

# Hyperparameter Sampling: Gibbs vs. NUTS

01/01/2021

## Objective

Experiment 1 (Vague Priors)

Experiment 2 (Less Vague Priors)

Experiment 3 (Less Vague Priors)

## 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

# Experiments Overview

- ▶ The experiments are conducted over **vague** and **less vague** priors for all groups of parameters

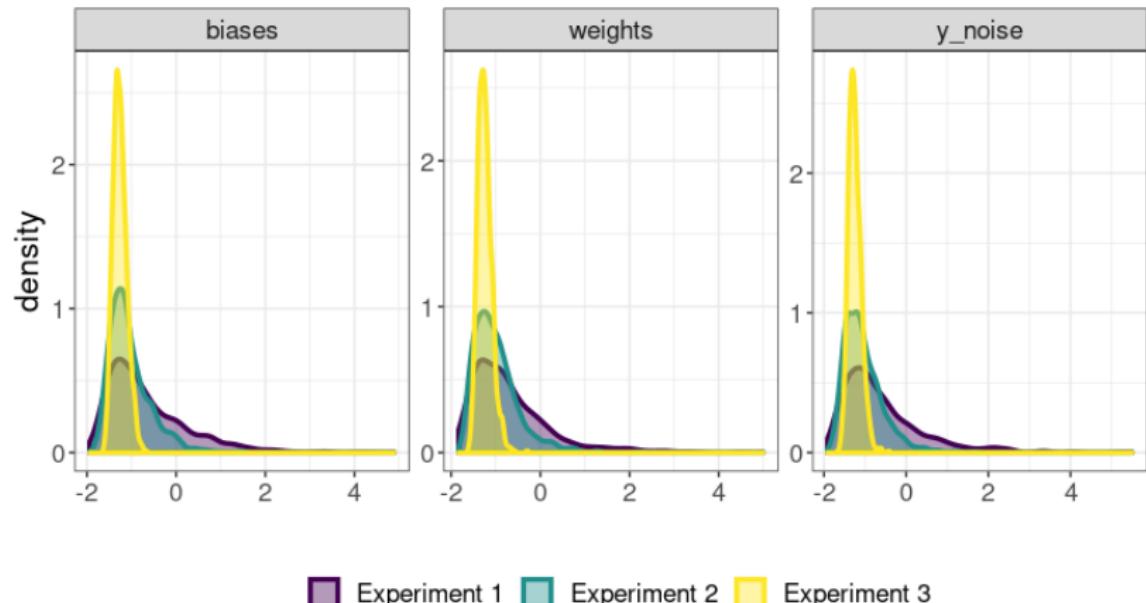


Figure 1: Prior Distributions

## Experiment 1 (Vague Priors)

# Assumptions

Architecture:

- ▶ 1 hidden layer, 8 hidden units
- ▶ tanh activation
- ▶ *parameter groups*: input-hidden weights, output-hidden weights, hidden biases, output bias

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 biases hyperparameter: 0.05:0.5
- ▶ output bias hyperparameter: 0.05:0.5
- ▶ target noise: 0.05:0.5

# NUTS - Weight Traces

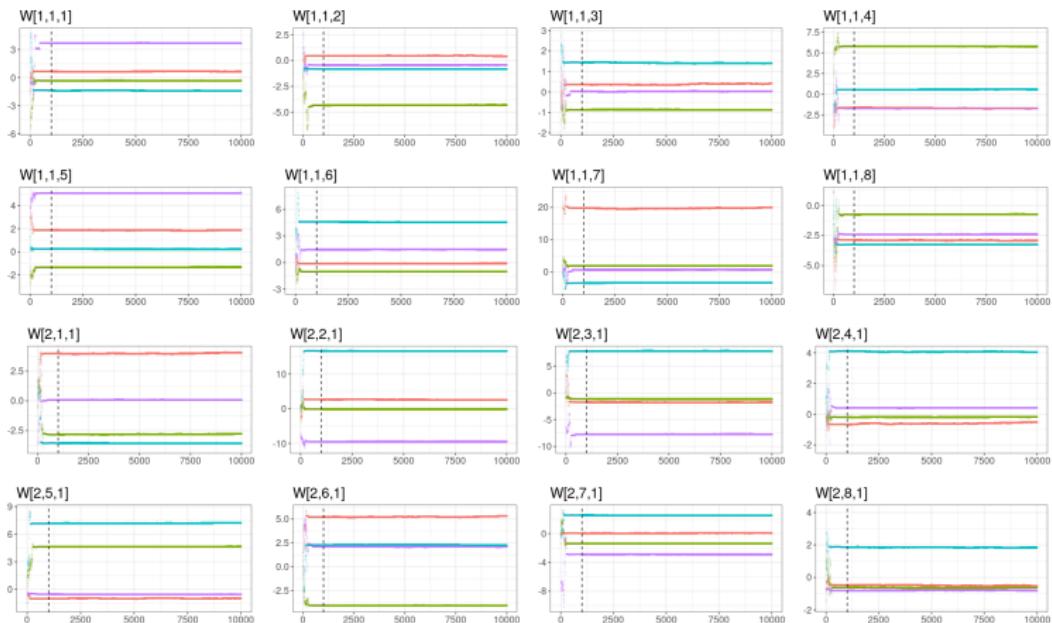


Figure 2: Weight Traces

# NUTS - Hyperparameter Traces

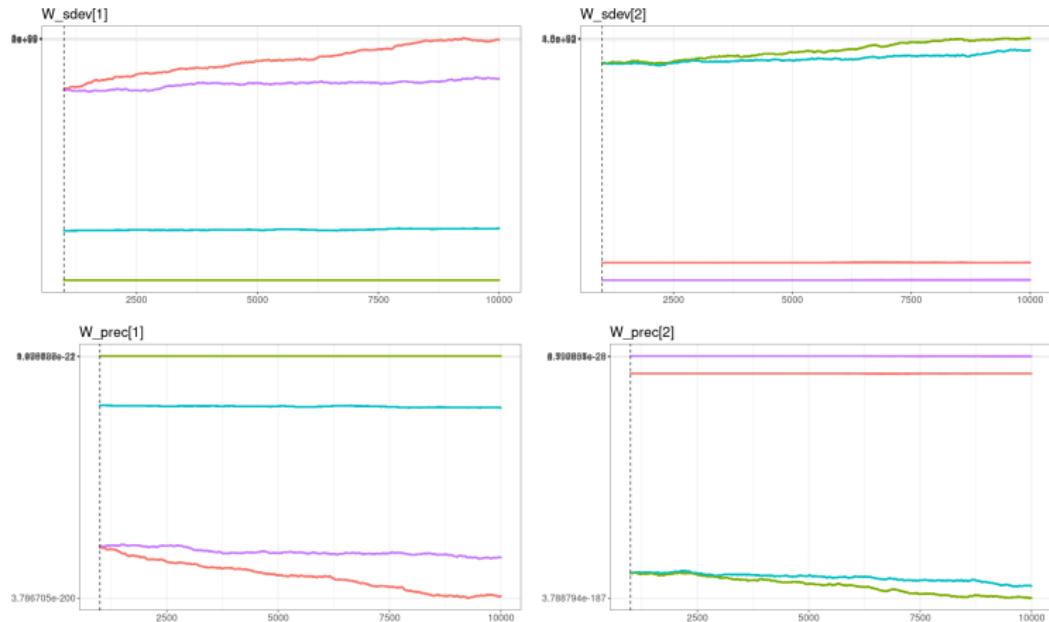


Figure 3: Hyperparameter Traces

# NUTS - Test Set Predictions

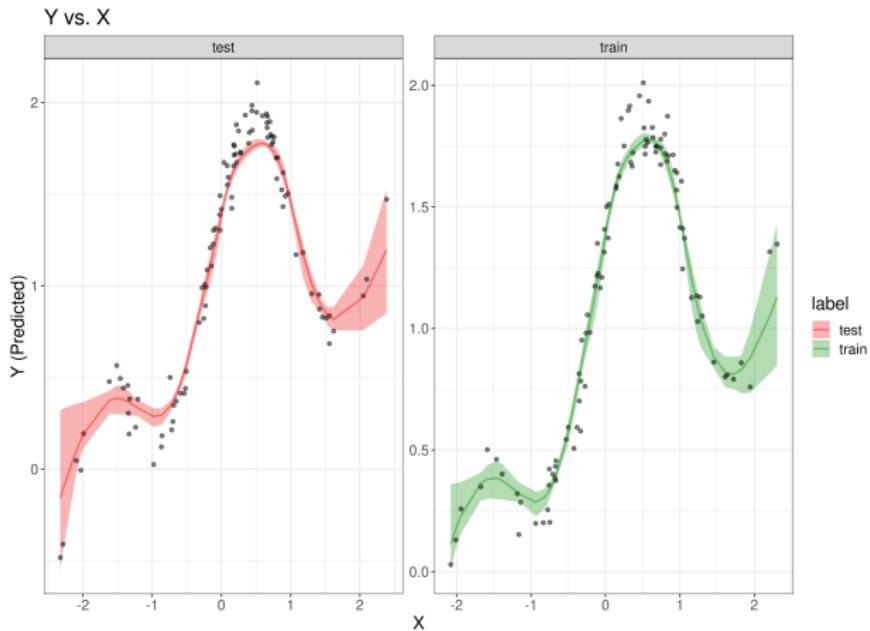


Figure 4: Predictive Quality

# NUTS - Chain statistics

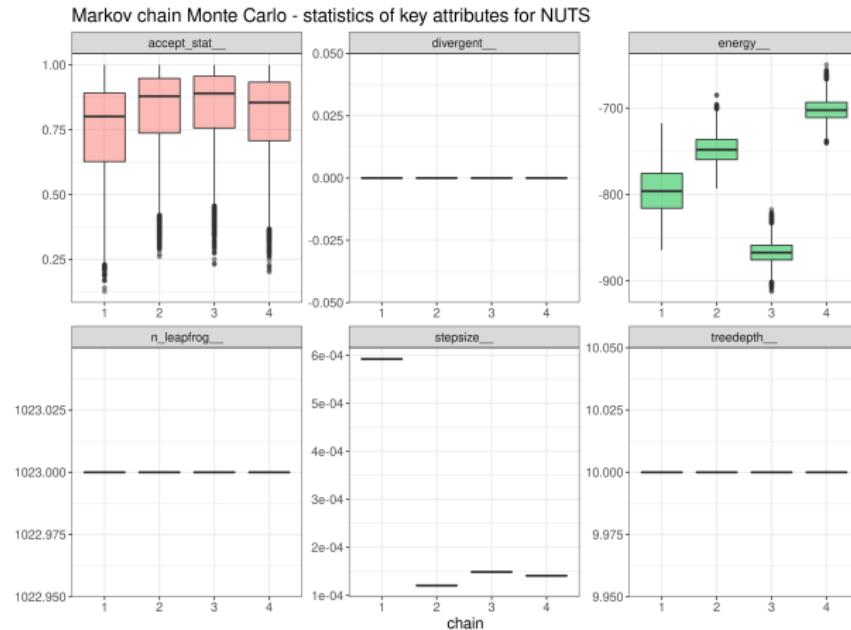


Figure 5: Chain Statistics

# FBM - Weight Traces

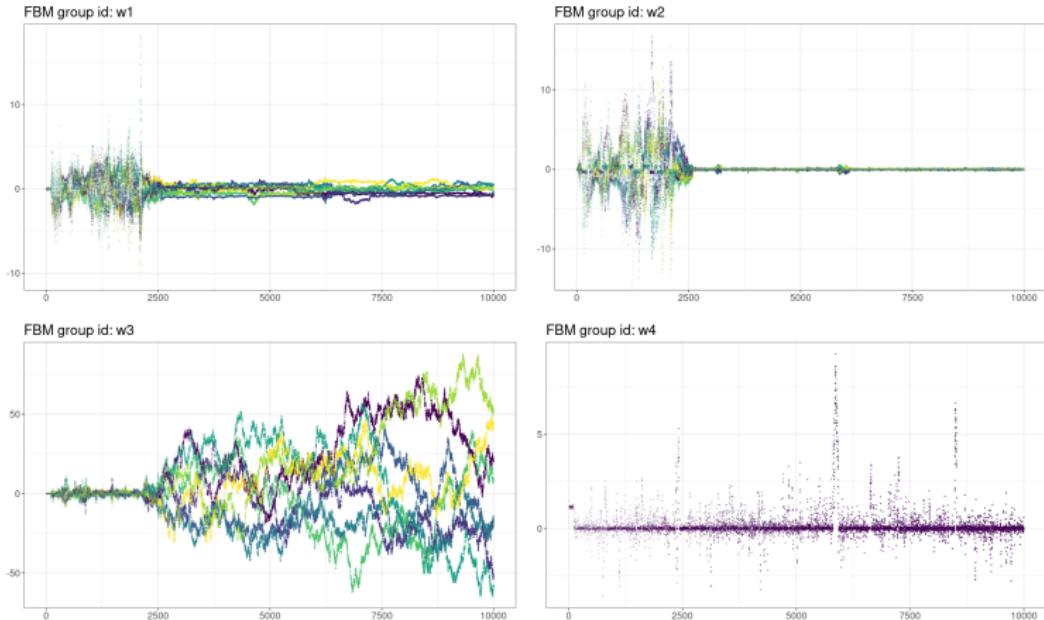


Figure 6: Weight Traces

# FBM - Hyperparameter Traces

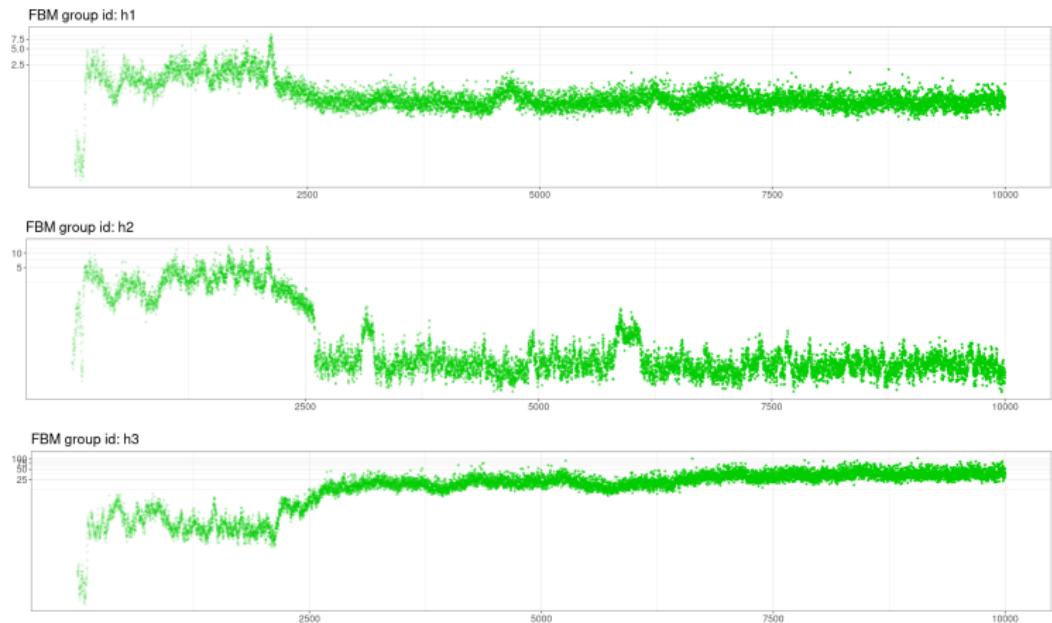


Figure 7: Hyperparameter Traces

## FBM - Test Set Predictions

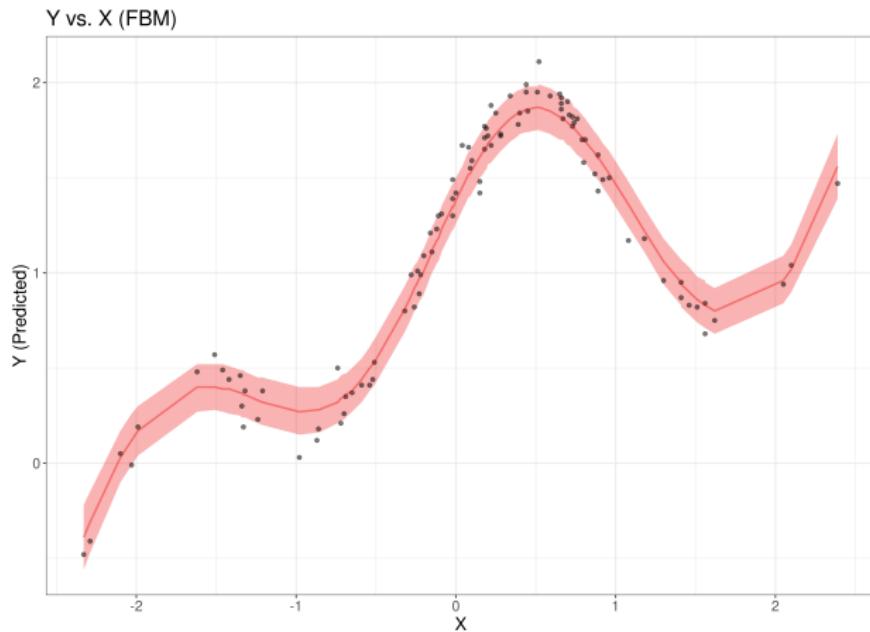


Figure 8: Predictive Quality

## Experiment 2 (Less Vague Priors)

# Assumptions

Architecture:

- ▶ 1 hidden layer, 8 hidden units
- ▶ tanh activation
- ▶ *parameter groups*: input-hidden weights, output-hidden weights, hidden biases, output bias

Data:

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

Hyperparameter Priors (FBM Notation):

- ▶ input-hidden weights hyperparameter: 0.05:1
- ▶ hidden-output weights hyperparameter: 0.05:1
- ▶ hidden biases hyperparameter: 0.05:1
- ▶ output bias hyperparameter: 0.05:1
- ▶ target noise: 0.05:1

# NUTS - Weight Traces

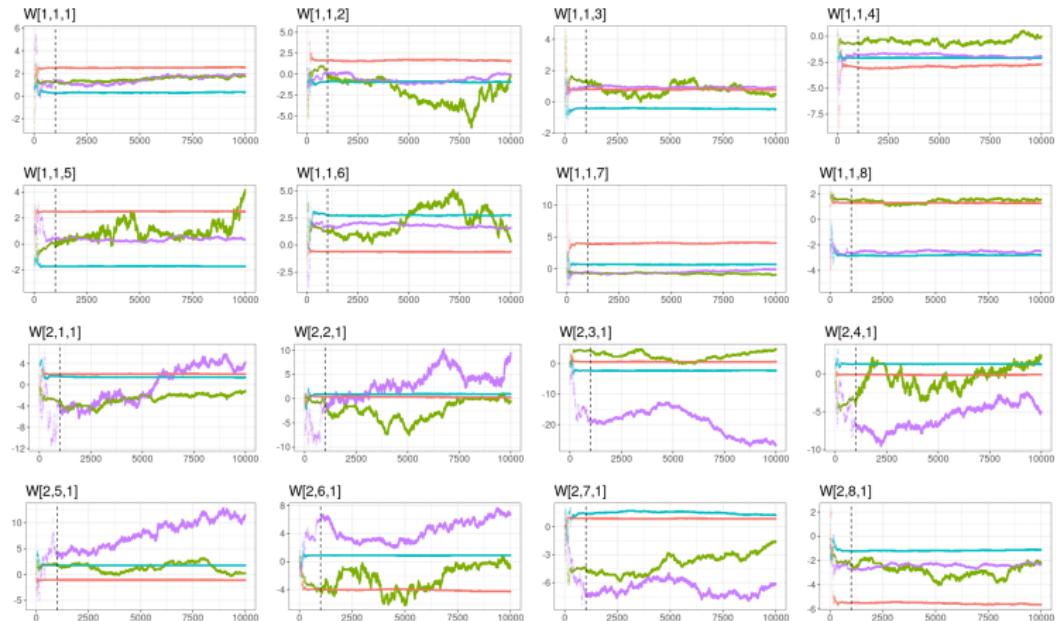


Figure 9: Weight Traces

# NUTS - Hyperparameter Traces

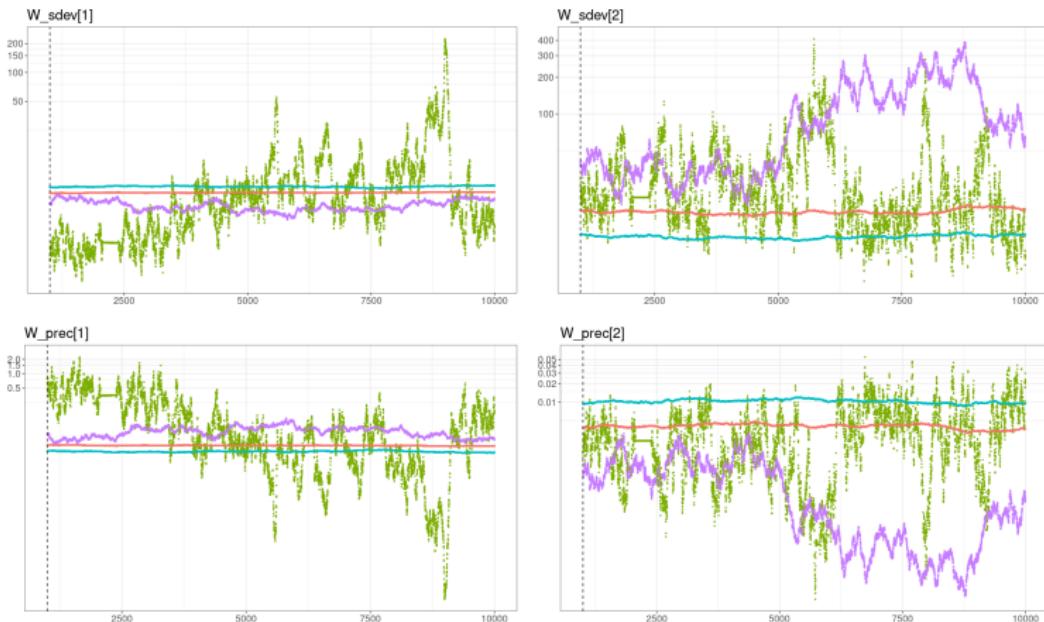


Figure 10: Hyperparameter Traces

# NUTS - Test Set Predictions

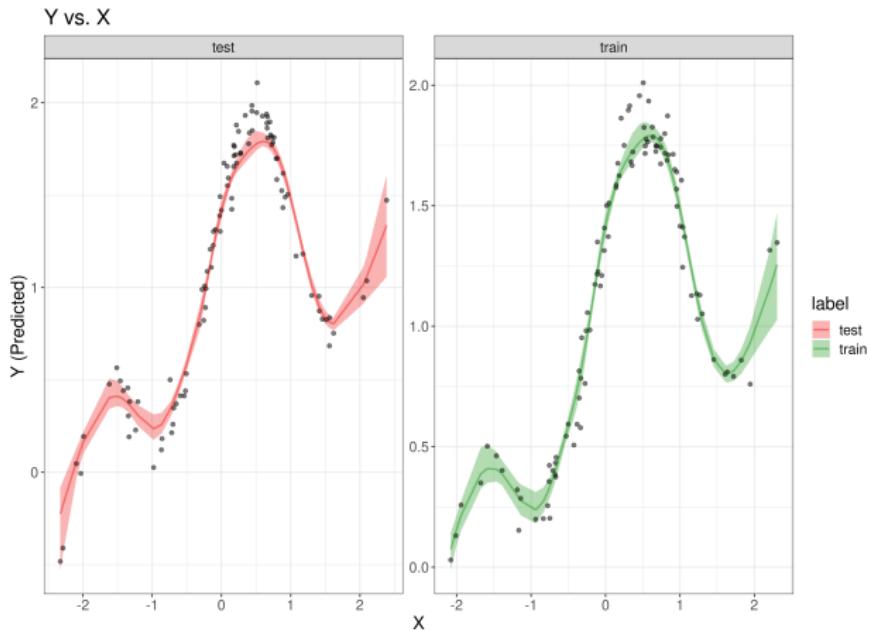


Figure 11: Predictive Quality

# NUTS - Chain statistics

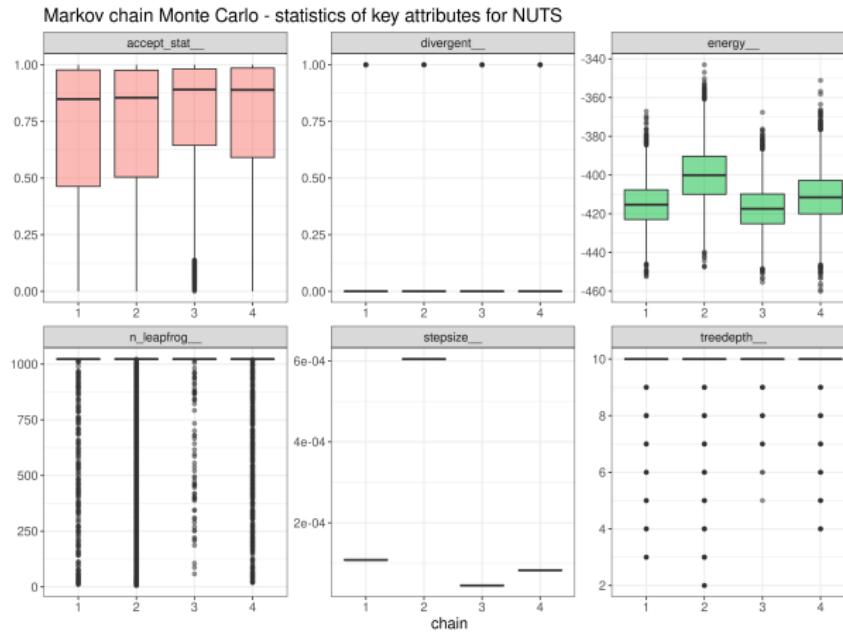


Figure 12: Chain Statistics

# FBM - Weight Traces

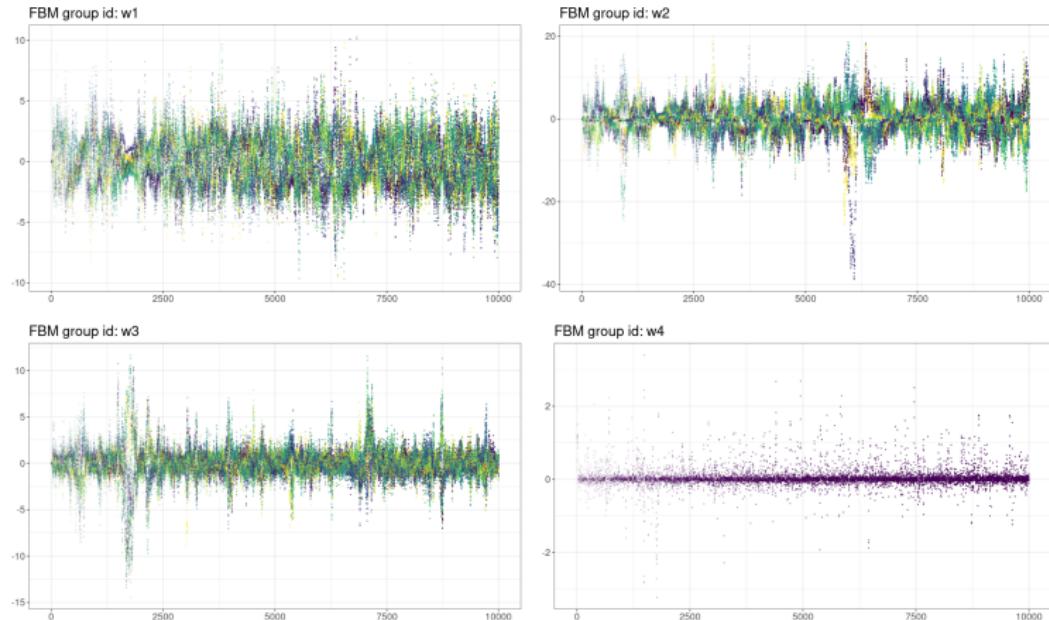


Figure 13: Weight Traces

# FBM - Hyperparameter Traces

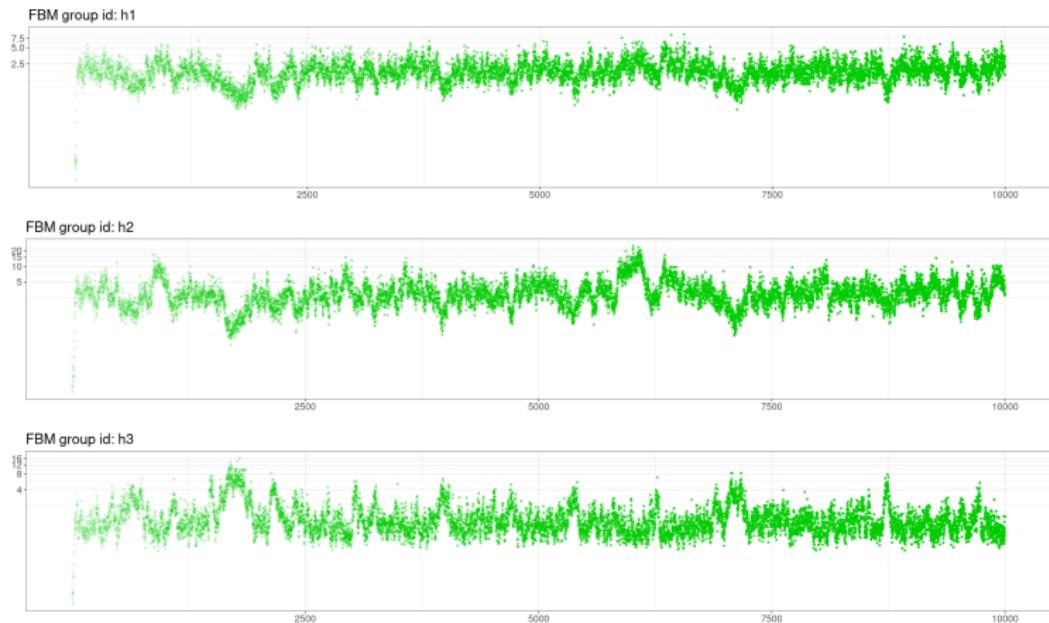


Figure 14: Hyperparameter Traces

## FBM - Test Set Predictions

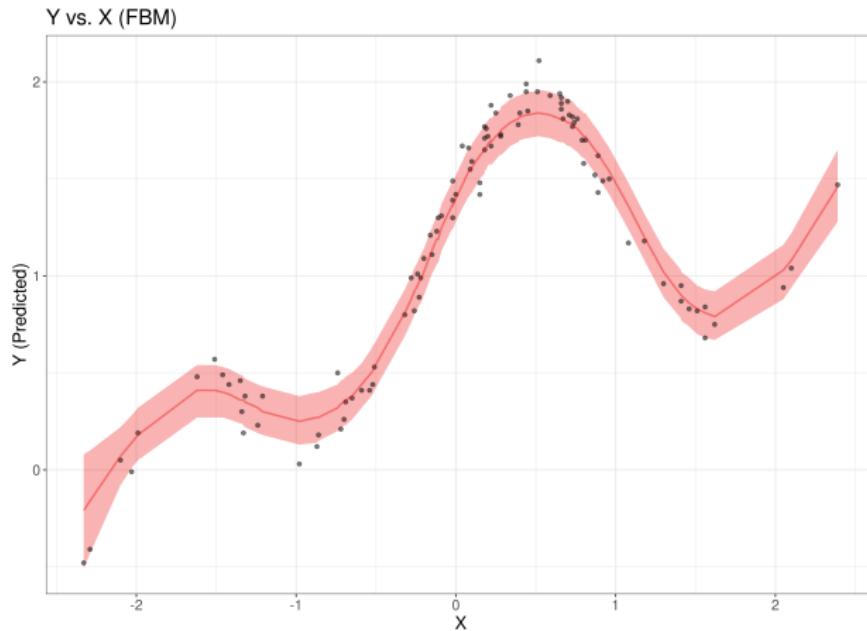


Figure 15: Predictive Quality

## Experiment 3 (Less Vague Priors)

# Assumptions

Architecture:

- ▶ 1 hidden layer, 8 hidden units
- ▶ tanh activation
- ▶ *parameter groups*: input-hidden weights, output-hidden weights, hidden biases, output bias

Data:

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

Hyperparameter Priors (FBM Notation):

- ▶ input-hidden weights hyperparameter: 0.05:5
- ▶ hidden-output weights hyperparameter: 0.05:5
- ▶ hidden biases hyperparameter: 0.05:5
- ▶ output bias hyperparameter: 0.05:5
- ▶ target noise: 0.05:5

# NUTS - Weight Traces

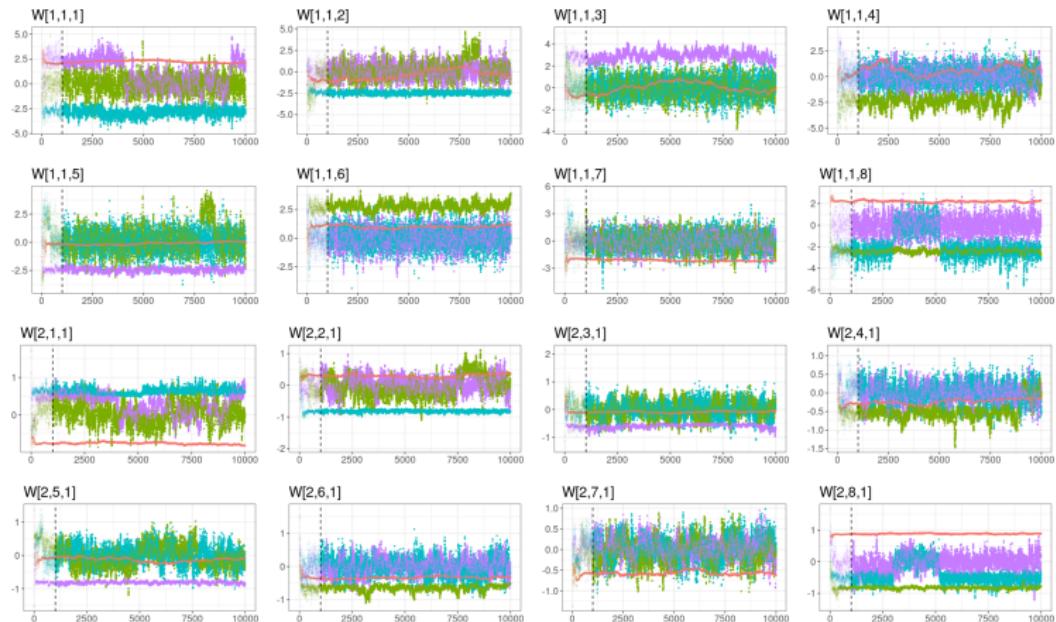


Figure 16: Weight Traces

# NUTS - Hyperparameter Traces

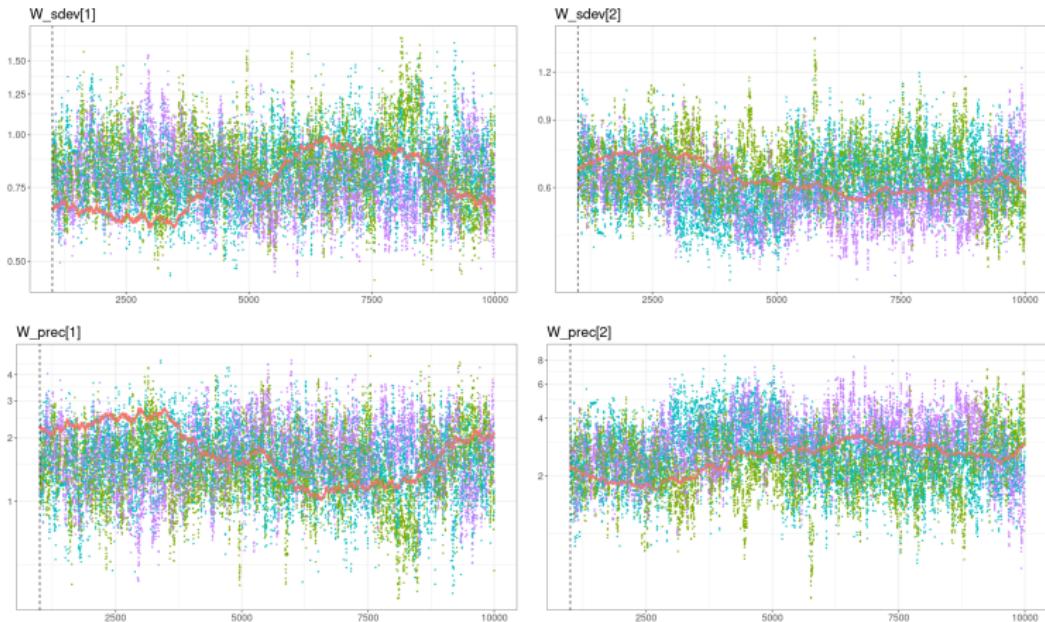


Figure 17: Hyperparameter Traces

# NUTS - Test Set Predictions

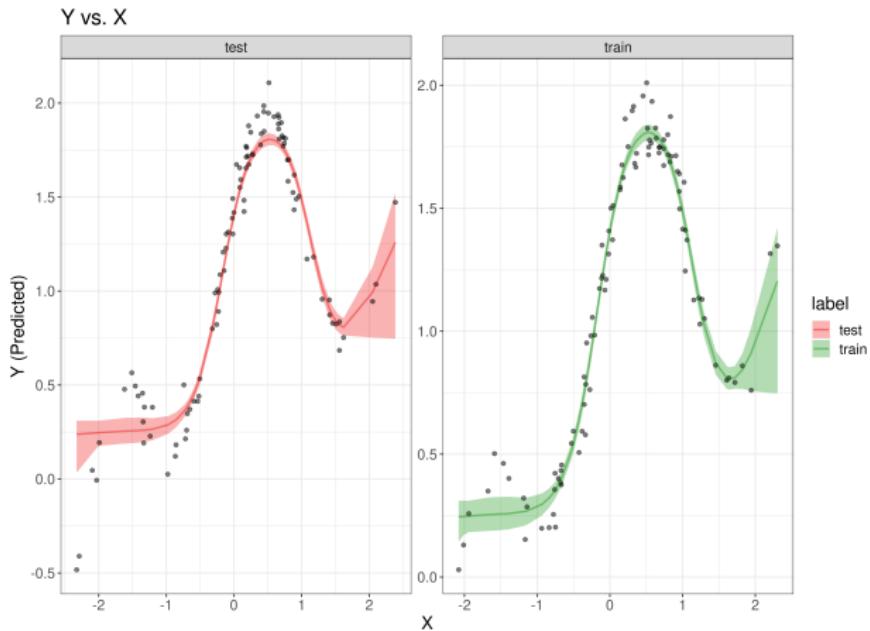


Figure 18: Predictive Quality

# NUTS - Chain statistics

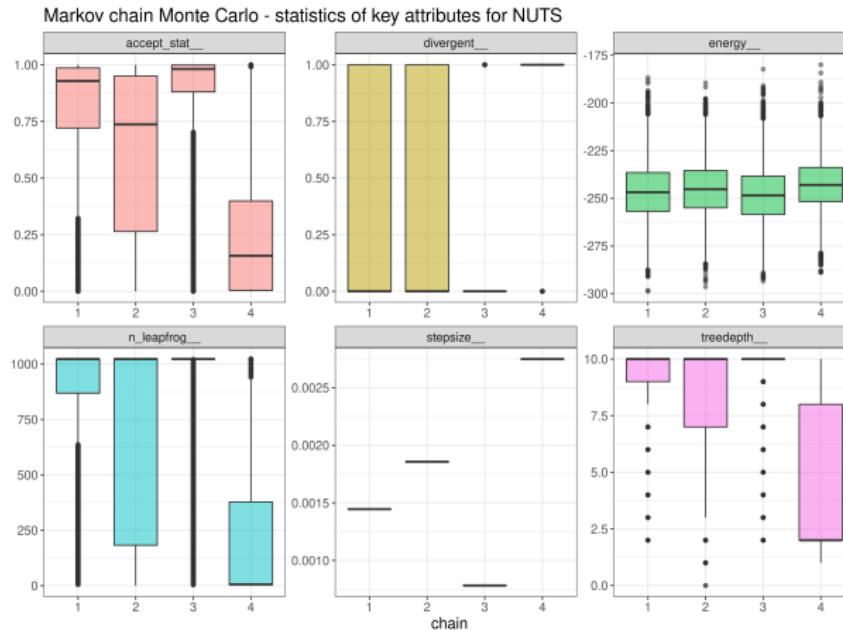
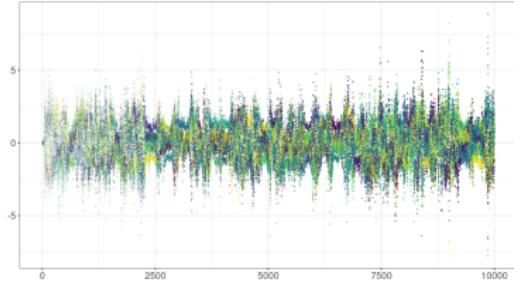


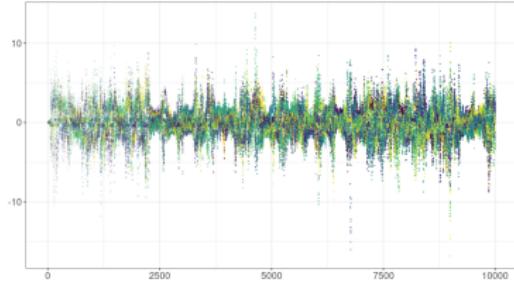
Figure 19: Chain Statistics

# FBM - Weight Traces

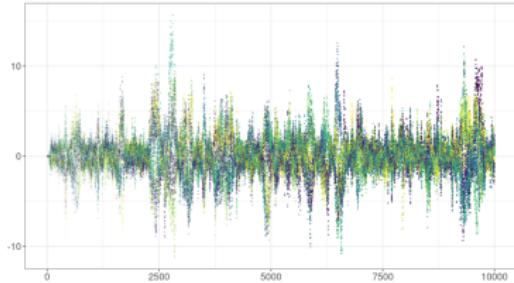
FBM group id: w1



FBM group id: w2



FBM group id: w3



FBM group id: w4

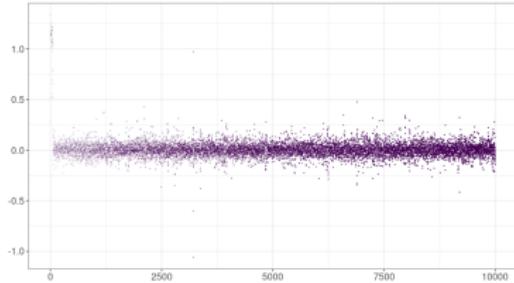


Figure 20: Weight Traces

# FBM - Hyperparameter Traces

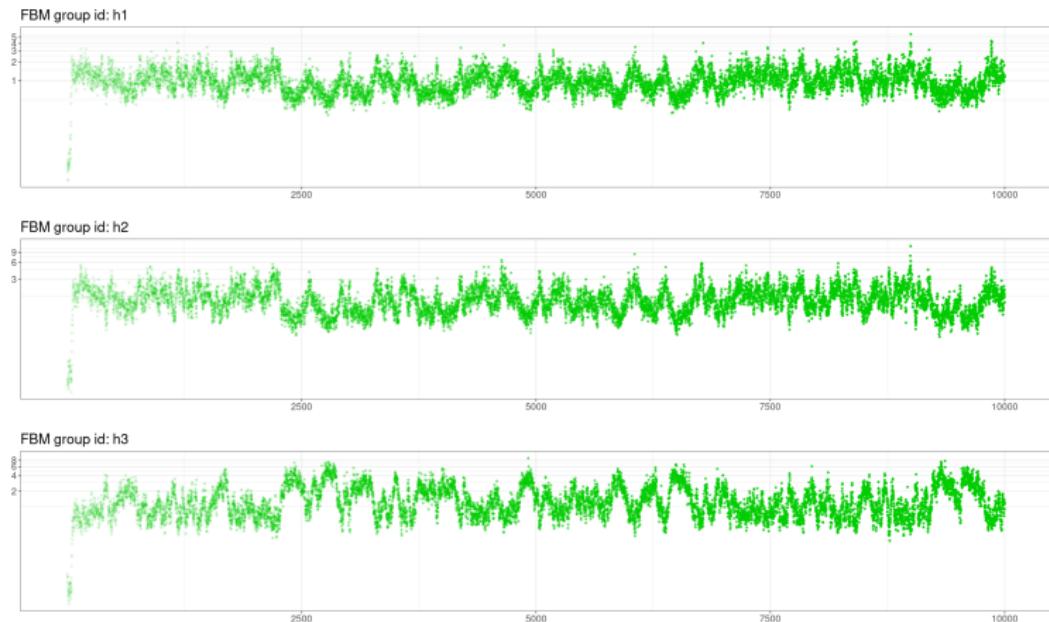


Figure 21: Hyperparameter Traces

## FBM - Test Set Predictions

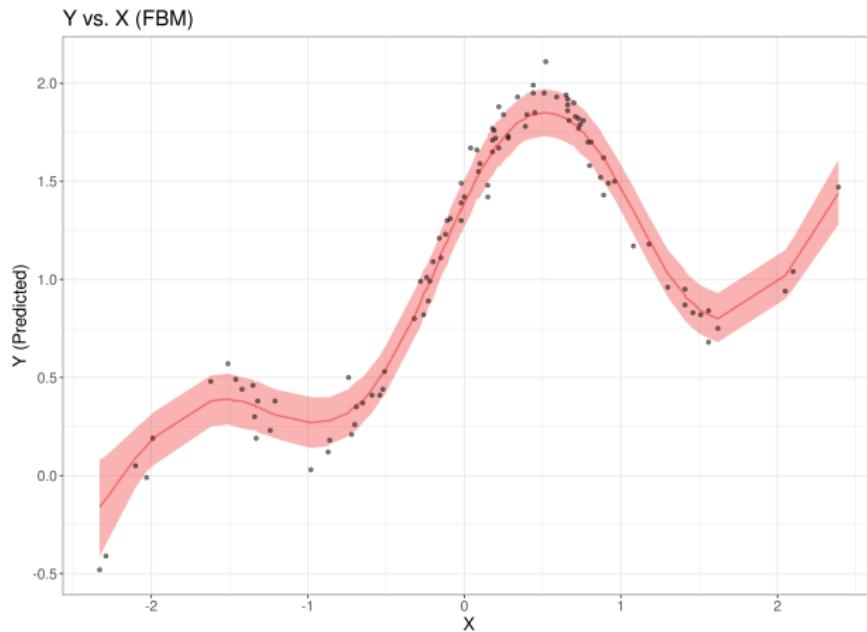


Figure 22: Predictive Quality