Dealing with nuisance parameters for Bayesian model calibration

PRESENTED BY

Kellin Rumsey[1] & Gabriel Huerta [1]

[1] Department of Statistical Sciences
Sandia National Laboratories

*co-authors: Lauren Hund (SNL) and Derek Tucker (SNL).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. SAND NO. 2019-4335 C

Statement

This work was supported by a Sandia National Laboratories Laboratory Directed Research and Development (LDRD) grant. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND 2019-4335 C

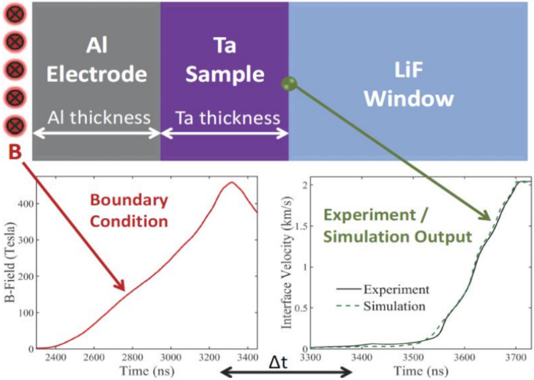
Background

- **Dynamic material properties experiments**: access to the most extreme temperatures and pressures attainable.
- **Sandia National Labs Z-machine**: pulsed power driver that can deliver massive electrical currents over very short timescales (of the order of 60MA over $1\mu s$)).
- **Goal:** Understanding of material models at extreme conditions by coupling computational simulations with experimental data.

Background

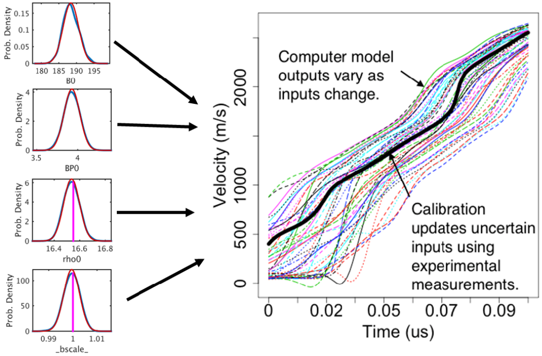
- **Goal**: Generalized solution for calibrating dynamic material models.
- **Physicists**: ideally want a solution that does not necessarily require a statistician in the loop.
- **Parameters of interest are physical**: material properties with ”true” value that is of interest.
- **Ideally:** robust algorithm for UQ parameter calibration.
- **Firstly**: Calibrate a well-understood model - two parameters of the equation of *state of tantalum*.

Experimental setup



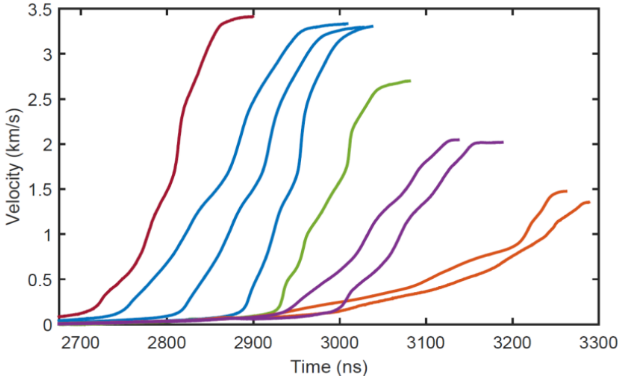
- "By coupling experimental and simulated velocity traces, parameters of the tantalum (Ta) equation of state (EOS) can be estimated".
- Massive electric currents treated as boundary conditions.
- Stress wave propagates thru system.

Calibration



- Uncertain inputs generate velocity curves using a computer model.
- Probability distributions look for "agreement" of outputs and measurements.
- Bayesian framework is a natural in this context...

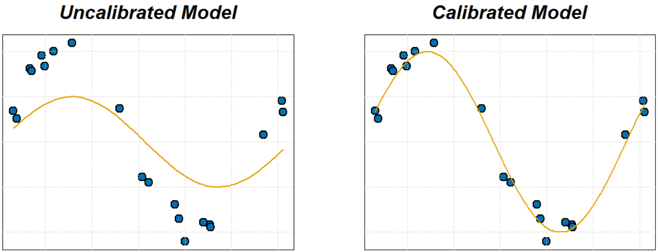
Challenges



- How to accurately estimate uncertainties?
- Calibration parameters have physical interpretation.
- Lots of *nuisance* parameters.

Approach

- Bayesian Model Calibration (BMC) (Kennedy & O'Hagan 2001) often used to "tune" computer model.
- *Calibrated* model for prediction (interpolation).
- Partitioned into *physical parameters* and *nuisance parameters*.



Approach

- Kennedy & O'Hagan 2001 model,

$$\begin{aligned} y(x_i) &= \eta(x_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + \epsilon_i \\ \epsilon_i &\overset{iid}{\sim} N(0, \sigma^2) \\ \delta(\cdot) &\sim GP(\boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta) \end{aligned}$$

- x_i are known *inputs* (experiment test conditions, time)
- $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\gamma})$ are calibration parameters.
- η is the true value of the outcome as a function of x and $\boldsymbol{\theta}$.
- ϵ_i is a measurement error.
- $\delta(\cdot)$ is a *discrepancy function term*.

Our Framework

- We model the i^{th} observation in the j^{th} experiment as,

$$y(x_{ij}) = \eta(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\gamma}_j) + \delta(x_{ij}) + \epsilon_{ij}$$

- $\boldsymbol{\alpha}$ are the (unknown) values of the calibration parameters.
- $\boldsymbol{\gamma}_j$ unknown values of experimental uncertainties for experiment j.
- $y(x_{ij})$ is the observed velocity at time x_{ij}.
- $\eta(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\gamma}_j)$ is the computer model output at x_{ij}.
- $\delta(x_{ij})$ is a G-P discrepancy term.
- ϵ_{ij} are measurement uncertainties at x_{ij}.

Dynamic material property calibration

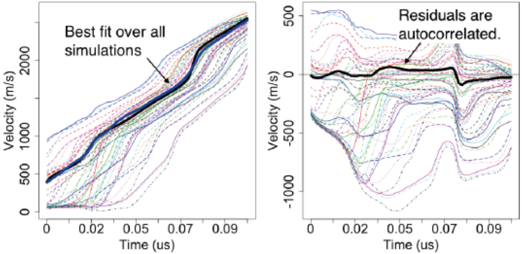
- BMC framework to obtain inference for two material properties of Tantalum.
- B_0 and B_0' are the Bulk modulus of tantalum and its pressure derivative.

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (B_0, B_0')$$

- Four nuisance that may vary across $p = 9$ experiments
  - Tantalum density - γ_1
  - Magnetic field scaling - γ_{2j}, $j = 1, 2, \cdots 9$
  - Aluminum thickness- γ_{3j}, $j = 1, 2, \cdots 9$
  - Tantalum thickness - γ_{4j}, $j = 1, 2, \cdots 9$
- Potential for overfitting and lack of identifiability.

Issues

- Model can fit well to data, solutions far from *true* parameter values.
- Can we diagnose such overfitting? Can we mitigated it?
- **Model discrepancy** can reduce the identifiability of the calibration parameters.



Model Discrepancy

- Without strong assumptions about discrepancy, KOH should not be expected to provide correct inferences.
- $\delta()$ and $\boldsymbol{\theta}$ are not jointly identifiable (Loeppky et al., 2006; Arendt et al., 2012; Brynjarsdóttir and OH́agan, 2014; Tuo and Wu, 2016).
- Robust alternatives to G-P discrepancy?
  - Brown and Hund (2018) use *power likelihoods*.

$$p(\boldsymbol{\theta}|Y) \propto exp\left(-wl(Y|\boldsymbol{\theta})\right) p(\boldsymbol{\theta})$$

  - Problems with fewer experimental curves and more nuisance parameter are harder.
  - Time series models?

Nuisance parameters and overfitting

- **Aluminum and Tantalum thickness parameters:** These nuisance parameters are measured with a device which we beleive to be well registered.
- Measurement error is *exclusive source* of uncertainty. The prior mean and variance of these nuisance parameters are well known.
- Nuisance parameters are standardized (mean 0, variance 1).
- The *standard informative (SI) prior* is:

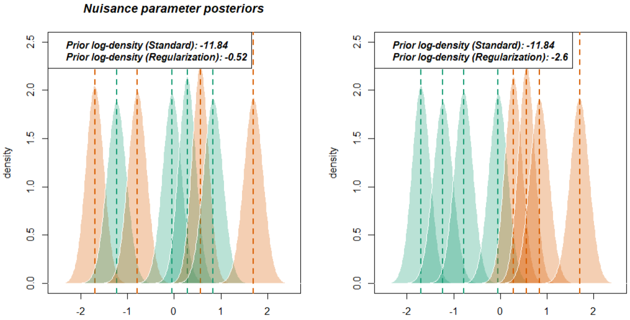
$$(\gamma_{k1}, \gamma_{k2}, \cdots \gamma_{k9}) \sim N(0, I_9), \; k = 2, 3, 4$$

- "True values" are expected to look like a draw from a $N(0, I_9)$ distribution.

Nuisance parameters and overfitting

- Three types of overfitting:
- **Overdispersion:** Posterior estimates are collectively too large.
  - Indicates a "calibration solution". Good fit to data but scientifically unreasonable.
  - Standard informative prior usually prevents this from occurring.
- **Underdispersion:** Posterior estimates are collectively too close to 0.
  - Can lead to underestimation of uncertainty in $\boldsymbol{\alpha}$.
  - Standard informative prior will *not* address this case.
- **Collective Bias:** The posterior estimates are collectively biased (i.e. all are negative).
  - Indicates a systematic bias *across experiments*.
  - Can lead to biased estimates of $\boldsymbol{\alpha}$ to compensate.

Collective Bias for 2 nuisance-sets



- Left: No grouping occurs.
- Right: Collective bias implies systematic overfitting across experiments.
- Standard prior assigns same values.

A metric for overfitting

- We define,

$$M_\gamma = \frac{1}{p}\sum_{j=1}^{p}\gamma_j \qquad V_\gamma = \frac{1}{p-1}\sum_{j=1}^{p}\left(\gamma_j - M_\gamma\right)^2$$

- Prior beliefs about problem structure suggests:

$$M_\gamma \approx 0 \qquad V_\gamma \approx 1$$

- Under standard normal,

$$\pi_{M_\gamma, V_\gamma}(m, v) = N(m \mid 0, 1/p) \times \left[(p-1)\chi^2(v(p-1) \mid p-1)\right]$$

- Reasonable to check that the estimates $\hat{M}_\gamma$ and $\hat{V}_\gamma$ are *coherent* with prior.

A metric for overfitting

- **Definition:** We say that (m, v) is *more coherent with the prior* than (m', v') if

$$\pi_{M_\gamma, V_\gamma}(m, v) > \pi_{M_\gamma, V_\gamma}(m', v')$$

- Define the set of all points which are less coherent with the prior than $(\hat{M}_\gamma, \hat{V}_\gamma)$

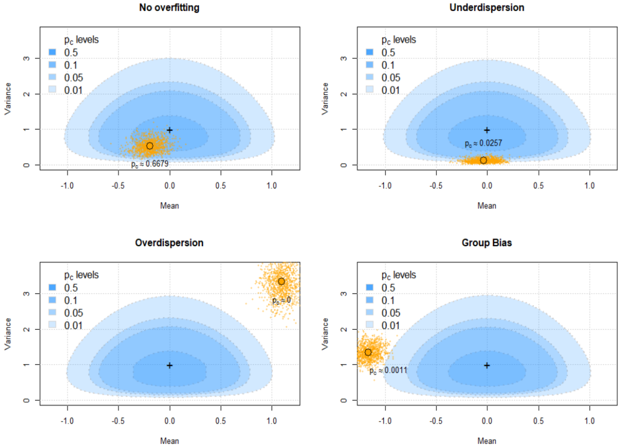
$$\Gamma_{\hat{M}_\gamma, \hat{V}_\gamma} = \left\{(m, v) \mid \pi_{M_\gamma, V_\gamma}(\hat{M}_\gamma, \hat{V}_\gamma) > \pi_{M_\gamma, V_\gamma}(m, v)\right\}$$

- *Probability of prior coherency* of $(\hat{M}_\gamma, \hat{V}_\gamma)$

$$p_c(\hat{M}_\gamma, \hat{V}_\gamma) = \int_{\Gamma_{\hat{M}_\gamma, \hat{V}_\gamma}} \pi_{M_\gamma, V_\gamma}(m, v)\, dm dv$$

$$\approx \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{1}\left(\pi_{M_\gamma, V_\gamma}(\hat{M}_\gamma, \hat{V}_\gamma) > \pi_{M_\gamma, V_\gamma}(m_\ell, v_\ell)\right)$$

Diagnostic plot for simulated case $p = 10$



- Orange: Point estimates and posterior draws of (M_γ, V_γ)
- Blue: Prior probability contours.

The moment penalization prior

- Overfitting of nuisance parameters leads to $(\hat{M}_\gamma, \hat{V}_\gamma)$ with low prior coherency.
- The *moment penalization (MP) prior* **penalizes** solutions with low prior coherency.
- Let $h_a(x)$ be a function which takes larger values when x is close to a.

$$\pi_\gamma^{MP}(\boldsymbol{\gamma}) \propto h_0(M_\gamma)h_1(V_\gamma)$$

- Tries to encourage solutions with

$$M_\gamma \approx 0 \qquad\qquad V_\gamma \approx 1$$

The moment penalization prior

- Simple and effective choice for $h_a(x)$: Gaussian kernels

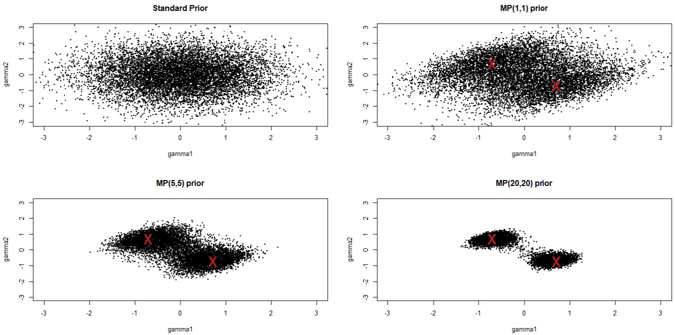
$$\pi_{\gamma}^{MP}(\boldsymbol{\gamma}) \propto \exp\left[-\lambda_1 M_{\gamma}^2\right]\ \exp\left[-\lambda_2 (V_{\gamma}-1)^2\right]$$

- λ_1 and λ_2 control how strongly we want to enforce constraints.
- *Reparameterize:* $\omega_1 = 2Var(M_{\gamma})\lambda_1$ and $\omega_2 = 2Var(V_{\gamma})$
- Write $\boldsymbol{\gamma} \sim MP(\omega_1, \omega_2)$ to mean that,

$$\pi_{\gamma}^{MP}(\boldsymbol{\gamma}) \propto \exp\left[-\frac{p\omega_1}{2} M_{\gamma}^2\right]\ \exp\left[-\frac{(p-1)\omega_2}{4}(V_{\gamma}-1)^2\right]$$

- $\boldsymbol{\gamma} \sim MP(1,1)$ is the *standard moment penalization prior.*

Samples from the Standard MP prior



- $10,000$ draws via M-H for $p = 2$.
- As $\omega \to \infty$ all density is placed on $\pm(1/\sqrt{2}, -1/\sqrt{2})$
- As p grows, the induced marginal priors become $N(0, 1)$.

Moment penalization in the limit

- **Z-Regularization:** Consider a set of p latent variables Z.

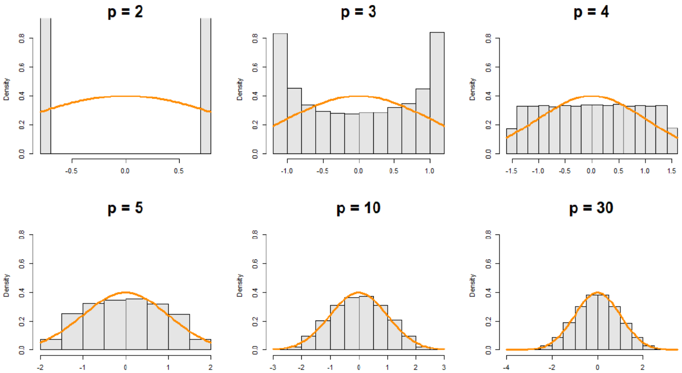
$$
\begin{aligned}
Z_1, \cdots Z_p &\overset{iid}{\sim} N(0,1) \\
\gamma_k &= \frac{Z_k - \bar{Z}}{S_Z}
\end{aligned}
$$

- We enforce that $M_\gamma = 0$ and $V_\gamma = 1$.
- This approximates the limit situation for $MP(\omega_1, \omega_2)$

$$\omega_1 \to \infty; \quad \omega_2 \to \infty.$$

- As p increases, marginal prior on γ_k goes to $N(0,1)$.

Z-Regularization: Marginal prior on γ_k



Data informed regularization

- The MP prior harnesses the known structure of the problem and forces each *group* to behave reasonably.
- Not appropriate for all cases, and a more general form of regularization is required.
- We consider the class of *Global-Local Gaussian scale mixtures*:
- For $k = 1, \cdots p$,

$$\gamma_k \mid (\tau, \psi_k) \overset{ind}{\sim} N(0,\ \tau\psi_k)$$

$$\tau \sim g() \qquad \text{and} \qquad \psi_k \sim g_k()$$

- Commonly used in sparse linear model settings.

Data informed regularization

- Horseshoe prior is obtained by setting

$$\tau \sim C_+(0, \sigma) \qquad \text{and} \qquad \psi_k \sim C_+(0, \sigma_k)$$

- *Shrink globally:* When regularization is required, global parameter τ becomes very small.
- *Act locally:* Active components are selected by allowing ψ_k to become very large.
- If p is large, this can significantly increase the cost of BMC.

Example: The simple machine

- Brynjarsdottir and O'Hagan (2014): The simple machine delivers work

$$\zeta(x) = \frac{E\,x}{1 + x/20}$$

  - x is the amount of *effort* put into the machine.
  - E is the *efficiency* of the machine.
  - Denominator accounts for loss of work due to *friction*.
- The naive simulator introduces model discrepancy

$$\eta(x, E) = Ex$$

Example: The simple machine

- We consider $p = 10$ simple machines, and introduce base efficiency G_j as a machine-dependent nuisance parameter.
- Inputs $x_1, x_2, \cdots x_n$ evenly spaced over $[1, 4]$
- Data generating process:

$$y_{ij} = G_j + \frac{E\, x_i}{1 + x_i/20} + \epsilon_i$$
$$G_j \sim N(0, 0.05^2)$$
$$\epsilon_i \sim N(0, 0.01^2)$$

- Naive simulator:

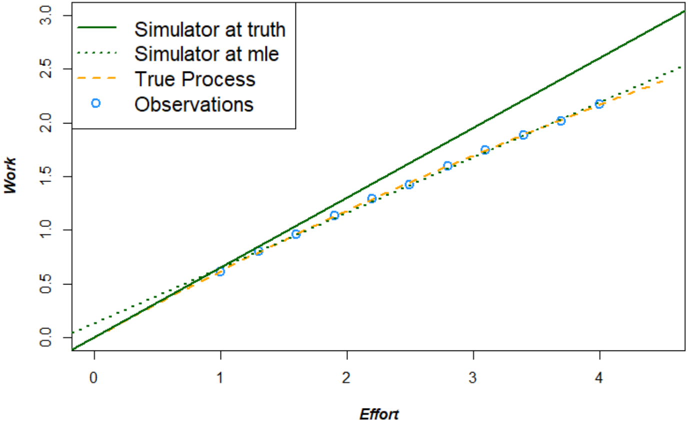
$$\eta(x, E, G) = G + E\, x$$

- True efficiency is $E = 0.65$. Standardize parameters:

$$\alpha = \frac{E - 0.65}{0.3} \sim N(0,1) \qquad \gamma_k = \frac{G_k - 0}{0.05} \sim N(0,1)$$

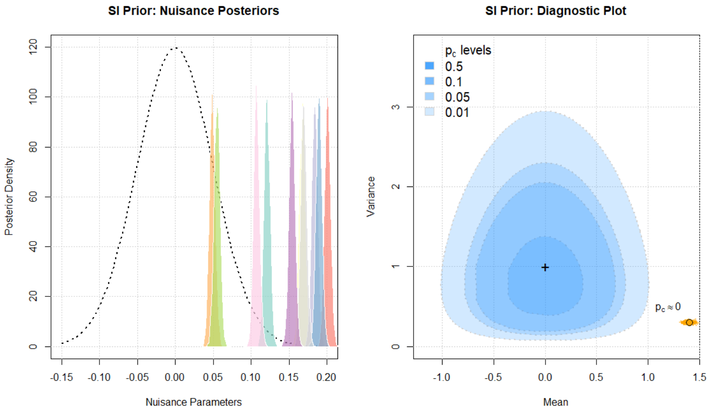
Example: The simple machine

- Model discrepancy leads to *systematic bias*.



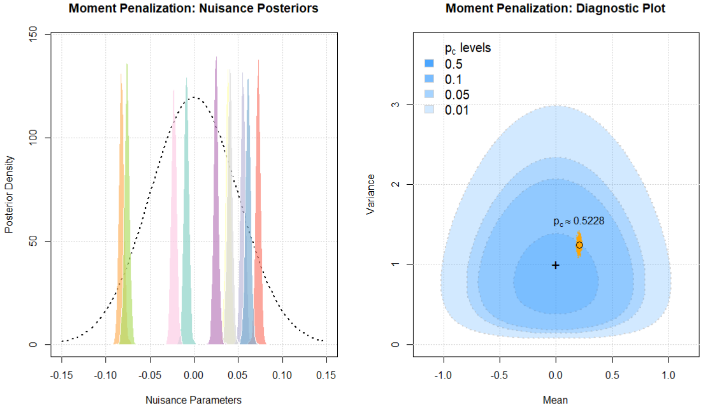
Example: The simple machine

- Under standard informative prior

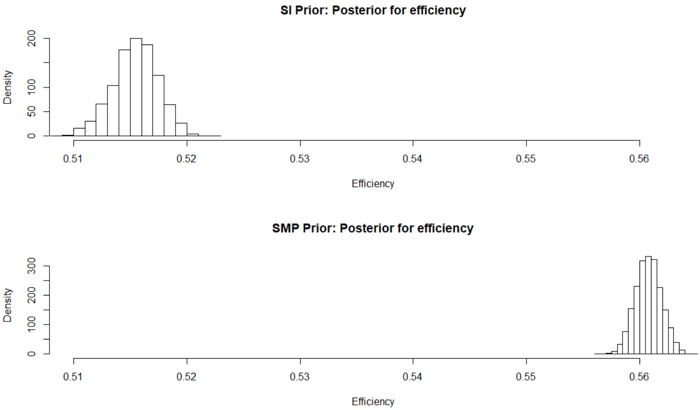


Example: The simple machine

- Under moment penalization prior



Example: The simple machine



- Posterior inference improves under MP, but is still far from truth.
- This is still valuable information! Model discrepancy is leading to biased inference on the parameter of interest.

Example: Borehole function

- Models water flow through a borehole (An & Owen, 2001; Harper & Gupta, 1983)
- The true process,

$$\zeta(x, \boldsymbol{\theta}) = \frac{2\pi T_u \Delta H}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

- Most of the inputs are treated as known
  - $r, T_u, T_l, \Delta H$ fixed at usual values (Surjanovic & Bingham, 2017).
- Compare the moment penalization prior to the standard informative prior.

Example: Borehole function

- $x = L$ known input where L is the length of the borehole (meters).
- The input r_w, radius of the borehole (nuisance parameter),

$$\gamma = \frac{r_w - 0.1}{0.0161812} \sim N(0,1)$$

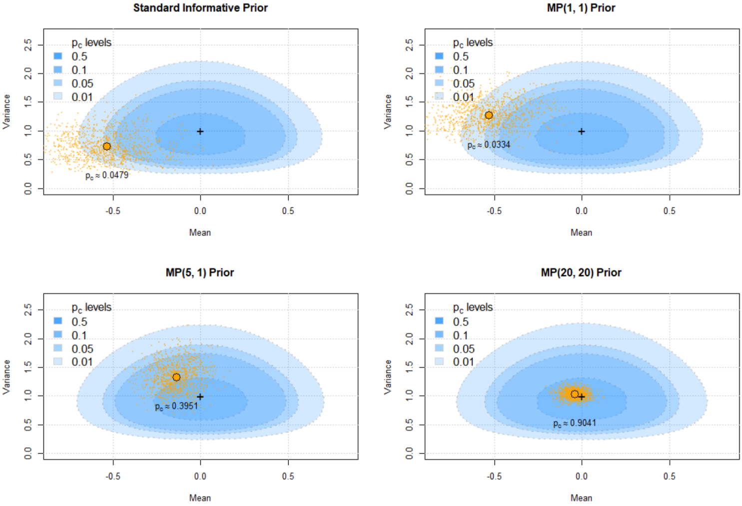
- The physical parameter K_w, hydraulic conductivity of the borehole (meters per year).

$$\alpha = \frac{K_w - 10950}{632.2} \sim N(0,1)$$

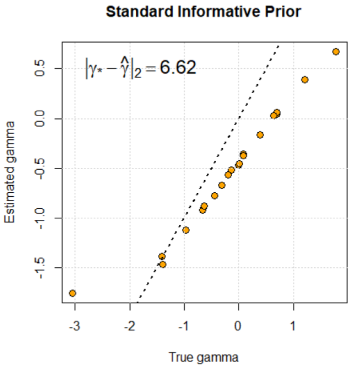
- A *low fidelity simulator*,

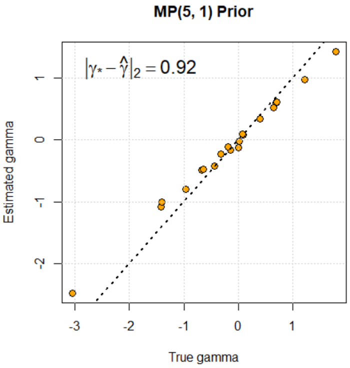
$$\eta(x, \boldsymbol{\theta}) = \frac{2\pi T_u \Delta H}{\ln(r/r_w)\left(1.5 + \frac{1.4 L T_u}{\ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

Example: Borehole function Diagnostic plot

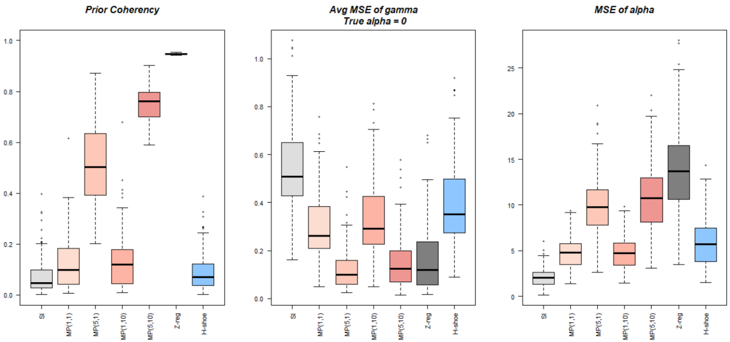


Example: Borehole function
Estimation of nuisance parameters





Example: Borehole function Simulation study



- $p = 5, n = 10$.
- $\alpha_\star \in \{-1, 0, 2\}$.

Example: Borehole function Simulation study

	Prior Coherency			Avg MSE of γ			MSE of α		
Prior	$\alpha_\star = 0$	$\alpha_\star = -1$	$\alpha_\star = 2$	$\alpha_\star = 0$	$\alpha_\star = -1$	$\alpha_\star = 2$	$\alpha_\star = 0$	$\alpha_\star = -1$	$\alpha_\star = 2$
SI	0.05	0.02	0.19	0.51	0.52	0.51	**2.07**	2.03	**2.07**
MP(1,1)	0.10	0.06	0.27	0.26	0.38	0.15	4.81	**1.97**	13.21
MP(5,1)	0.50	0.45	0.59	**0.10**	**0.12**	**0.06**	9.79	6.5	18.71
MP(1,10)	0.12	0.06	0.34	0.29	0.41	0.17	4.76	2.28	13.29
MP(5,10)	0.76	0.73	0.83	0.12	0.15	0.08	10.72	7.29	19.42
Z-reg	**0.95**	**0.95**	**0.95**	0.12	0.15	0.09	13.66	10.18	22.26
H-shoe	0.07	0.04	0.18	0.35	0.48	0.27	5.73	2.57	14.64

**Summary of the Borehole simulation results. Reported value is the median across 100 simulations. Bold value indicates "best" value in the column.*

- Posterior inference on α gets worse as inference on nuisance parameters improves.
- Still valuable information! Model discrepancy is leading to biased inference on the parameter of interest.

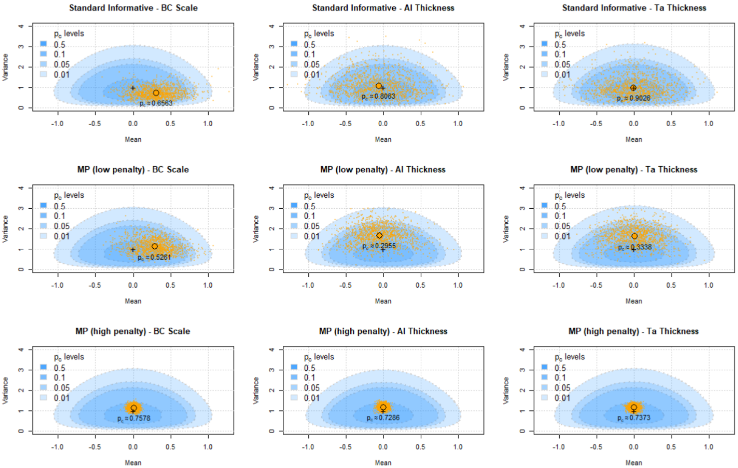
Dynamic material property calibration revisited

- Inference for two material properties of Tantalum.
- B_0 and B'_0 are the Bulk modulus of tantalum and its pressure derivative.

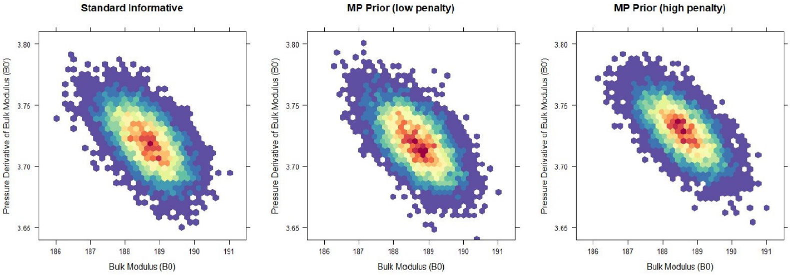
$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (B_0, B'_0)$$

- Four nuisance that may vary across $p = 9$ experiments
  - Tantalum density - γ_1
  - Magnetic field scaling - γ_{2j}, $j = 1, 2, \cdots 9$
  - Aluminum thickness- γ_{3j}, $j = 1, 2, \cdots 9$
  - Tantalum thickness - γ_{4j}, $j = 1, 2, \cdots 9$
- Perform BMC for SI, SMP and MP(20, 40) priors.

Dynamic material property calibration Diagnostic plots



Dynamic material property calibration
Physical parameter posteriors



- Similar posterior inference in all cases.
- Indicates that model discrepancy is unlikely to be causing bias in the parameters of interest.

Conclusions

- Overfitting of nuisance parameters leads to systematic bias which is often a symptom of model discrepancy.
- In complex high-dimensional problems, with appropriate problem structure, we can:
  - **Identify:** Probability of prior coherency identifies many types of overfitting, should it occur.
  - **Reduce:** The moment penalization prior reduces the systematic bias of the nuisance parameters.
  - **Diagnose:** Examine the sensitivity of posterior inference in order to diagnose the presence and effect of model discrepancy on the parameters of interest.

References

- Arendt, P.D., Apley, D.W. and Chen, W. (2016) A preposterior analysis to predict identifiability in the experimental calibration of computer models. IIE Trans.
- Bayarri, M. J., et al. (2007a) Computer model validation with functional output. Ann. Statist.,
- Bhattacharya, A., et al. (2015) Dirichlet Laplace Priors for Optimal Shrinkage, JASA.
- Brown, J. L., et al. (2014) Flow strength of antalum under ramp compression to 250 GPa. J. Appl. Phys
- Brown, J. L. and Hund, L. B. (2018) Model calibration via deformation. J.R. Statisti. Soc. C.
- Brynjarsdottir, J. and OHagan, A. (2014) Learning about physical parameters: the importance of model discrepancy. Invrs. Probl.
- Kennedy, M. C. and OHagan, A. (2001) Bayesian calibration of computer models (with discussion). J. R. Statist. Soc. B.

Selection of Hyper-parameters

- Adequacy of prior depends on selection of ω_1 and ω_2.
- **Update or estimate with MAP.** Weakly informative priors allow likelihood to dominate the selection. Problem of overfitting may not be addressed.
- **Cross validation.** Prediction or posterior based criteria leads to overfitting. Computationally difficult.
- **Sequential approach:** Use the diagnostic plot to increase ω_1 and ω_2 sequentially until prior coherency is reasonable.

Comparison: SI vs SMP

- For a given set of p nuisance parameters $(\gamma_1, \cdots \gamma_p)$ we compute:

$$\log \pi_{SI}(\boldsymbol{\gamma}) = \sum_{k=1}^{10} \log\left(N(\gamma_k \mid 0, 1)\right) = -\frac{1}{2}\log(2\pi) - \sum_{k=1}^{10} \frac{\gamma_k^2}{2}$$

$$\log \pi_{MP}(\boldsymbol{\gamma}) = c - \frac{p\omega_1}{2}\left(M_\gamma\right)^2 - \frac{(p-1)\omega_2}{4}\left(V_\gamma^{(m)} - 1\right)^2$$

where M_γ and V_γ denote the mean and variance of $\boldsymbol{\gamma}$.

- Think about these prior log-densities as penalties (small values) and rewards (large values).
- Compare penalty assigned by each prior over a wide range of potential nuisance sets.

Comparison: SI vs SMP

- Compare penalty assigned by each prior over a wide range of potential nuisance sets.
- **No overfitting:** Consider candidates for which overfitting is unlikely to be present. $\boldsymbol{\gamma} \sim N(0, I_{10})$.
- **Overdispersion:** Explore regions of the nuisance space in which magnitude of nuisance parameters is larger than expected. $\boldsymbol{\gamma} \sim N(0, 4\ I_{10})$.
- **Underdispersion:** Magnitude is smaller than expected. $\boldsymbol{\gamma} \sim N(0, \frac{1}{4} I_{10})$.
- **Collective Bias:** We explore regions where nuisance parameters are collectively biased compared to our expectations. $\boldsymbol{\gamma} \sim N(-1, I_{10})$

Comparison: SI vs SMP