

BNNs Hyperparameter Sampling: Comparing Gibbs vs. HMC (NUTS)

01/28/2021

Objective

Overview

Results

Prior Checks / Conformance

Objective

Background

- ▶ The parameters (weights and biases) of a Bayesian Neural Network (BNN) are generally assigned to **groups**
- ▶ The prior distributions for parameters in a groups share **common hyperparameters**
 - ▶ For example, all weights in the j^{th} layer might be assigned priors: $P(w_{i,j}) = N(0, \sigma_j)$, $P(1/\sigma_j^2) = Gamma(\alpha, \beta)$
- ▶ The hyperparameters are also parameters of the probabilistic model, whose posterior distribution is a pre-requisite for the predictive distribution of target variable on unseen data
 - ▶ Thus, sampling the posterior distribution of the hyperparameters is arguably as important as the posterior of the low-level parameters (weights, biases)

Objective

- ▶ For BNNs with normal priors on weights/biases, a gamma distribution is oft-used to sample the precision of the normal distribution in a group
- ▶ Neal (1995) pioneered this idea, and used Gibbs sampling for sampling hyperparameters ($P(\sigma|w_{i,j})$ has an analytical form)
 - ▶ (The scheme couples Gibbs sampling on hyperparameters and Hamiltonian Monte Carlo updates on the weights)
- ▶ Our objective is to test if contemporary adaptive HMC methods (such as No-U-Turn sampler) offer a competitive replacement for Gibbs sampling for hyperparameter sampling in BNNs

Overview

Experimental Overview

Three sampling techniques for parameters/hyperparameters:

- ▶ HMC (Centered Parametrization, NUTS for adaptation)
- ▶ HMC (Non-Centered Parametrization, NUTS for adaptation)
- ▶ FBM (Gibbs sampling for hyperparameters, HMC for all others)

Centered vs. Non-Centered Parametrization

- ▶ Centered:

$$f(W|\mu, \tau) = \text{Normal}(\mu, \tau^{-0.5})$$

- ▶ Non-Centered:

$$W = \mu + \tau^{-0.5} W_{norm}, f(W_{norm}) = \text{Normal}(0, 1)$$

- ▶ Applies to all low-level weights and biases

Global Assumptions

Architecture:

- ▶ 1 hidden layer, 8 hidden units
- ▶ tanh activation
- ▶ *parameter groups*: input-hidden weights, hidden-output weights, hidden biases, output bias
- ▶ variance of hidden-output weights scaled by number of units in hidden layer, to approach Gaussian Processes as number of units grows

Data:

- ▶ from FBM example
- ▶ input dimensions = 1, output dimensions = 1

Hyperparameter Priors (FBM Notation):

- ▶ input-hidden weights hyperparameter: 0.05:0.5
- ▶ hidden-output weights hyperparameter: 0.05:0.5
- ▶ hidden layer biases hyperparameter: 0.05:0.5
- ▶ output bias hyperparameter: 100

Assumptions

Tuning Parameters for NUTS:

- ▶ adapt_delta: 0.8 (default = 0.8)
- ▶ max_treedepth: 10 (default = 10)
- ▶ adapt_gamma: 0.05 (default = 0.05)

Initial Values:

- ▶ Weight Precision: 1
- ▶ Biases Precision: 1
- ▶ Target Noise Precision: 1
- ▶ Weights: 0
- ▶ Biases: 0

4 separate chains were used with the same initializations but different random seeds.

Results

Predictive Uncertainty

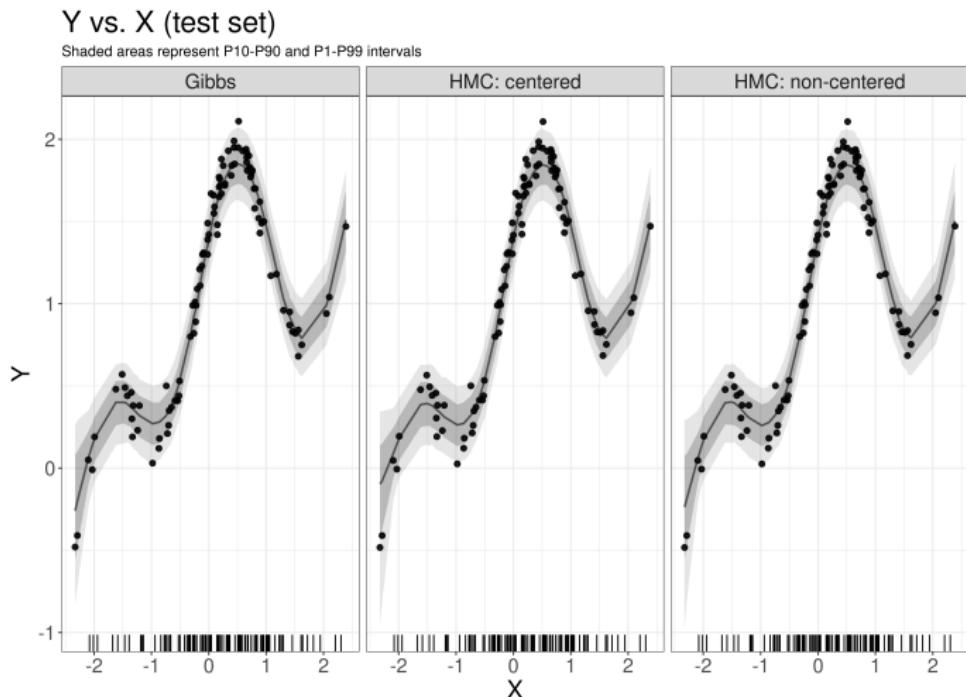


Figure 1: Predictive Quality

Prediction Median

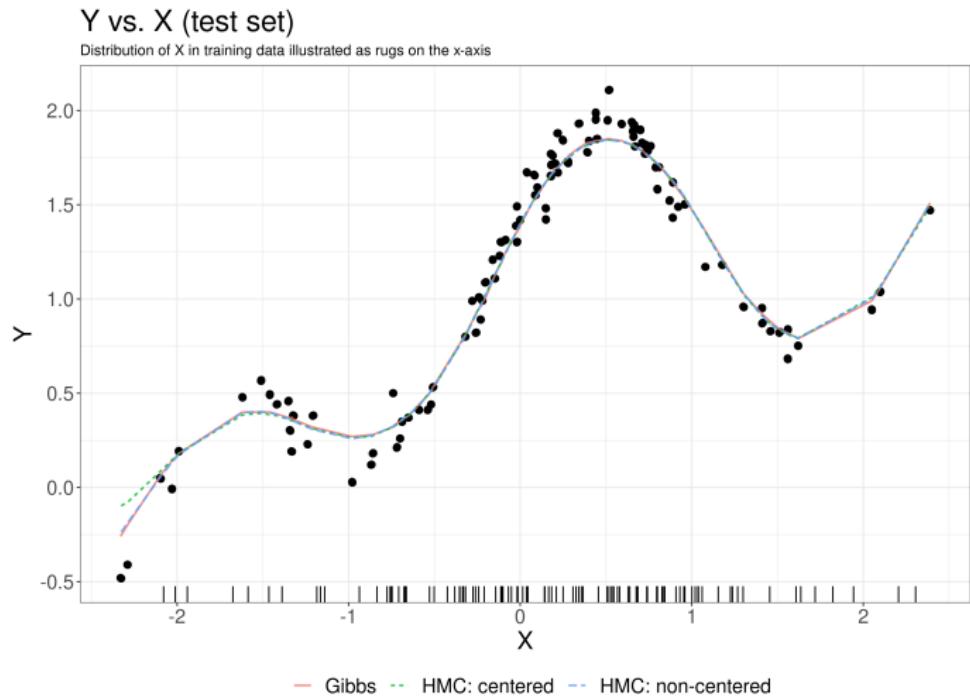


Figure 2: Predictive Median

Hyperparameters: Measurement Noise - Standard Deviation

Standard Deviation Hyperparameter: Target Noise

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

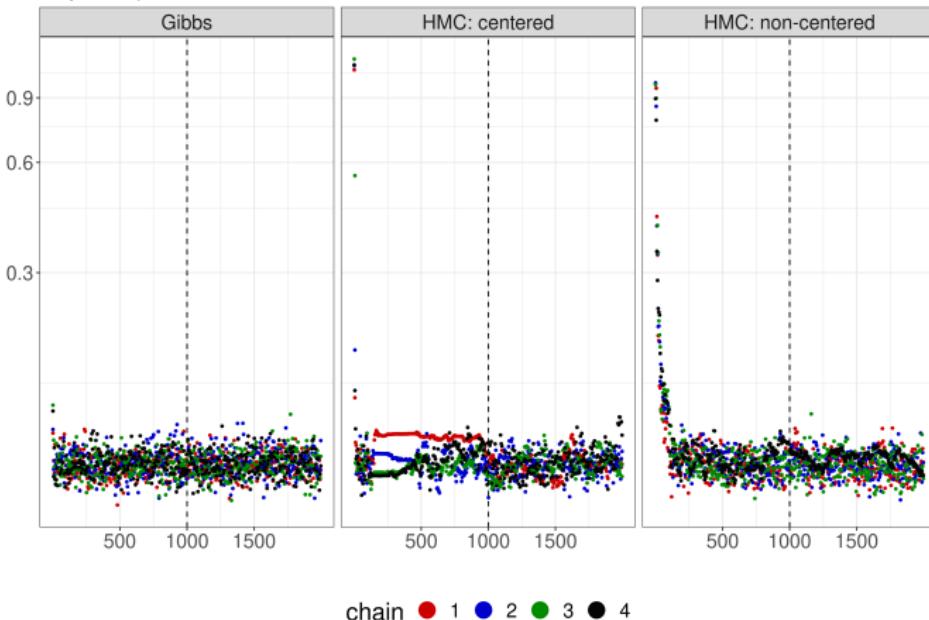


Figure 3: Target Noise Standard Deviation

Hyperparameters: Inputs-to-Hidden Weights - Standard Deviation

Standard Deviation Hyperparameter: Input-to-Hidden Weights

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

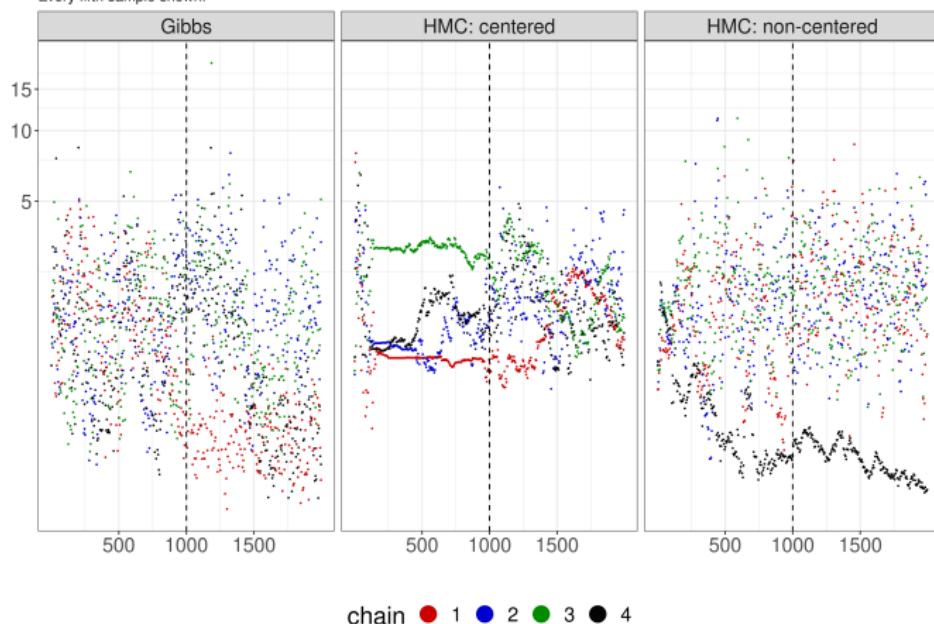


Figure 4: Inputs-to-Hidden Weights - Standard Deviation

Hyperparameters: Hidden-to-Output Weights - Standard Deviation

Standard Deviation Hyperparameter: Hidden-to-Output Weights

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison. Every fifth sample shown.

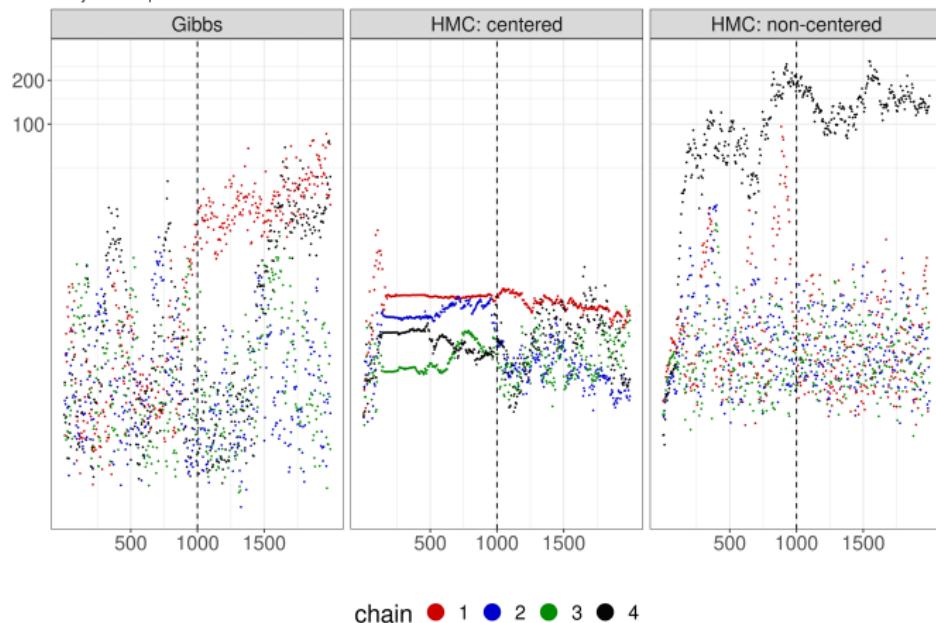


Figure 5: Hidden-to-Output Weights - Standard Deviation

Hyperparameters: Hidden Unit Biases - Standard Deviation

Standard Deviation Hyperparameter: Hidden Unit Biases

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

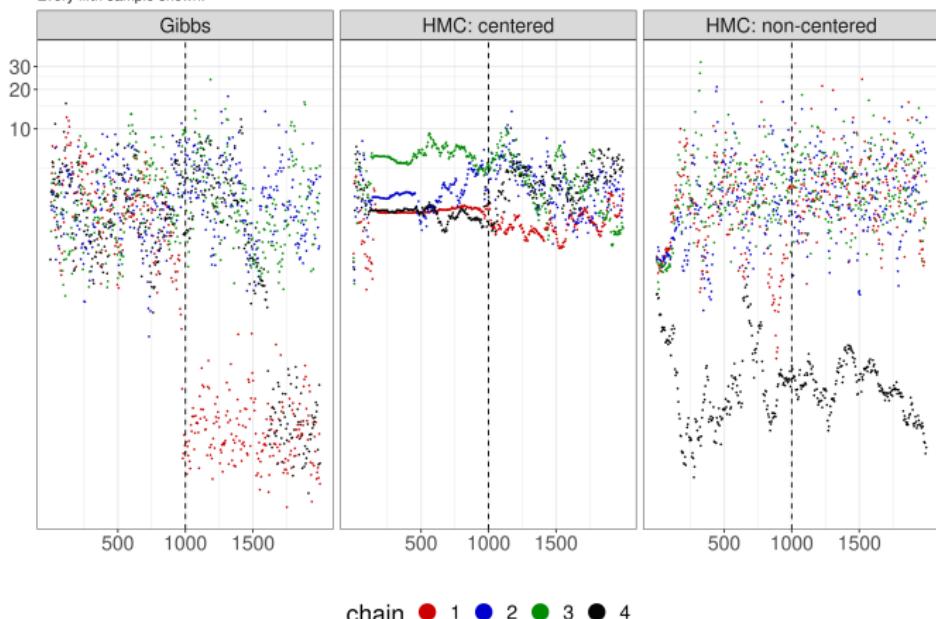


Figure 6: Hidden Unit Biases - Standard Deviation

Parameters: Inputs-to-Hidden Weights

Input-to-Hidden Weights

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

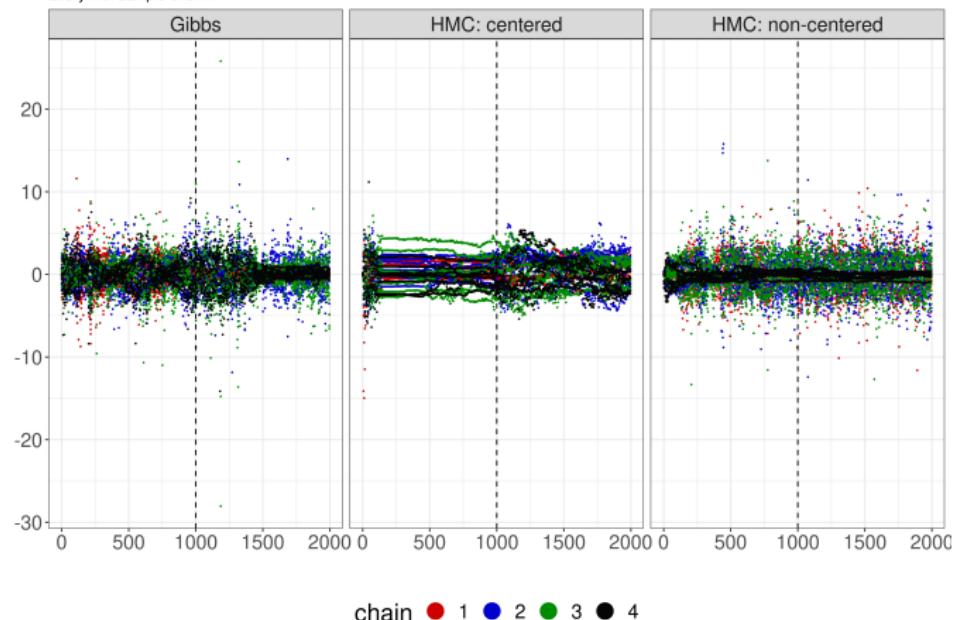


Figure 7: Inputs-to-Hidden Weights

Parameters: Hidden-to-Output Weights

Hidden-to-Output Unit Weights

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

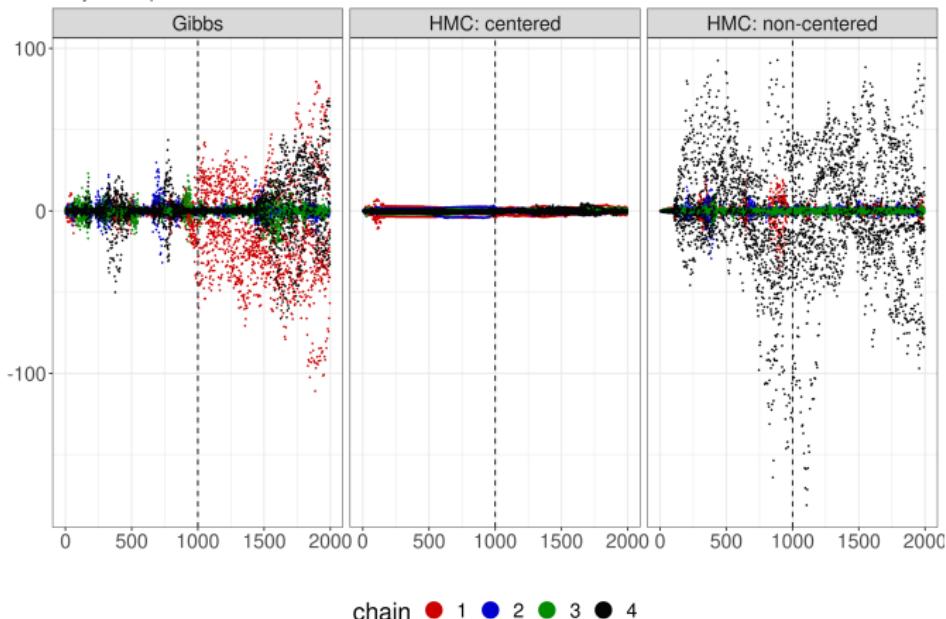


Figure 8: Hidden-to-Output Weights

Parameters: Hidden Units Biases

Hidden Unit Biases

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

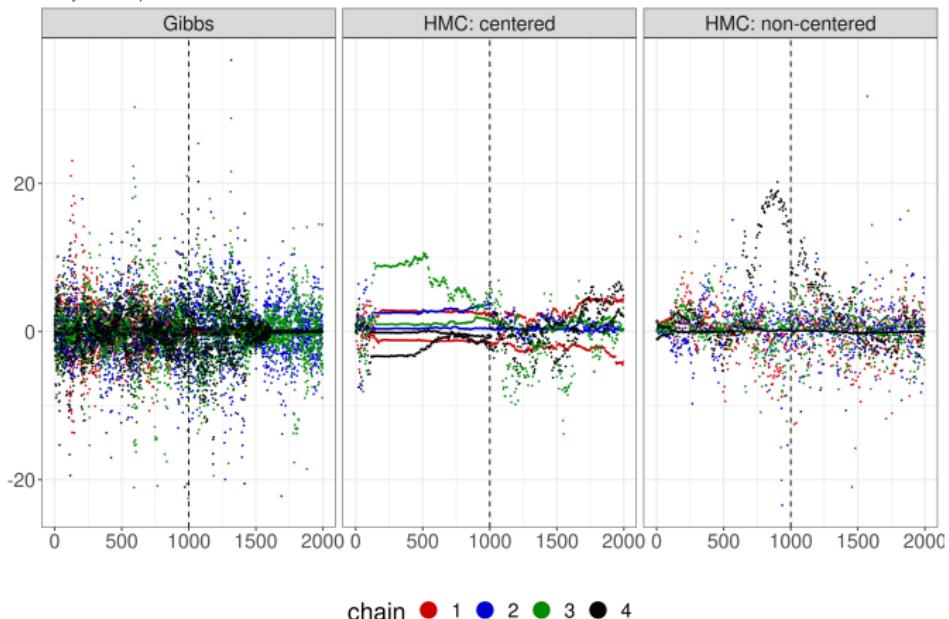


Figure 9: Hidden Units Biases

Parameters: Output Unit Bias

Output Unit Biases

Vertical line indicates starting point of values used as representative samples. Four chains are used for consistent comparison.
Every fifth sample shown.

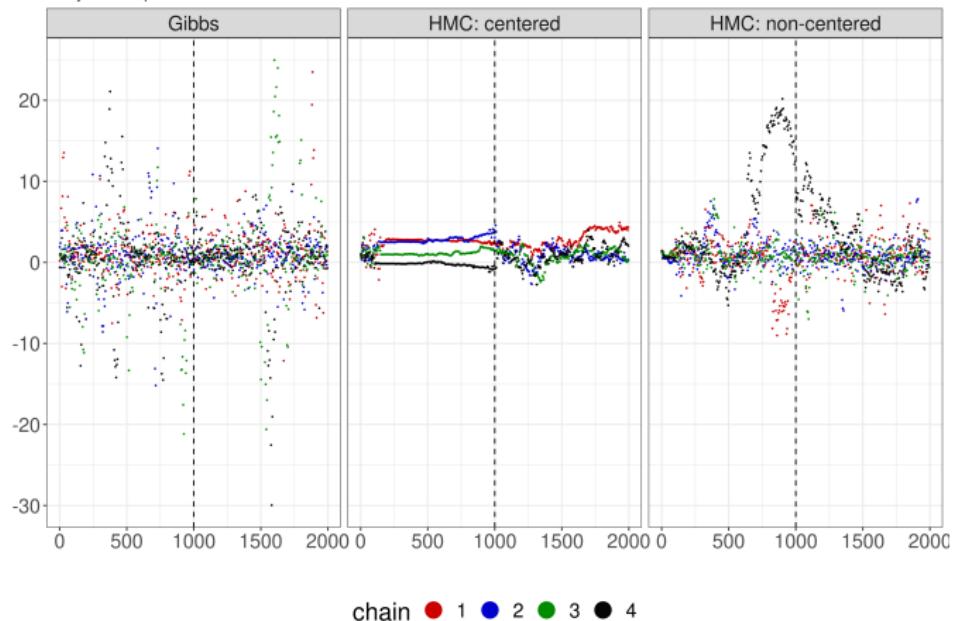


Figure 10: Output Unit Bias

Stepsize Comparison

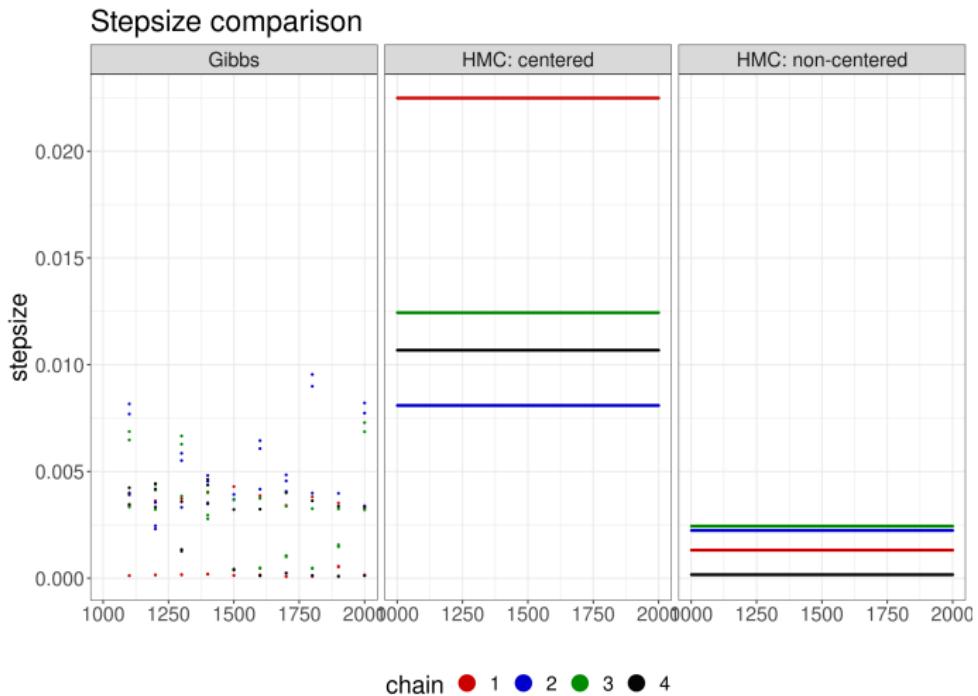


Figure 11: Step Size Comparison

Prior Checks / Conformance

Predictive Uncertainty

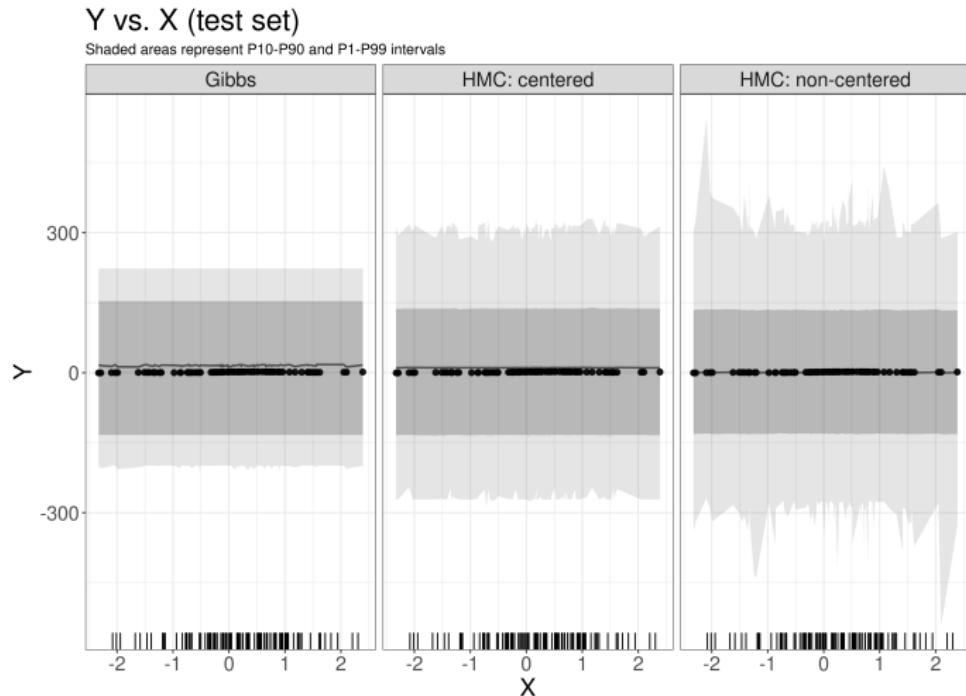


Figure 12: Predictive Quality

Hyperparameters: Measurement Noise - Standard Deviation

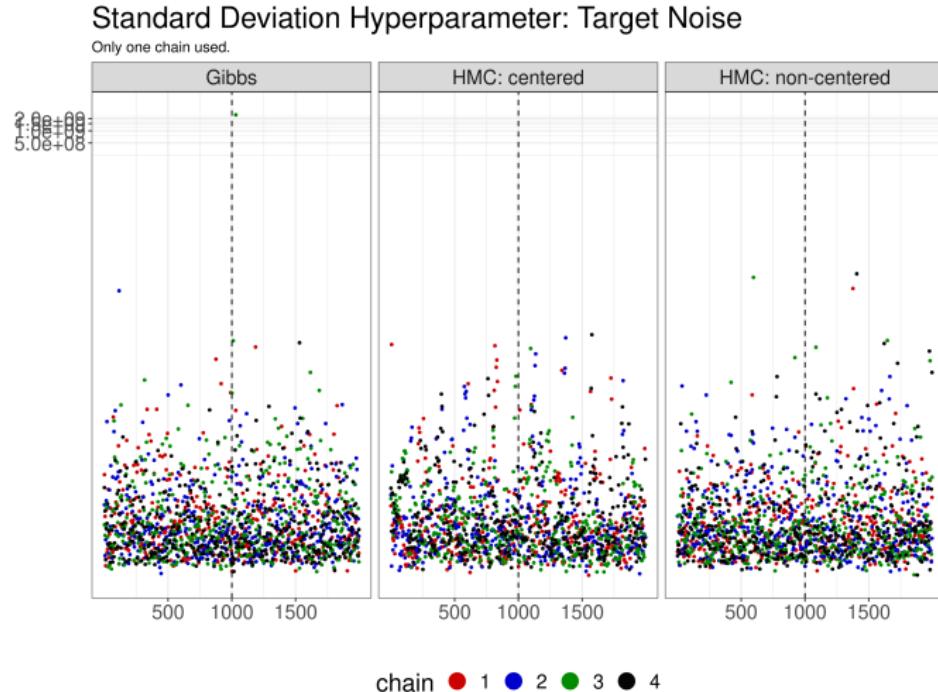


Figure 13: Target Noise Standard Deviation

Hyperparameters: Inputs-to-Hidden Weights - Standard Deviation

Standard Deviation Hyperparameter: Input-to-Hidden Weights

Only one chain used.

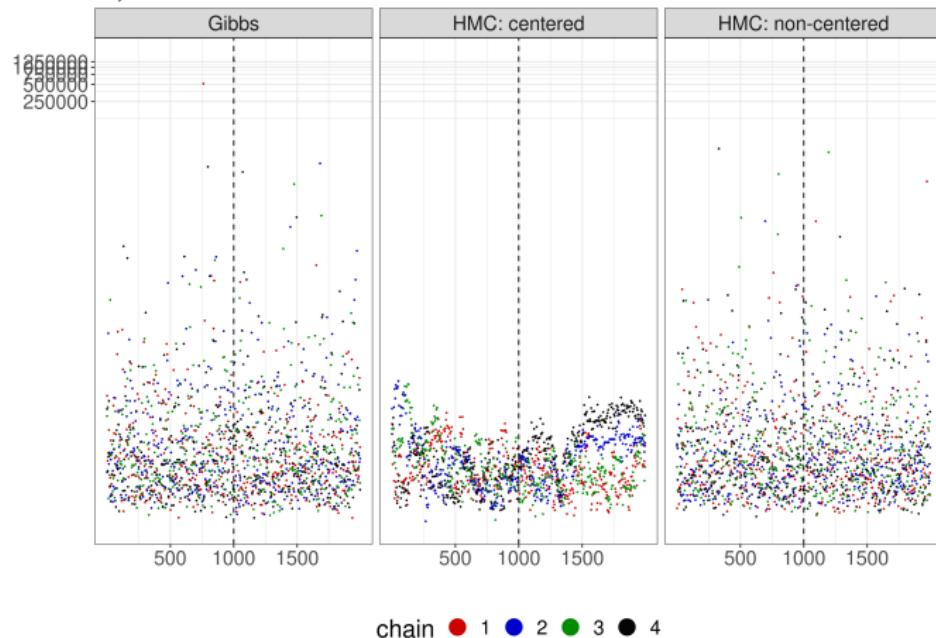


Figure 14: Inputs-to-Hidden Weights - Standard Deviation

Hyperparameters: Hidden-to-Output Weights - Standard Deviation

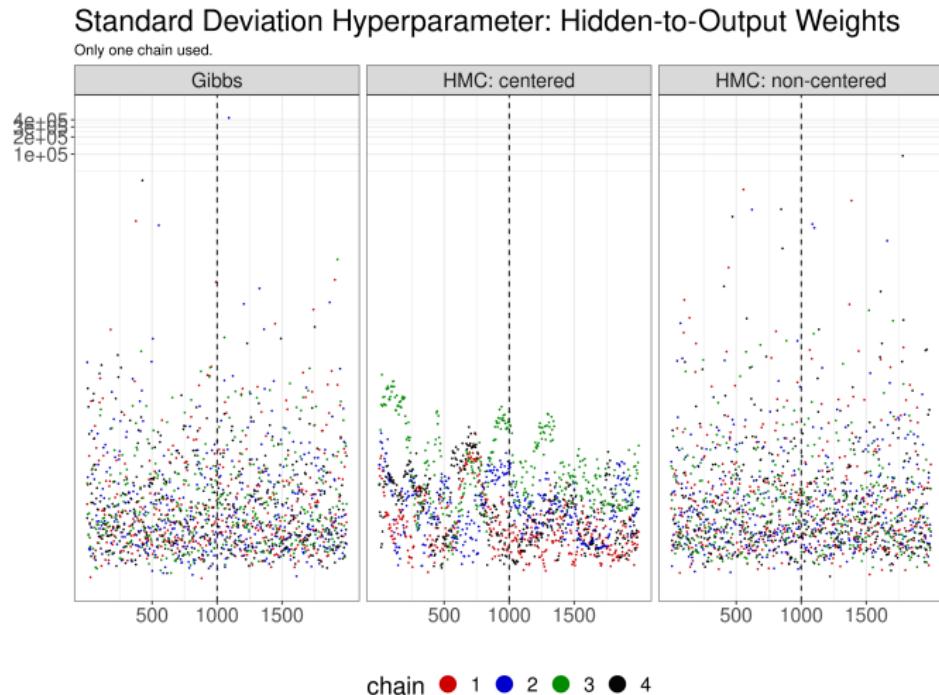


Figure 15: Hidden-to-Output Weights - Standard Deviation

Hyperparameters: Hidden Unit Biases - Standard Deviation

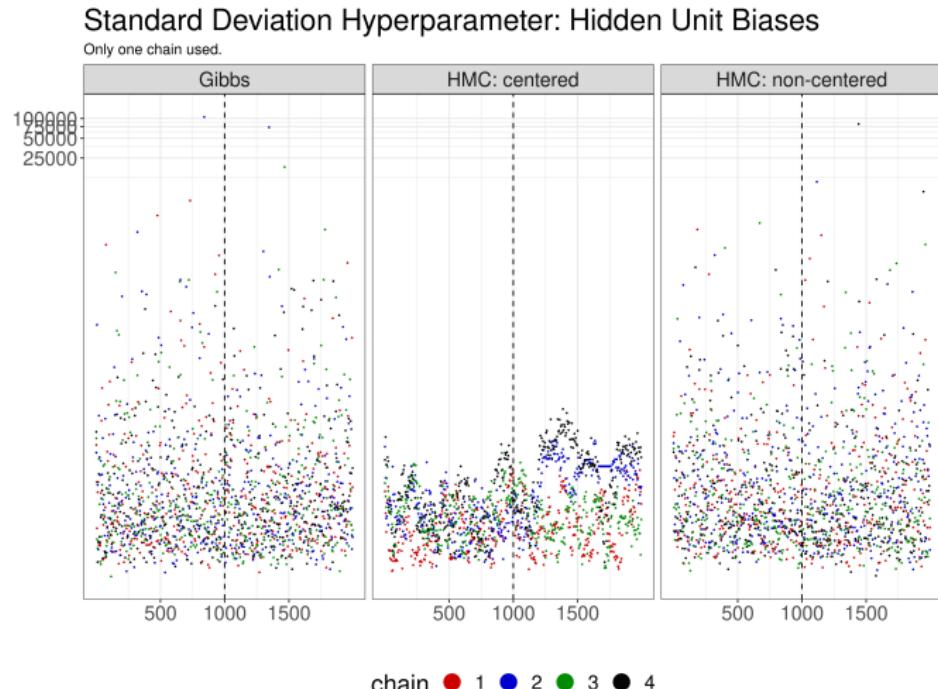


Figure 16: Hidden Unit Biases - Standard Deviation

Parameters: Inputs-to-Hidden Weights

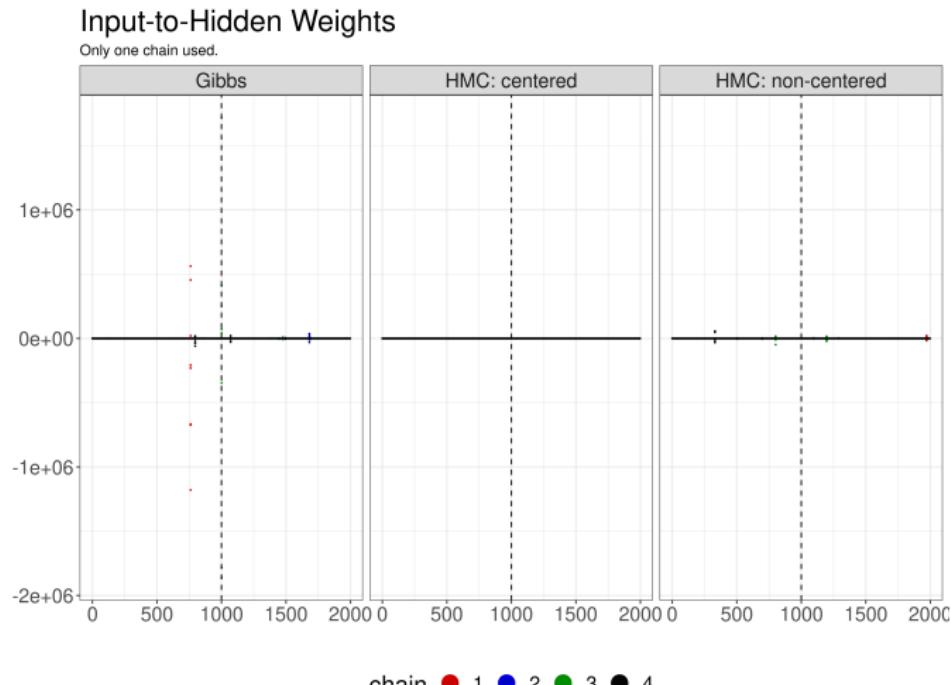


Figure 17: Inputs-to-Hidden Weights

Parameters: Hidden-to-Output Weights

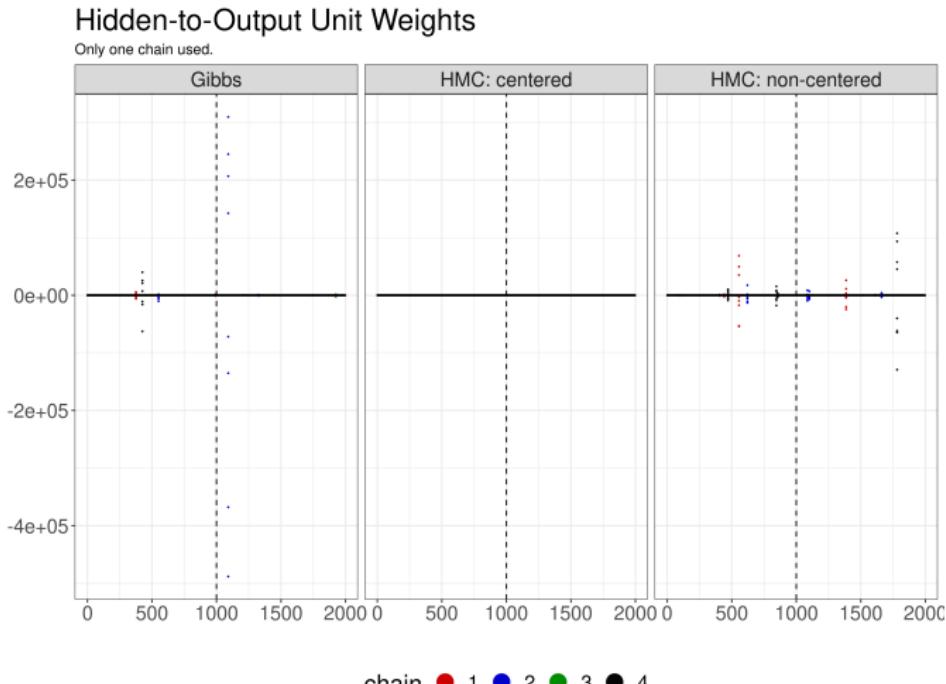


Figure 18: Hidden-to-Output Weights

Parameters: Hidden Units Biases

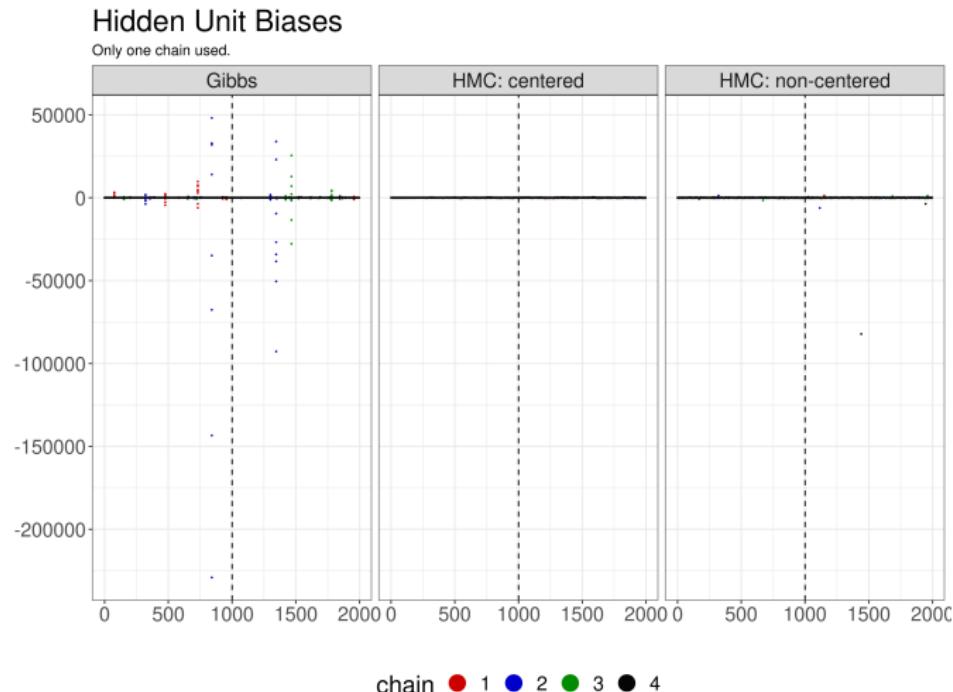


Figure 19: Hidden Units Biases

Parameters: Output Unit Bias

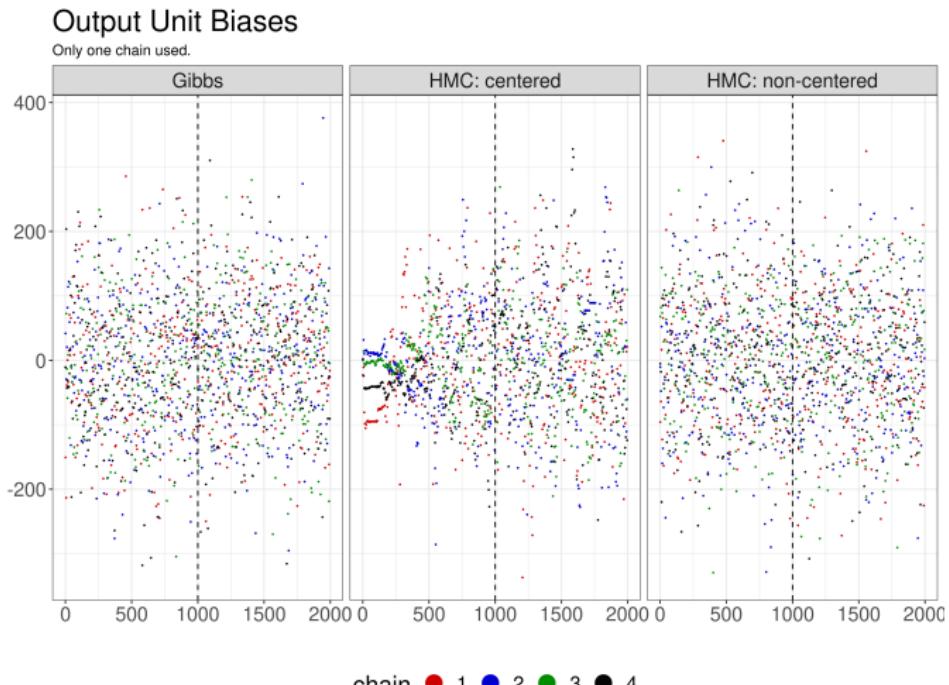


Figure 20: Output Unit Bias