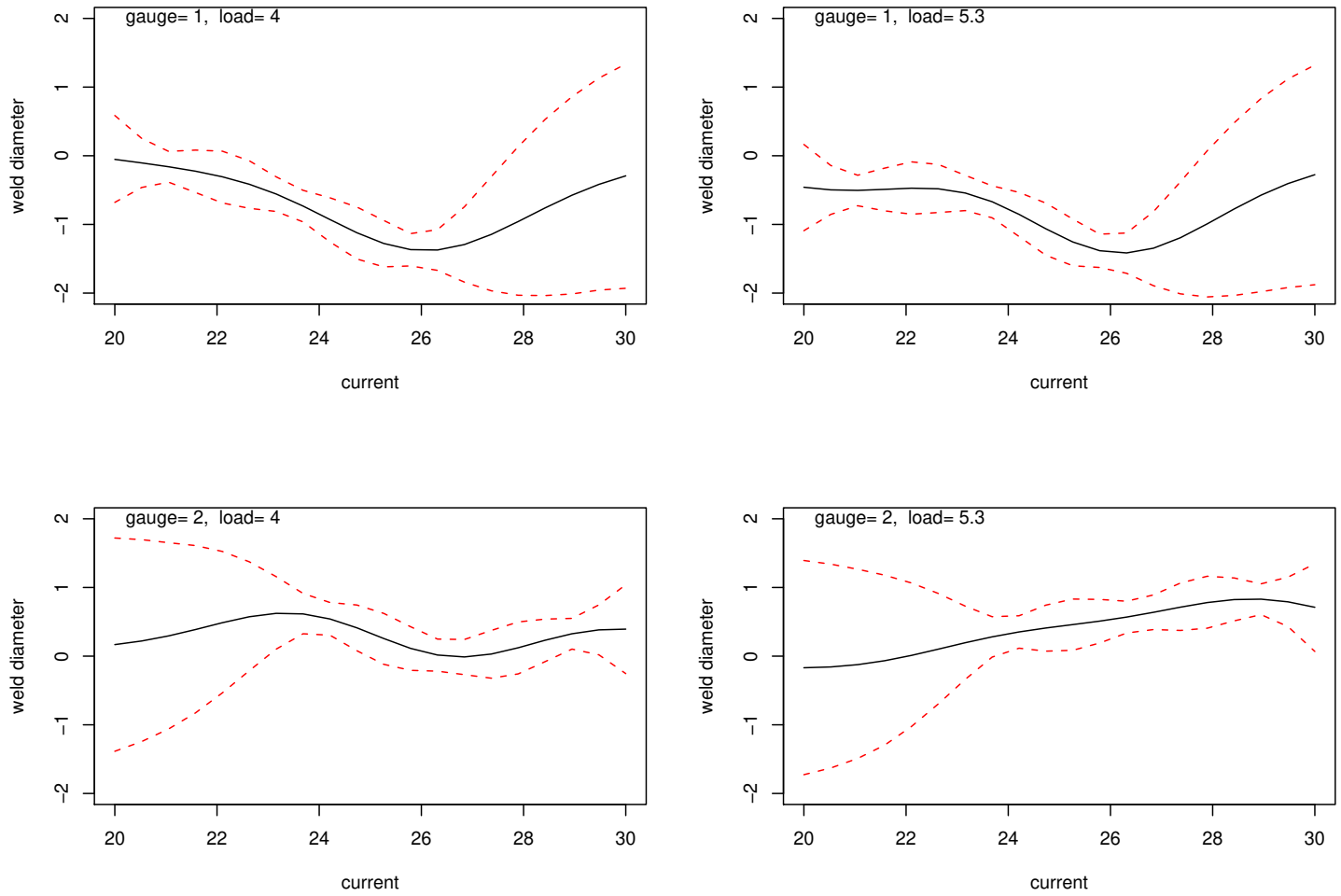
Bias–corrected predictions – Spotweld Data



Pure–model prediction – 90% tolerance bounds



Bias of Pure−model prediction



- Combining model and field data to simultaneously calibrate and validate a computer model
- Model inadequacy as a key concept
- Mix of Bayesian and likelihood techniques to produce methodology that can be used by non-experts
- Bias-corrected and pure-model predictions of reality
- Tolerance bounds