BNN Hyperparameters Integration: Gibbs vs. NUTS

01/01/2021

Overview

Overview

Three experiments:

- NUTS (Centered Parametrization)
- NUTS (Non-Centered Parametrization)
- FBM (Gibbs sampling for hyperparameters, HMC)

Centered vs. Non-Centered Parametrization

- Centered:

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

- Non-Centered:

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

- 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 weights for better limiting behavior

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
- target noise: 0.05:0.5

NUTS-specific Assumptions

Tuning Parameters:

- `adapt_delta`: 0.9 (default = 0.8)
- `max_treedepth`: 11 (default = 10)
- `adapt_gamma`: 0.01 (default = 0.05)

Initial Values:

- `Weight Precision`: 1
- `Biases Precision`: 1
- `Target Noise Precision`: 100
- `Weights`: Uniform(-5e-6, +5e-6)
- `Biases`: Uniform(-5e-3, +5e-3)

Note that for both centered and non-centered parametrization, the exact same initial values were used.

NUTS (Centered Parametrization)

NUTS (Centered) - Test Set Predictions

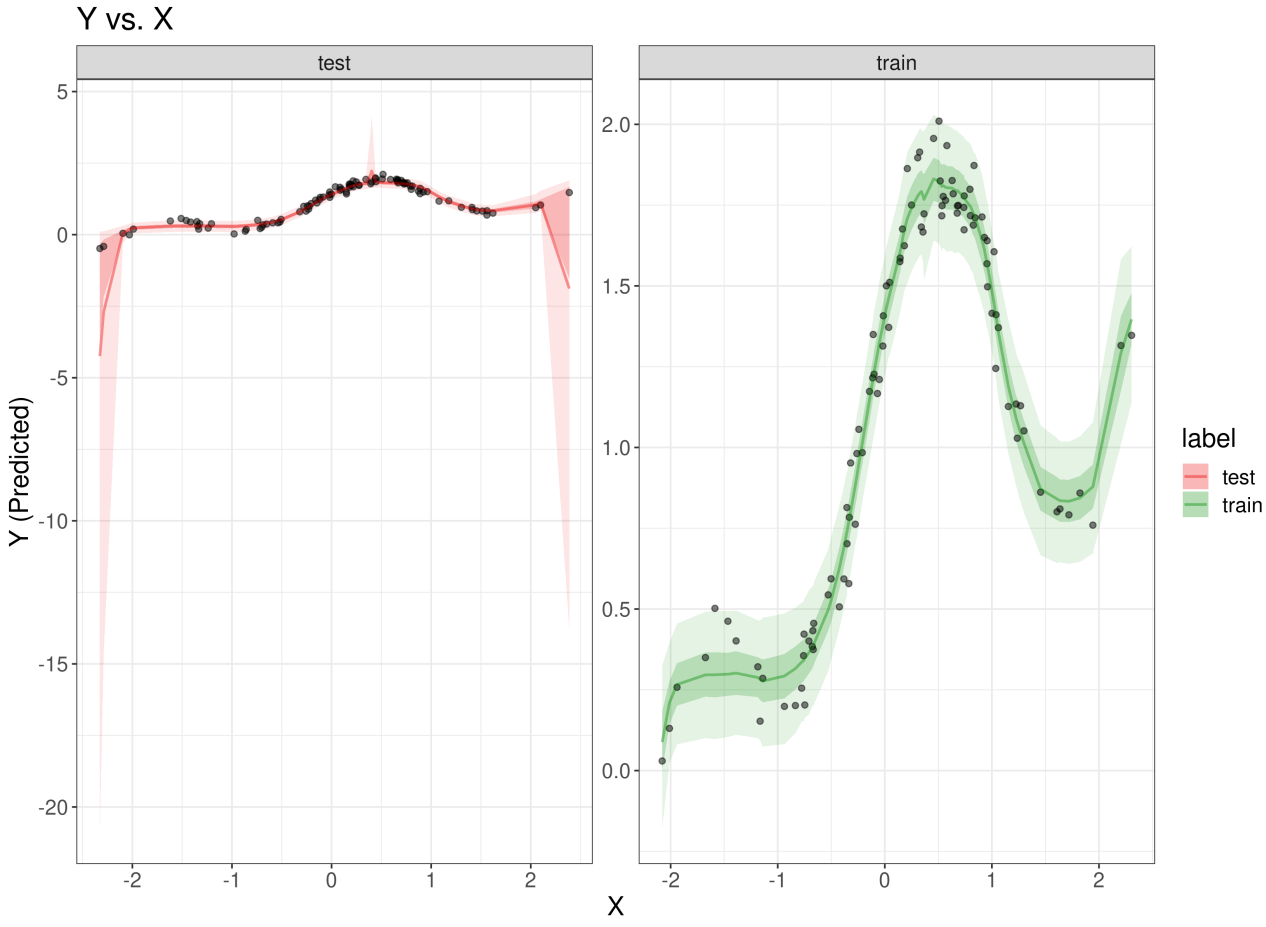

Figure 1: Predictive Quality

NUTS (Centered) - Weight Traces

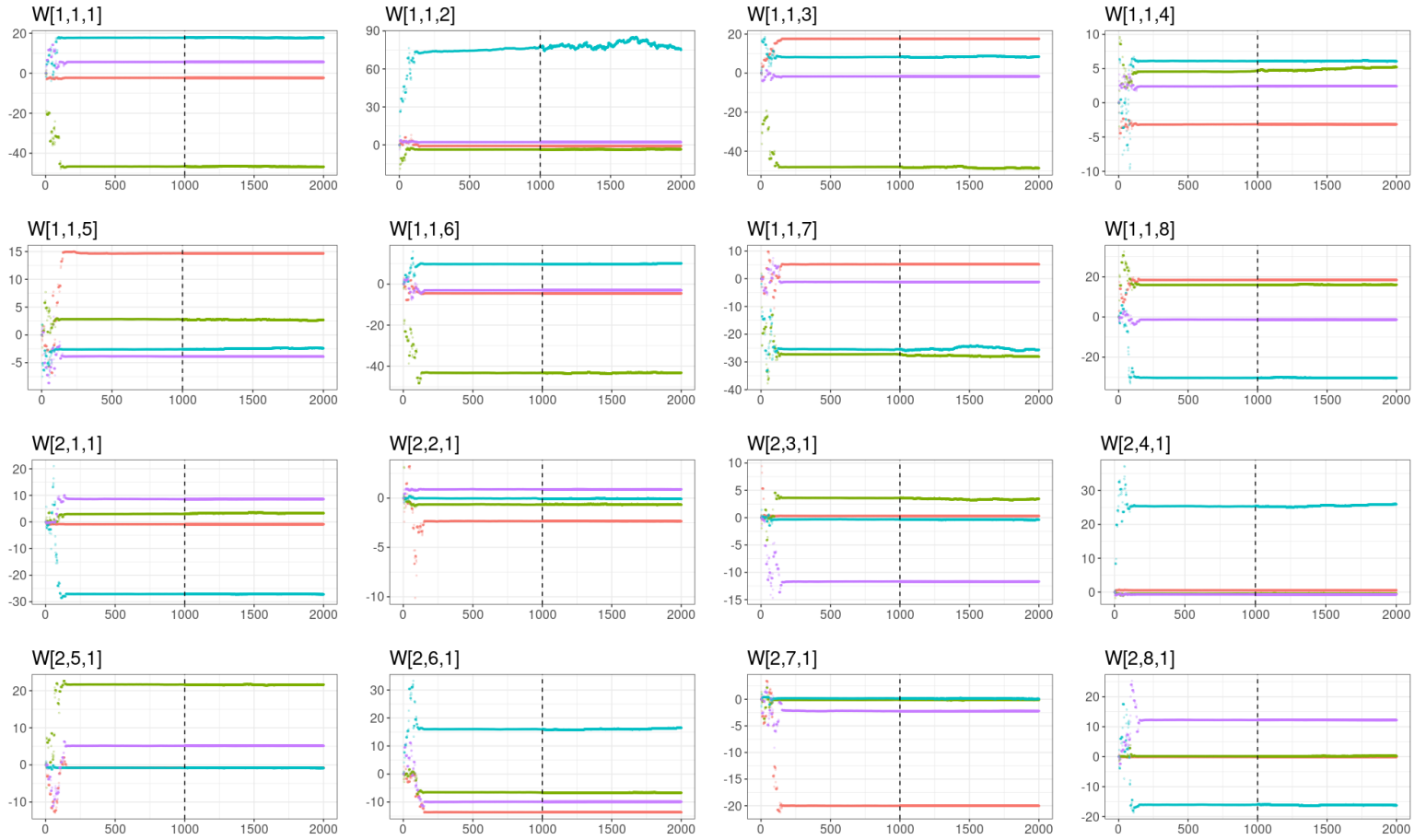

Figure 2: Weight Traces

NUTS (Centered) - Hyperparameter Traces

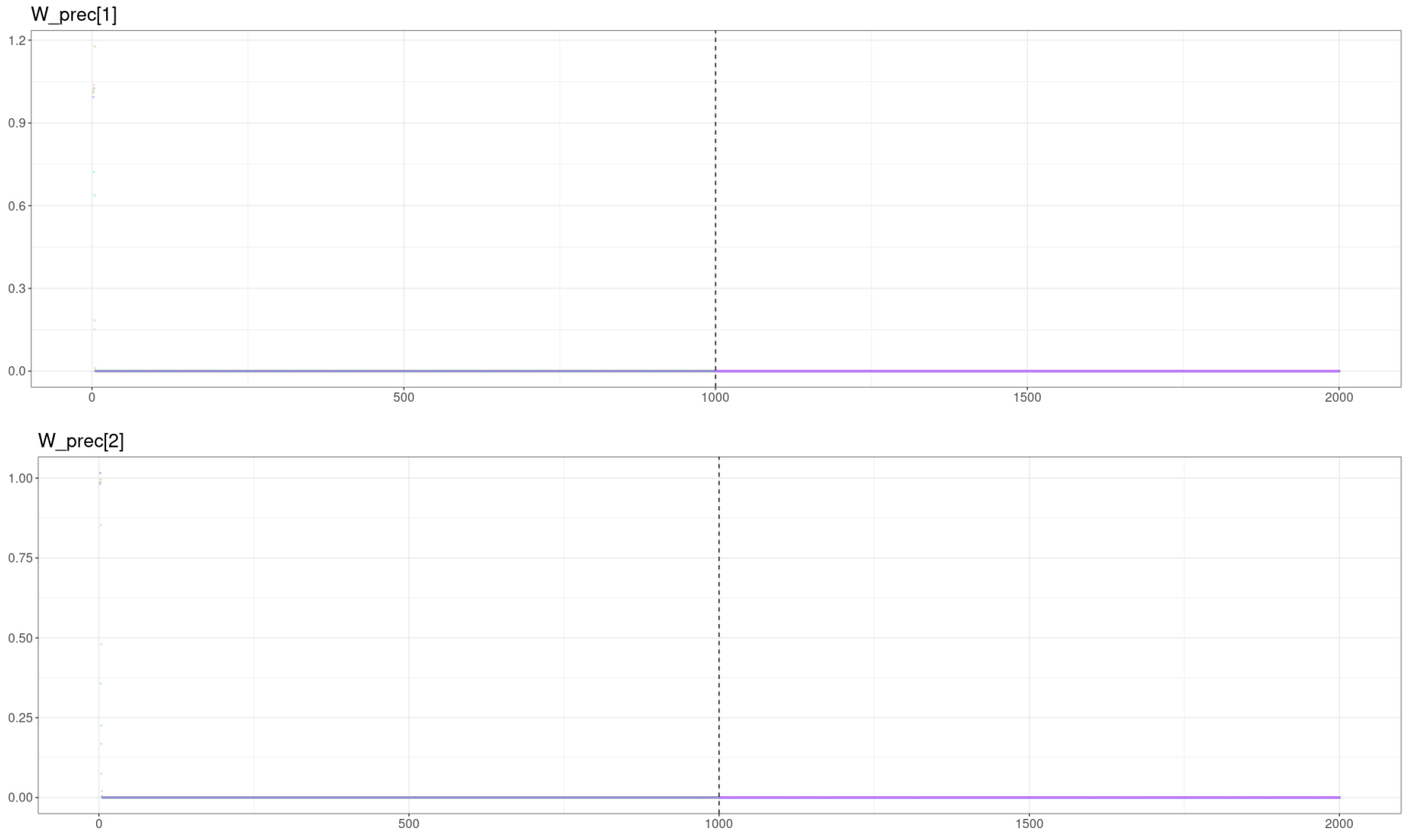

Figure 3: Hyperparameter Traces

NUTS (Centered) - Chain statistics

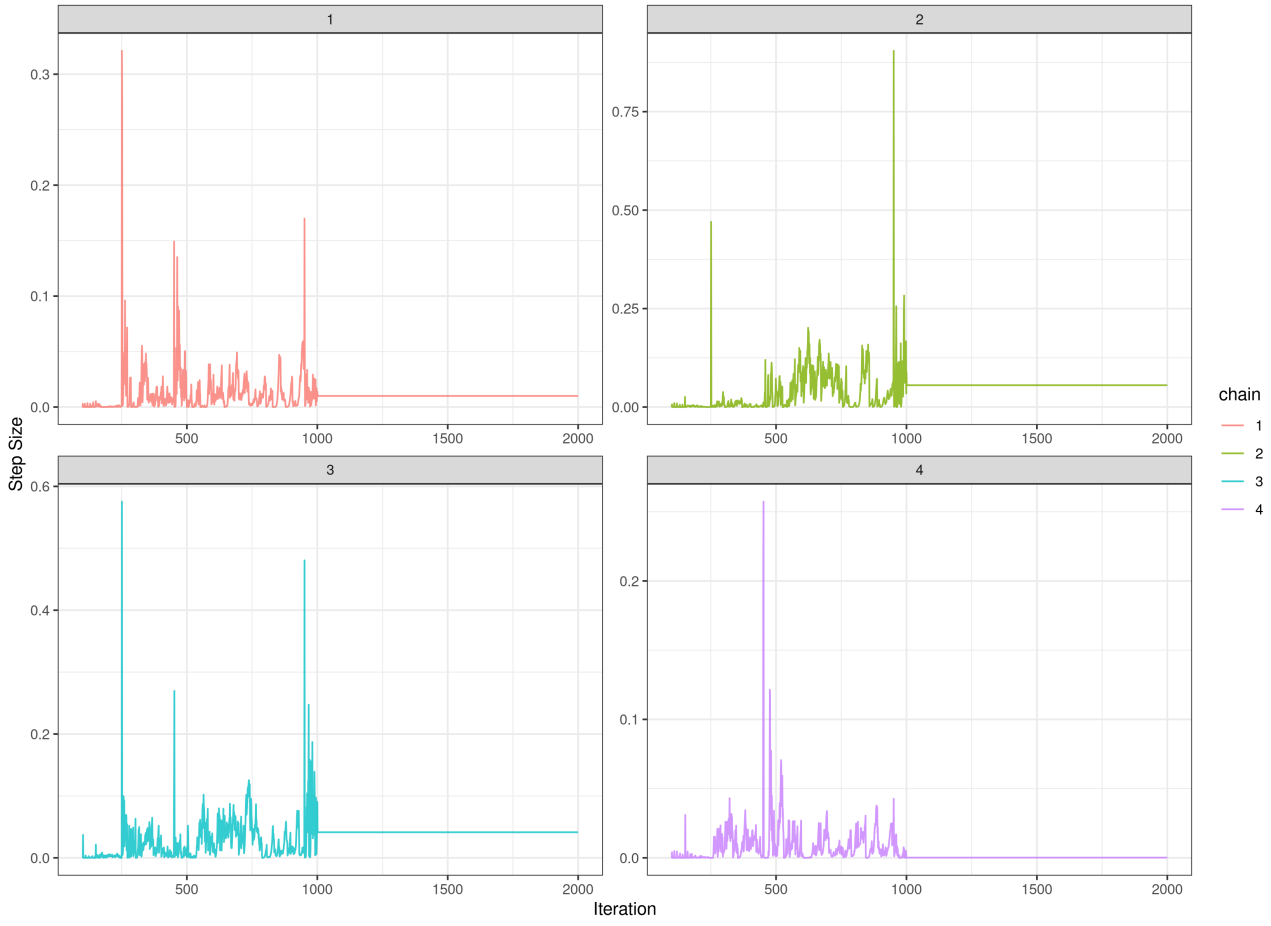

Figure 4: Chain Stepsizes

NUTS (Centered) - Chain statistics

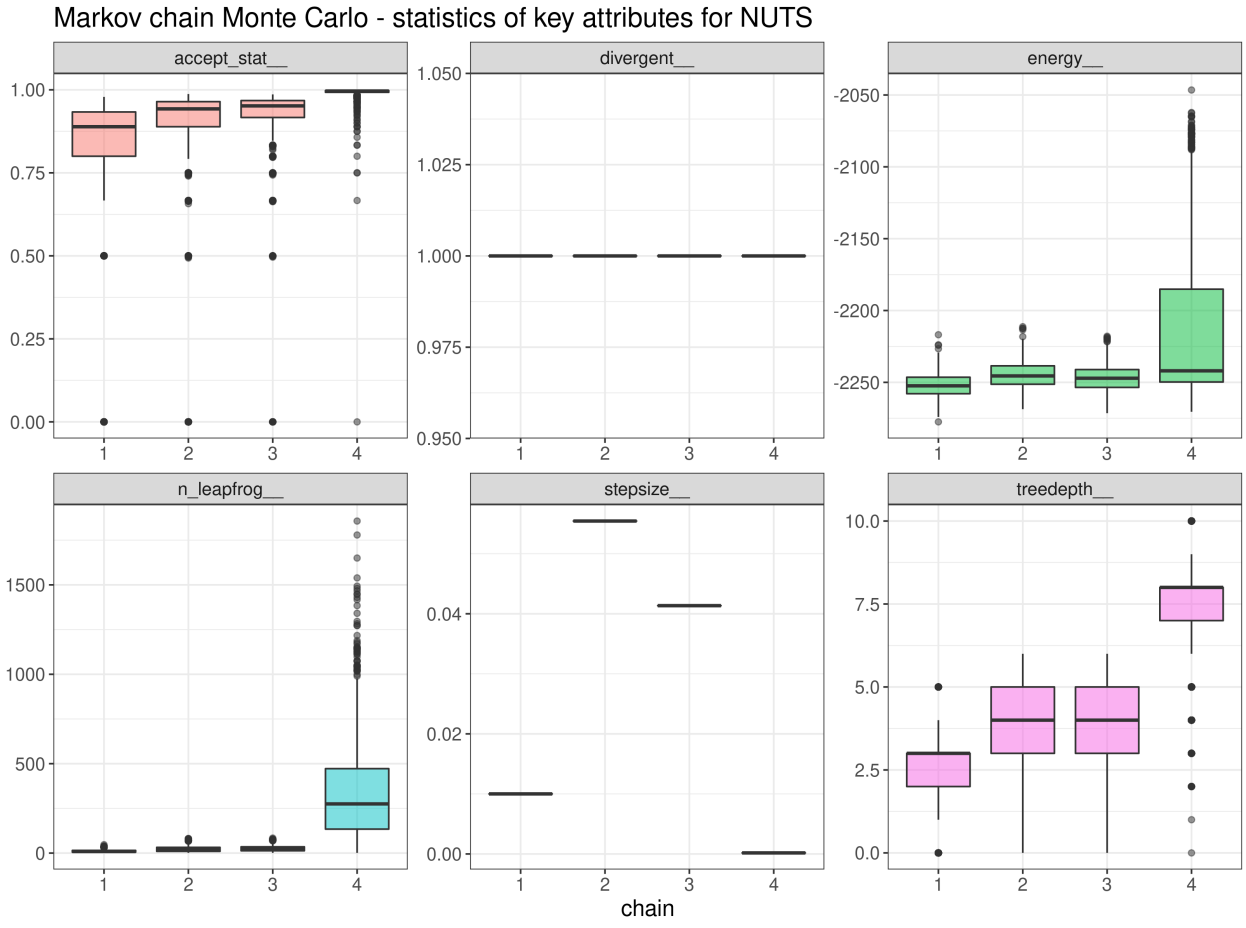

Figure 5: Chain Statistics

NUTS (Non-Centered Parametrization)

NUTS (Non-Centered) - Test Set Predictions

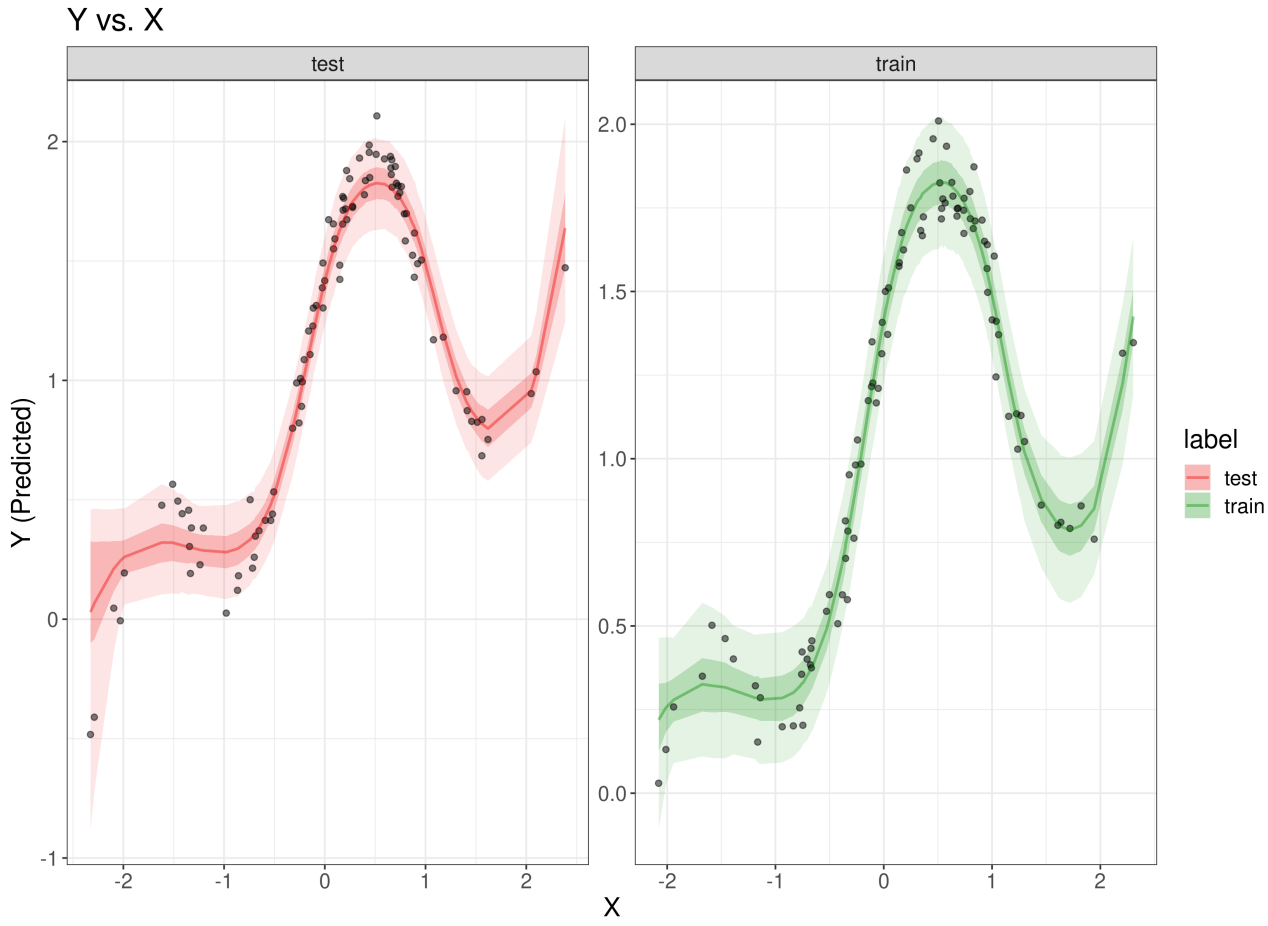

Figure 6: Predictive Quality

NUTS (Non-Centered) - Weight Traces

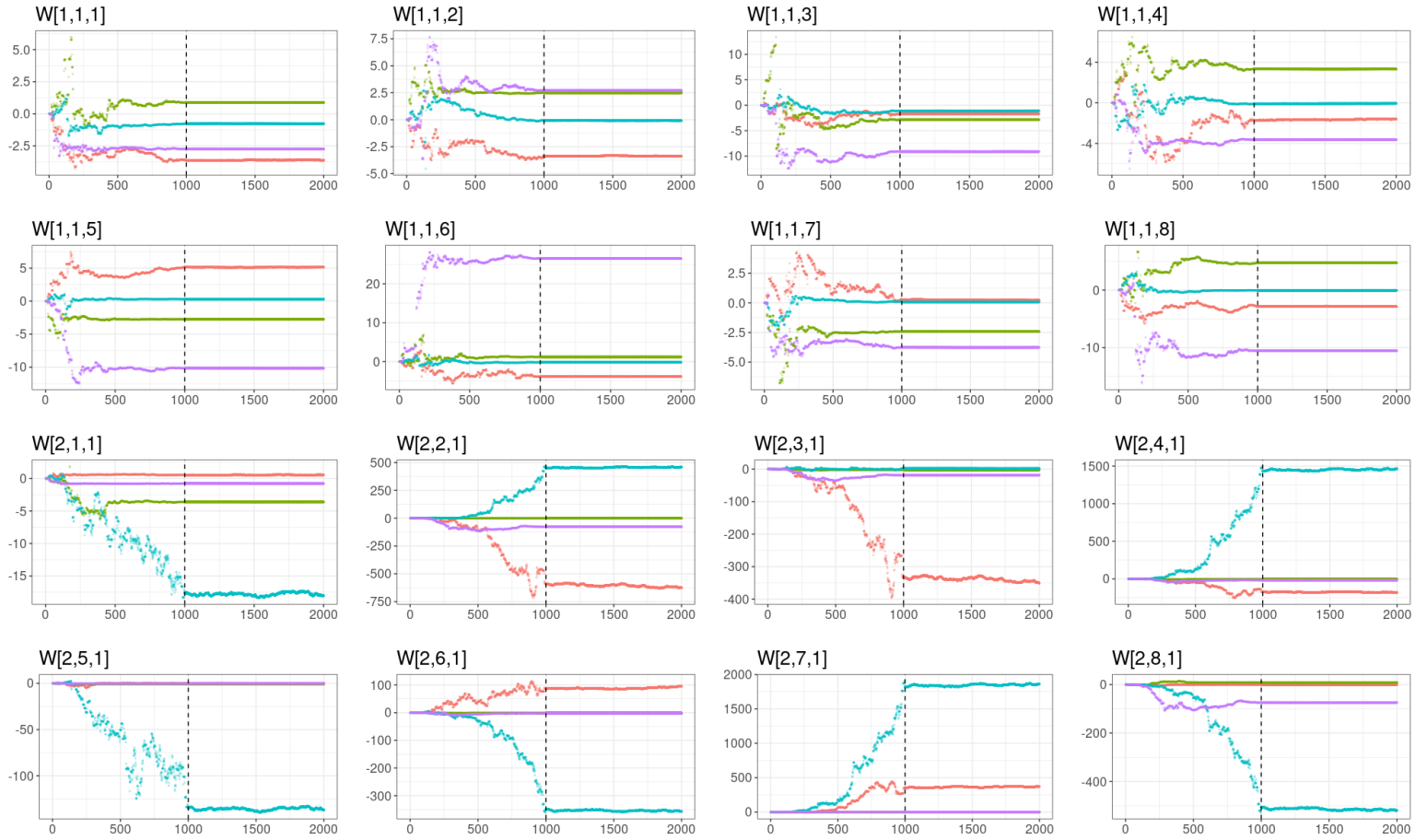

Figure 7: Weight Traces

NUTS (Non-Centered) - Hyperparameter Traces

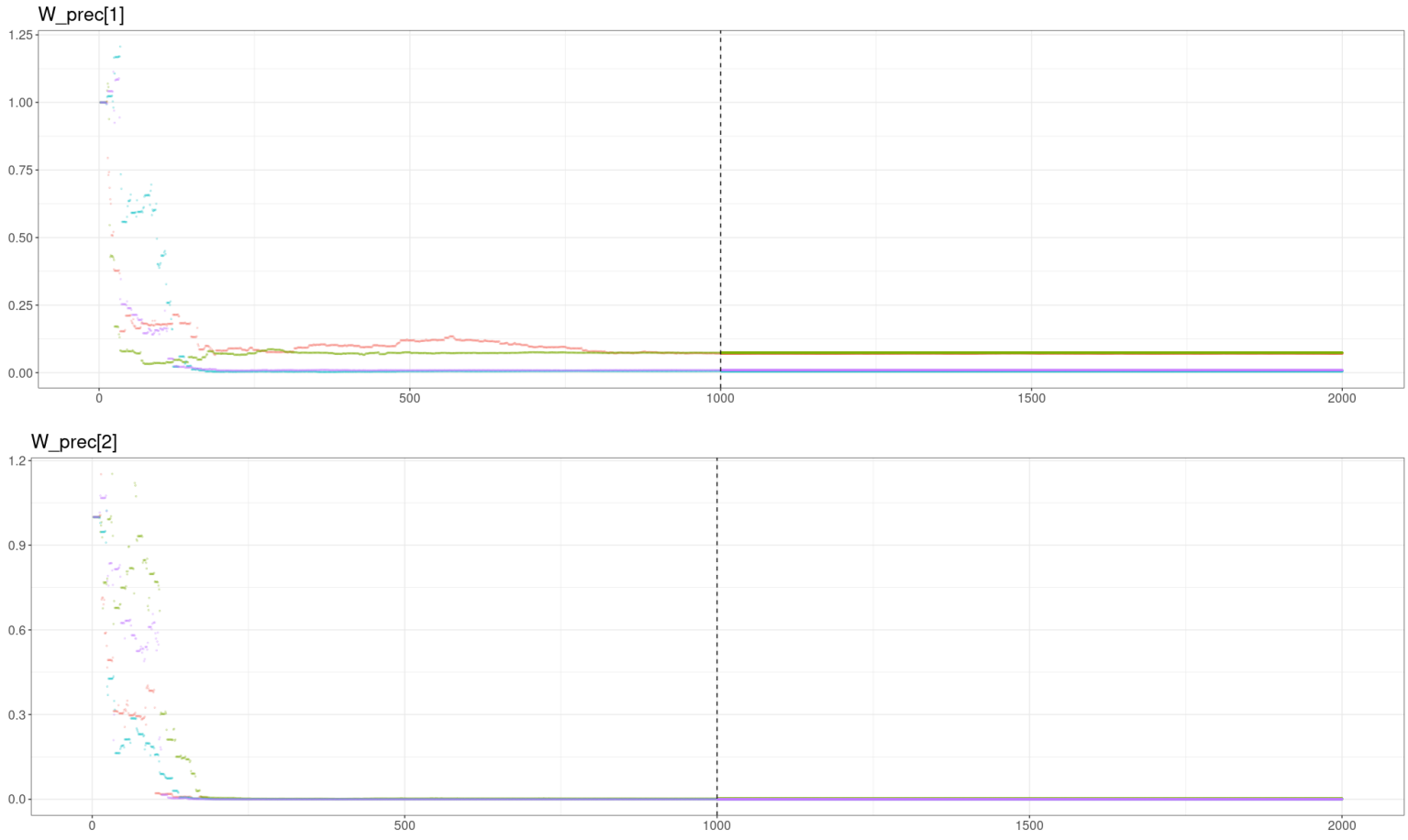

Figure 8: Hyperparameter Traces

NUTS (Non-Centered) - Chain statistics

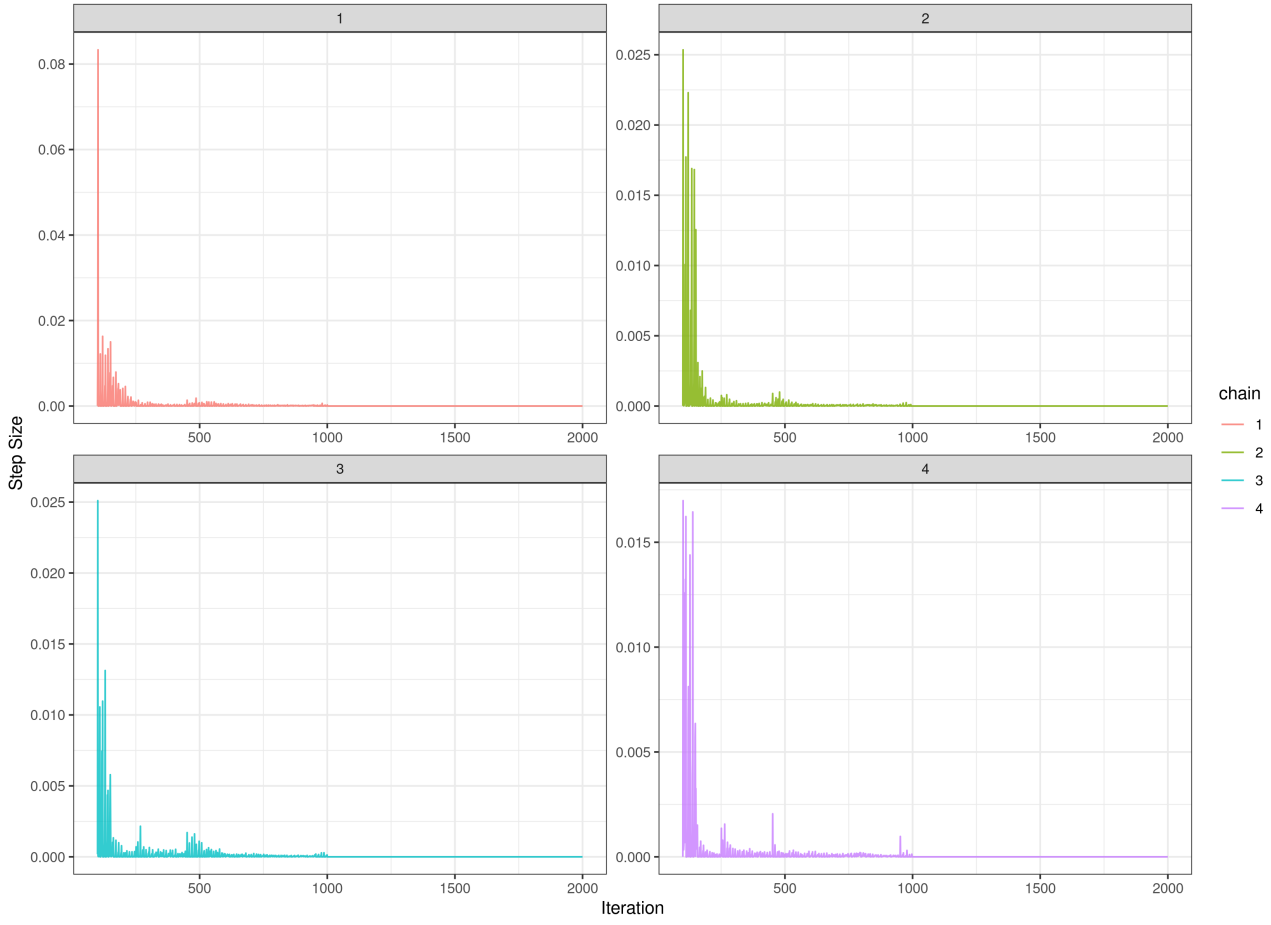

Figure 9: Chain Stepsizes

NUTS (Non-Centered) - Chain statistics

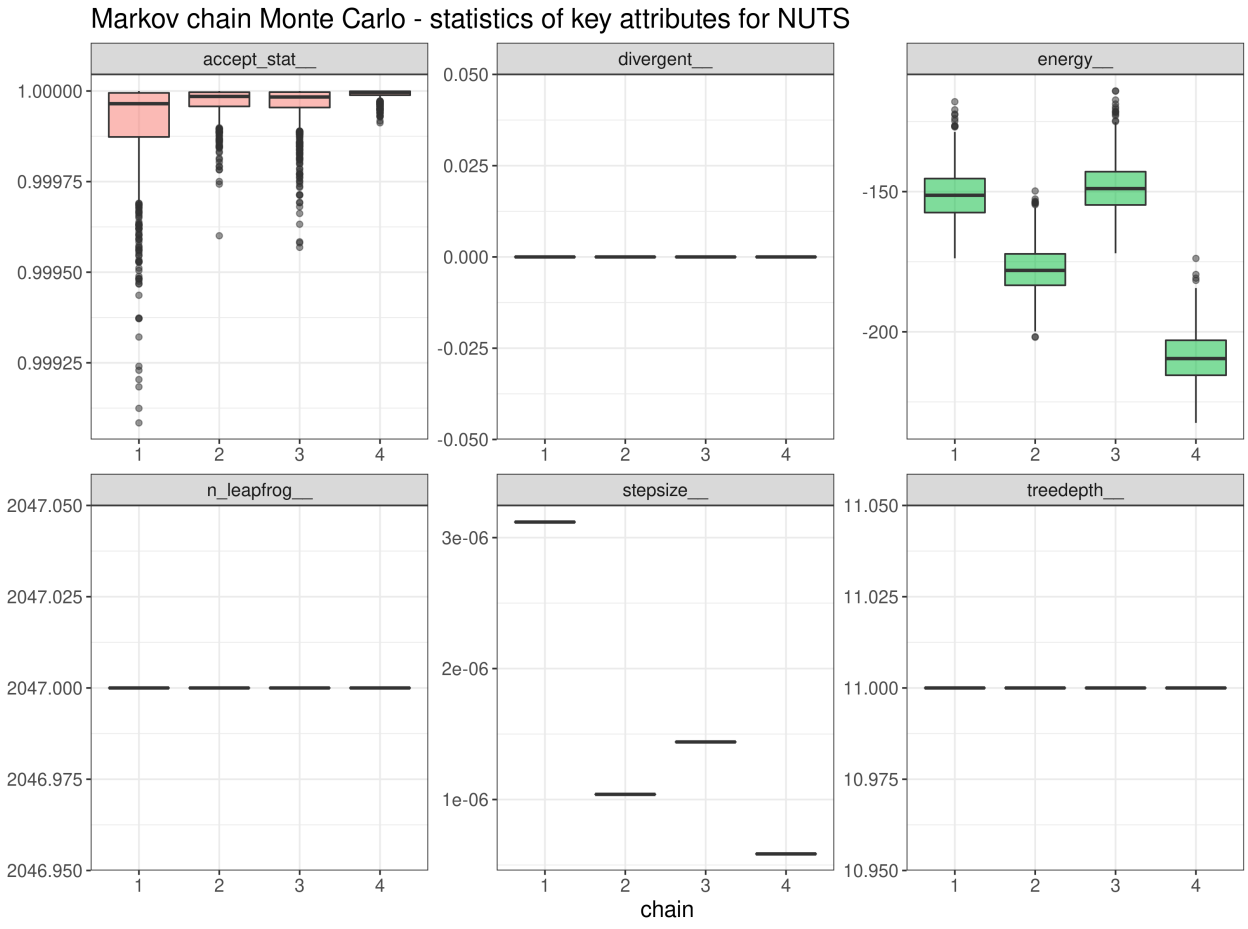

Figure 10: Chain Statistics

FBM

FBM (Non-Centered) - Test Set Predictions

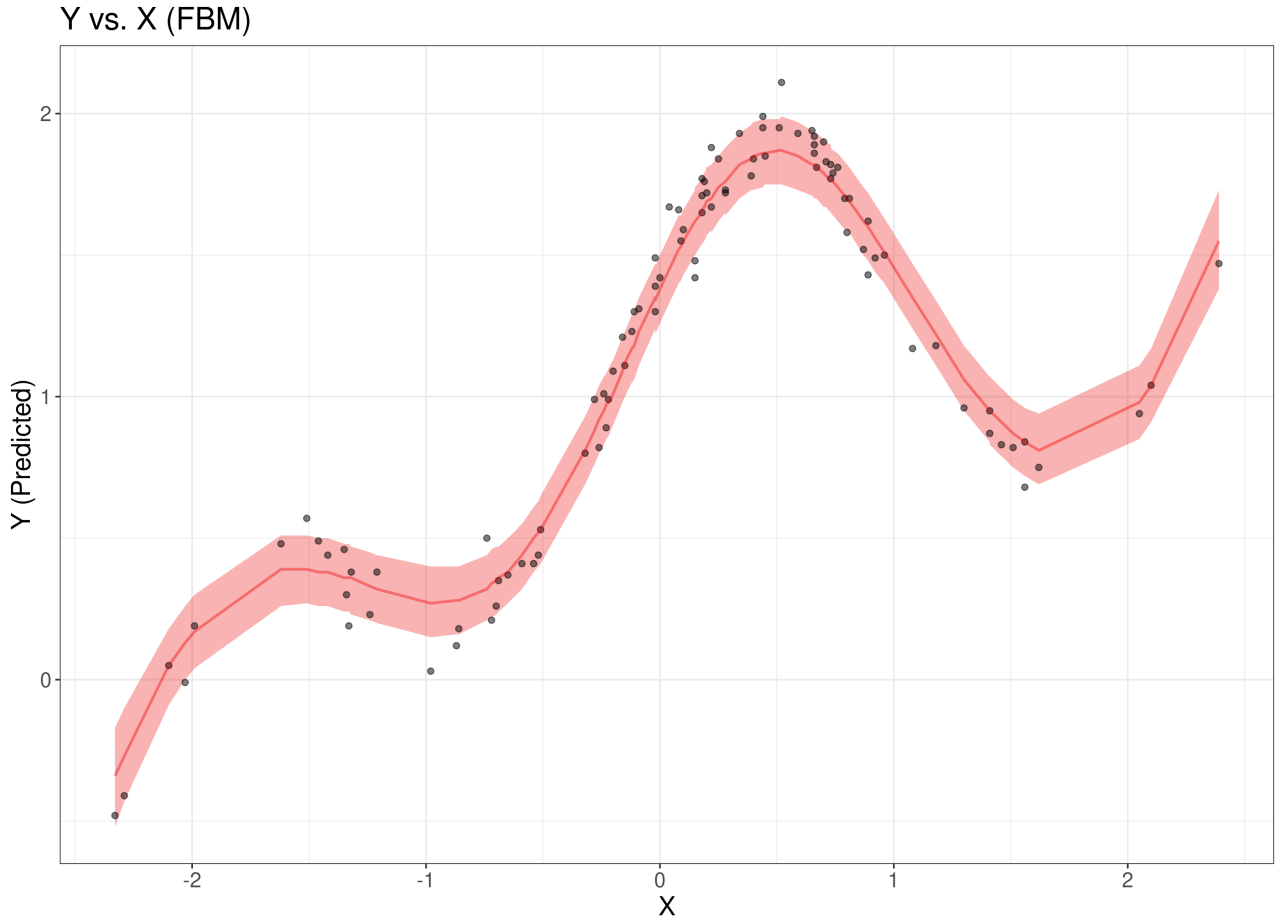

Figure 11: Predictive Quality

FBM (Non-Centered) - Weight Traces

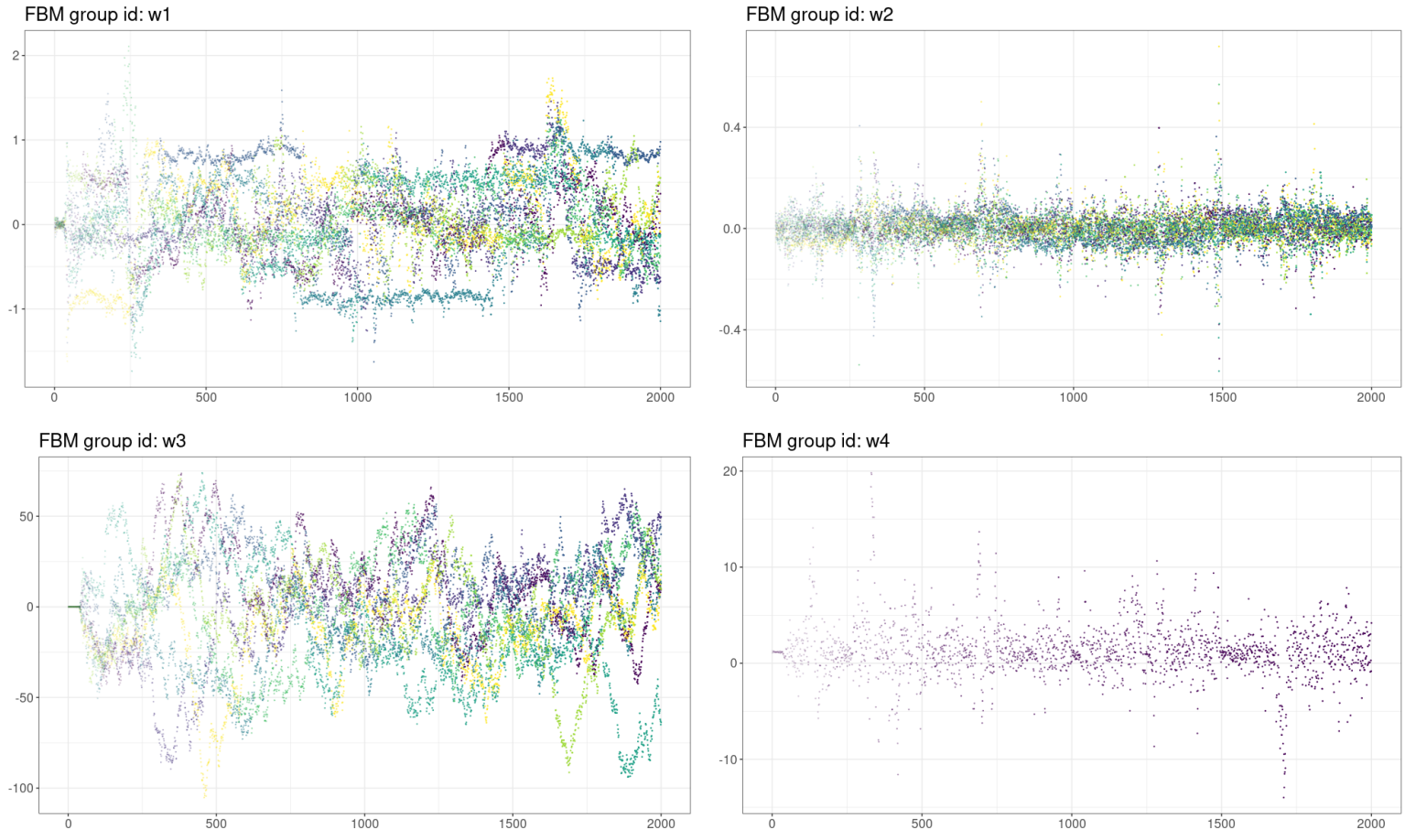

Figure 12: Weight Traces

NUTS (Non-Centered) - Hyperparameter Traces

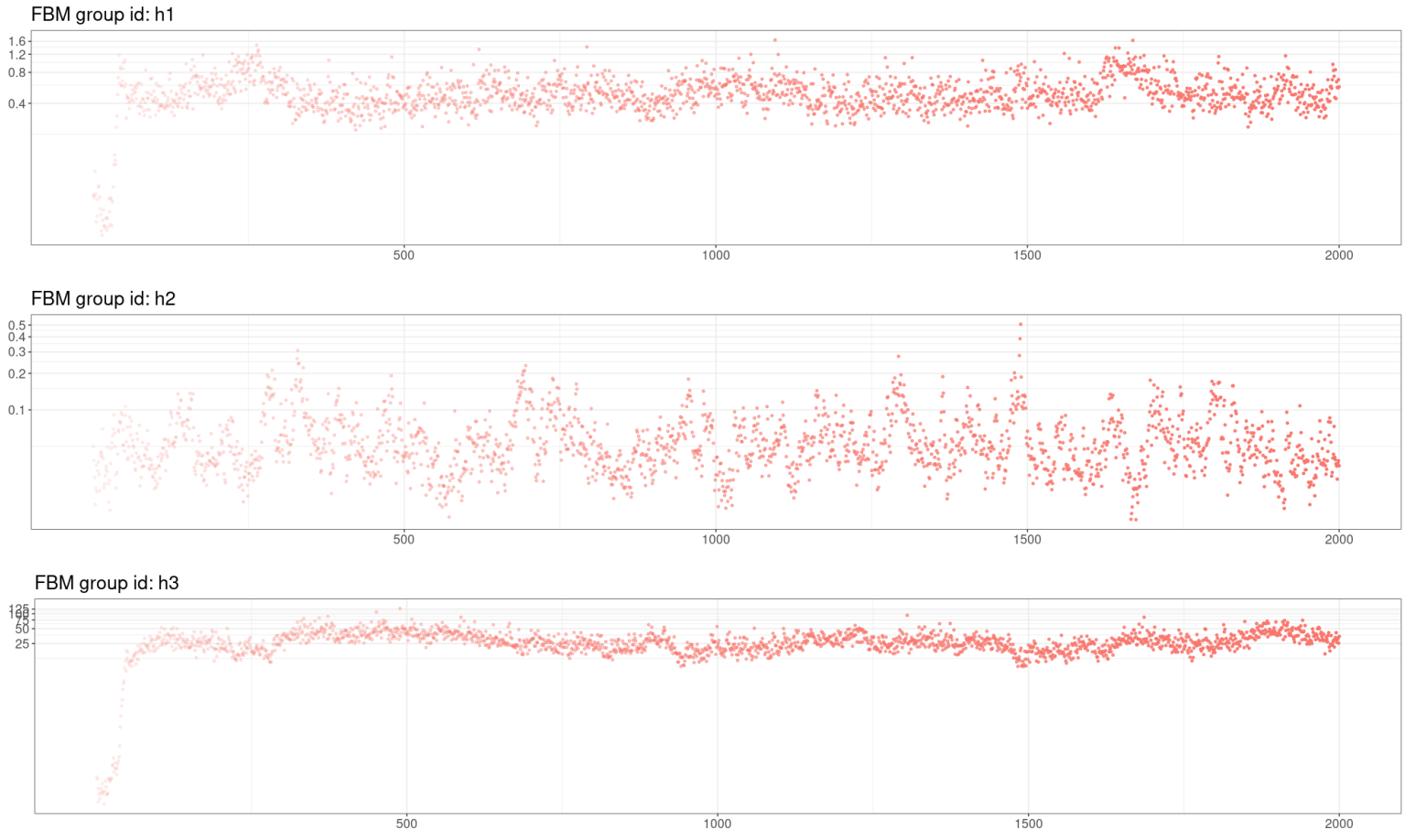

Figure 13: Hyperparameter Traces

Step size comparison

Step size comparison

Table 1: Step Size comparison

	Average Stepsize	Standard Deviation (stepsize) - % of mean
Centered NUTS	2.67e-02	0.00
Non-Centered Nuts	1.55e-06	0.00
FBM (Group 1)	3.78e-04	25.70
FBM (Group 2)	3.57e-04	25.20
FBM (Group 3)	8.83e-03	7.06