Bayesian inference for motor adaptation

M.F.C. van Leeuwen

This page was left intentionally blank

Bayesian inference for motor adaptation

Mark van Leeuwen
Student number: 4670655
11-11-2025

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)
Dept. of Neuroscience, Erasmus Medical Centre Rotterdam
01-09-2025 – 11-11-2025
Feasibility study (15 ECTS): 01-09-2025 – 11-11-2025

Supervisors:

Prof. dr. Maarten Frens
Dr. ir. Rick van der Vliet

Thesis committee members:

Prof. dr. Maarten Frens, Erasmus MC (Chair)
Dr. ir. Rick van der Vliet, Erasmus MC
Dr. ir. Winfred Mugge, TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Universiteit Leiden The Netherlands

TU Delft Delft University of Technology

Erasmus University Rotterdam

Preface

This thesis marks the end of my Master's Technical Medicine and with that the end of my time as a student. Since starting my bachelors in 2017, I have had so many great experiences and I have learnt and grown a lot both from my studies and from everything I have experienced with all the amazing people I met along the way.
Writing this thesis has been a challenging, yet rewarding process for me from which I have learnt a great deal. I especially want to thank Rick for his invaluable help, guidance, and patience during this final project. I would also like to thank Maarten and Jos for having a place for me at their research group. I also want to thank Caro and all the Master's students in the lab who turned this long journey into an enjoyable and fun experience.

Mark van Leeuwen
November 2025

Summary

Introduction: Motor adaptation is the process of adapting a movement plan to unexpected results due to a changing environment or changes in physical performance. Using the theory of optimal forward control this process can be described using using a forward control model with two learning parameters and two noise factors. This study will use Bayesian inference and a No-U-Turn sampler to estimate these learning and noise parameters from movement data from a reaching task in a small (N=60) and a large dataset (N=2226).
Methods: In total, six models were created and tested following a state-space model for adaptation. Three models used the same hierarchical design for the learning parameters and compared different hierarchical approaches for priors for planning noise (σ_η and execution noise (σ_ϵ). The other three models used the same non-hierarchical design for the noise parameters and compared different hierarchical hyperpriors and a non-hierarchical design for the learning rate (A) and adaptation rate (B).
Results: For the smaller dataset, the same issue was seen for all model designs, where the posterior distributions of σ_η are heavily skewed toward 0, with the HMC unable to converge ($\hat{R} > 1.1$). For the large dataset a non-hierarchical approach for both the learning parameters and noise parameters was able to converge for the majority of subjects for all four parameters.
Discussion: While none of the models performed perfectly, in this paper a model is created that is able to quantify motor adaptation and motor noise from visuomotor task data in a large dataset of 2226 subjects. In the future this model can be used for further research into neurological and genetic factors that influence motor adaptation.

Contents

1. Introduction

Motor adaptation is the process through which a learned movement is adapted to a changing environment. This process can even be seen in simple motor tasks such as reaching for an object, where the pattern of movement is constantly being adapted to changes in physical performance of the individual resulting for example from fatigue, and to changes in the environment resulting from for example wind. (1, 2) The speed of this process differs greatly from person to person and can be influenced by a combination of neurological and genetic factors. (3, 4) This combined field of population neuroscience has seen growing interest over the past decade. (5, 6) The near infinite number of combinations of genes and environments means that a very large sample size is necessary in order to reasonably draw conclusions from a study into their effects on cognitive neuroscience. The Generation R (GenR) study is such a large cohort study that studies the role of environmental and genetic causes in the growth, development and health from fetal life until young adulthood. (7–9) Within this cohort, data was gathered on the individual genetics and environmental exposures of the participants, and, among other tests, a motor adaptation reaching task was performed and MRI imaging was performed. This study will describe how motor adaptation can be quantified and will apply this method to the large GenR dataset so the adaptation data can be used for further research.

The theory of optimal feedforward control (OFC) describes motor adaptation as a computational feedforward model, where the nervous system uses a combination of a predicted movement plan and sensory feedback received during movement to optimize movement control and to correct for movement errors. (10) The rate at which motor adaptation takes place can be described using two learning parameters related to retention and adaptation. Alongside these learning parameters, the motor adaptation rate is also influenced by variability introduced in the execution of the movement plan and in the processing of sensory information. (11) The variability in movement execution is referred to as execution noise and stems from variability introduced in the sensorimotor pathway in the recruitment of motor units. (12) The variability in the processing of sensory information is referred to as planning noise and stems mainly from processing of sensory information in the premotor cortex. (13)

While these noise factors and the learning parameters can not be observed, it is possible to estimate them using Bayesian inference. Bayesian inference modelling uses a joint probability distribution of all observed and unobserved parameters in a problem to estimate a probability density function for each of these parameters. (14) Because the Bayesian approach results in a probability interval, it is well suited to estimate the noise and learning parameters, which are inherently stochastic in nature. (15)

Bayesian inference can also be used to give the model a hierarchical nature, where parameters are simultaneously estimated on a group level, on a per-subject basis and for each trial of a specific subject. (16) This means that the model more closely portrays the hierarchical nature of motor adaptation, where some learning parameters change depending on the conditions of the specific trial, while others are used to describe the behaviour of the individual participant. (17) The outcomes of the model are also more robust to variance between subjects.

The model outcomes are inferred using a Markov Chain Monte Carlo (MCMC) method. MCMC methods create sequences of samples that converge in a target probability distribution. Several chains of samples are created, starting in random points that substantially differ from each other. New samples are added to the chain based on a likelihood that is assigned by an algorithm. (18)

Several algorithms can be used to obtain this sequence of samples. In this study, the No-U-Turn Sampler (NUTS) algorithm was used. (19) NUTS is an extension of the Hamiltonian Monte Carlo (HMC) algorithm, that uses a leap frog integrator to propose new point to sample in the state-space. (20, 21) This reduces the correlation between steps when compared to other MCMC algorithms such as the random walk Metropolis and the GIBS sampler. This means that fewer samples are needed in order to converge to a target distribution, which improves efficiency. A drawback of HMC is that manual tuning is necessary for the algorithm to work correctly. NUTS removes the need for manual, making it easier to implement without sacrificing efficiency. (19, 22)

Van der Vliet et al. 2018 (11) has previously shown that Bayesian modelling can be used to extract planning and execution noise, retention rate and adaptation rate from a visuomotor adaptation task using a GIBS sampler. However, the lower efficiency of this MCMC algorithm hampers the usability of this model for a larger dataset like the Generation R study. This study will improve on this model by redesigning the model using a NUTS sampler in order to raise the model's efficiency, which will allow for the use of a large dataset.

2. Methods

2.1 Model design

The Bayesian model that was used is based on the following first-rate model:

$$x[n+1] = Ax[n] - Be[n] + \eta[n] \tag{1}$$

$$e[n] = y[n] + p[n] \tag{2}$$

$$y[n] = x[n] + \epsilon[n] \tag{3}$$

Here, y[n] is the measured movement angle of an individual subject at trial n. Movement angle y[n] is dependent on the intended movement angle x[n] and a noise factor ϵ, which signifies execution noise. This is a factor for variability stemming from noise in the sensorimotor pathway. (12)

The planned movement x[n+1] is dependent on the previous movement plan x[n], the retention rate A, adaptation rate B, the error e observed in the previous trial and the planning noise η. The observed error is determined as a combination of the movement angle y and the applied perturbation p. Planning noise stems from variability in neuronal processing of sensory information.

The model was created and tested using the PyMC library (v5.16.2) for Python (v3.12.4). (23) . Several model designs were tested using hierarchical and non-hierarchical models for the estimation of the noise- and learning parameters. A general visualization of the model is shown in fig. 1. This visualization uses a hierarchical approach for the noise and learning parameters. In this study several designs were used that adapted parts of the model as explained in the following sections. Visualisations for each of these models and the distribution shape of the model's priors can be found in Appendix A.

2.1.1 Non-hierarchical noise models

Four models were designed using a non-hierarchical prior for the noise parameters, with differing priors for the learning parameters. The noise parameters η and ϵ were modelled as zero-mean Gaussians with standard deviation σ_η and σ_ϵ. σ_η was defined as $\Gamma(1, 0.5^2)$ and σ_ϵ was defined as $\Gamma(4, 1)$.

The distribution of the learning parameters A and B were specified in the model as a normal distribution in the logistic space using a standard logistic function. This distribution was sampled using three separate specifications. The first set of priors used informated hyperpriors μ_A, μ_B, σ_A and σ_B as shown in table 1. These priors were chosen in order to represent information from previous studies (11, 17), but without constraining the results of the model. In order to assess the robustness of the posterior for the chosen priors and to check for possible bias introduced by the priors, a second set of priors was used for the same general model design. These priors (Table 1) were much wider than the previous prior set, while still retaining the expected relative size of A and B, with the retention rate (A) expected to be greater than the adaptation rate (B).

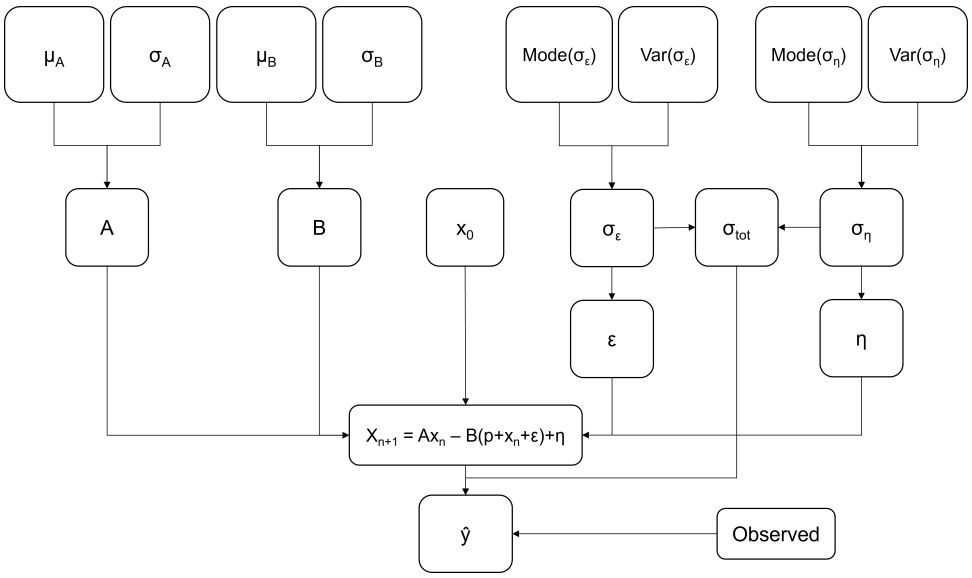

Figure 1: Visualization of the designed Bayesian model, showing the prior distributions of each parameter in the model and their codependence

Table 1: Hyperprior distributions of learning parameters

	μ_A	μ_B	σ_A	σ_B
Informed priors	$N(4, 0.25^2)$	$N(-2, 0.25^2)$	$\Gamma(0.5, 0.25^2)$	$\Gamma(0.5, 0.25^2)$
Wide priors	$N(2, 1)$	$N(-2, 1)$	$\Gamma(1, 0.5)$	$\Gamma(1, 0.5)$

In a third hyperprior specification, a non-hierarchical design was used for the learning parameters, defining the logit-normal distributions A and B directly as

$$A \sim P(N(2, 0.5^2))$$

$$B \sim P(N(-2, 0.5^2))$$

2.1.2 Hierarchical noise models

The models with a hierarchical description of the noise parameters used the determined the learning parameters using a hierarchical design with informed priors as described in table 1. Several methods were used to define the hyperpriors for the noise parameters σ_η and σ_ϵ.

In the first model, σ_η and σ_ϵ were defined as a gamma distribution with gamma hyperpriors on the mode and variance.

$$Var(\sigma_\eta) \sim \Gamma(0.5, 0.25^2) \qquad Mode(\sigma_\eta) \sim \Gamma(1, 0.5^2)$$

$$Var(\sigma_\epsilon) \sim \Gamma(4, 0.5^2) \qquad Mode(\sigma_\epsilon) \sim \Gamma(0.5, 0.25^2)$$

Second, a non-centred parametrization was used on the prior of the planning noise. (24) This was done in order to guide the posterior distribution of the planning noise away from 0. Using this method, σ_η was sampled from a gamma distribution with a mode sampled from a gamma prior and a variance defined as $mode * v_{norm}$, where v_{norm} is a gamma distribution with a fixed high density interval (HDI) between 0.1 and 5. Generally, a non-centred parametrization is used when dealing with areas of high curvature in the probability space, so called funnels. These funnels can prohibit the

algorithm from efficiently exploring the posterior distribution. (25) Alongside this effect, the chosen v_{norm} ensures that the variance of the prior for σ_η at minimum will be 10% of the mode, which is positive and greater than 0.

In the third model, σ_η and σ_ϵ were not sampled directly but were defined as components of the total noise σ^2_{total}, where $\sigma^2_\eta = p * \sigma^2_{total}$ and $\sigma^2_\epsilon = (1-p) * \sigma^2_{total}$. The variance ratio p was sampled from a logistic normal distribution to ensure a range from 0 to 1. σ_{total} was sampled from a gamma distribution with hyperpriors for μ and σ defined as

$$\mu \sim \Gamma(20, 1)$$

$$\sigma \sim \Gamma(1, 0.25)$$

2.2 Dataset

For the design and testing of the model two separate datasets were used as observed data for y[n] in eq. 3. The first dataset consisted of reaching data gathered for Van der Vliet et al. 2018. (11) This dataset consisted of 60 participants with 900 trials for each subject. The second dataset contained data gathered in the Generation R study. This dataset consisted of 2226 participants with 110 trials for each subject. The first dataset was used to create and test the model designs. The second dataset was then applied to the same models in order to test the robustness of the models.

2.2.1 Experimental design

The experiments used for both datasets followed similar design for the reaching task. The participant is seated in front of a horizontal opaque surface, while holding a robotic handle under the surface obscured from view. On this surface a cursor and a target are projected. The cursor is controlled by moving the robotic handle. The participants were instructed to perform a straight shooting movement and only decelerate when the cursor has moved past the target. When the cursor has moved a distance from its origin equal to the distance between the origin and the target, the movement is dampened by a force cushion and the handle moves back to the origin position.

Set 1 consisted of 450 baseline trials and 450 perturbed trials, in which the perturbation is applied in a stepwise manner increasing by 1.5° every 8 to 12 trials until reaching 9° and then switching in the opposite direction until -9°. (Fig. 2a) Set 2 consisted of 20 baseline trials and 90 perturbed trials, with the perturbation increasing by 2° every 5 trials up to 8° and switching in the opposite direction until -8°. (Fig. 2b)

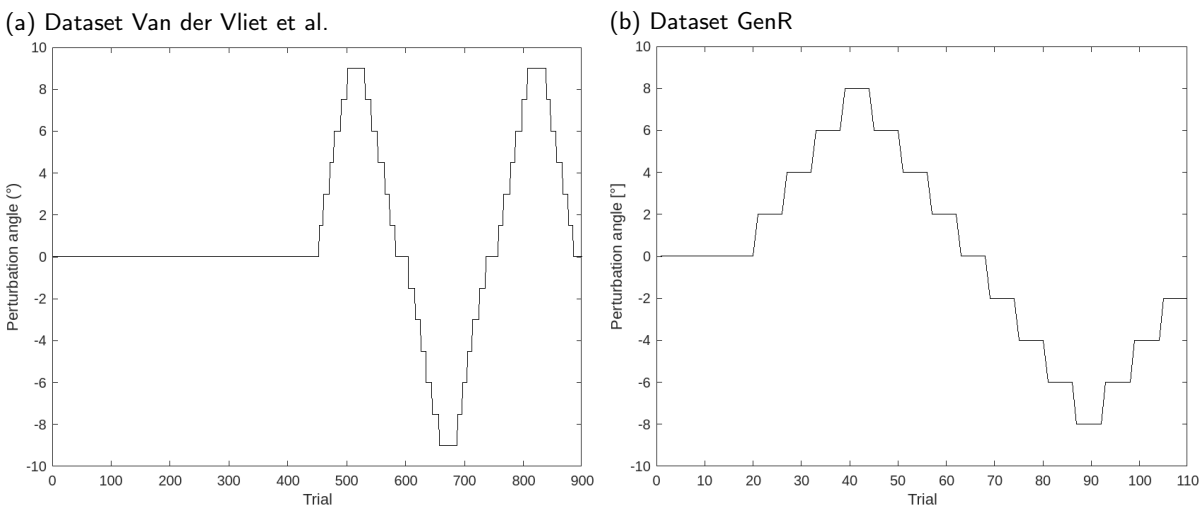

Figure 2: Experimental designs of reaching tasks

2.3 Model evaluation

The models were evaluated on an individual level by assessing the reliability and robustness of the results, the model efficiency and the predictive accuracy of the model.

The reliability and robustness of the results were assessed using the Gelman-Rubin diagnostic $\hat{R}$. (26) This is a score for the convergence of chains. When all chains converge to the same distribution, $\hat{R}$ converges to 1. Generally, it is recommended to aim for $\hat{R} \leq 1.1$. (27, 28)

Besides the posterior distributions that follow from the models, the models were evaluated based on efficiency. In Bayesian modelling this can be done by visually comparing the transition energy density and the marginal energy density distributions. Hamiltonian Monte Carlo uses changes in energy induced by what is called momentum resampling to explore the target distribution. When the transition energy distribution induced by this momentum resampling is similar to the marginal energy distribution corresponding to the target probability distribution, the algorithm can easily explore the relevant energies, improving robustness and leading to small auto-correlations in the Hamiltonian Markov Chain (fig. 3). (29) The comparison is quantified using the Bayesian Fraction of Missing Information (BFMI). When BFMI $\rightarrow$ 0 the momentum resampling induces slow exploration and BFMI $\rightarrow$ 1 showing that the momentum resampling effectively generates draws from the marginal energy distribution. (30)

The predictive accuracy of each model was determined using Pareto Smoothed Importance Sampling Leave One Out Cross Validation (PSIS LOO-CV). The expected log pointwise predictive density (ELPD) that is produced by this method can be used both to evaluate individual models as well as a way to compare the different models. (14, 31)

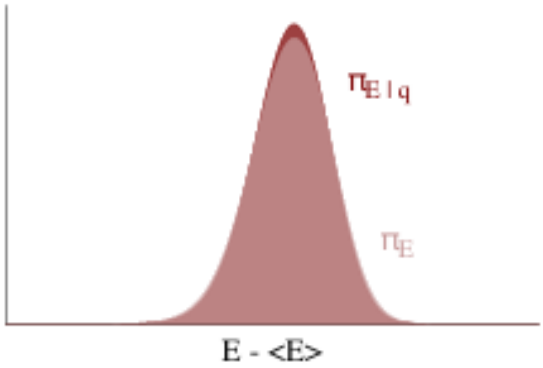

Figure 3: Example of the marginal energy distribution π_E and energy transition density $\pi_{\Delta E}$ of an efficient Hamiltonian Monte Carlo Chain. From Betancourt 2016 (29)

3. Results

3.1 Dataset evaluation

Outliers were previously removed from the dataset from Van der Vliet et al. as described in the paper. For the GenR dataset, trials were removed when the reaching angle was greater than 45°. Due to recording errors that would often occur during the early trials, the first five trials of each subject were removed. Analysis of the datasets show that the GenR data contained a lot of missing data (fig. 4B).

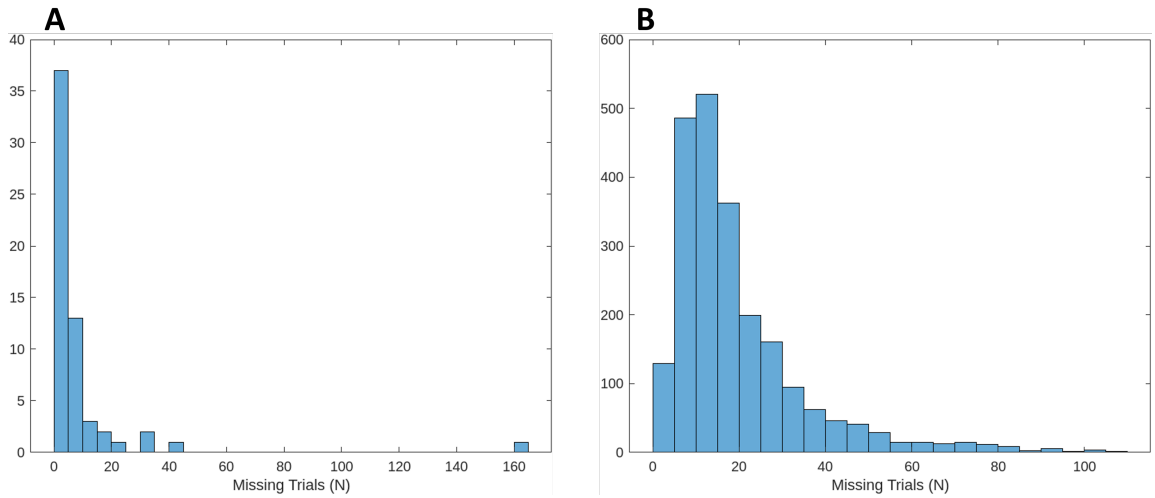

Figure 4: Number of missing data points per subject for the dataset from Van der Vliet et al. (A) and the GenR dataset (B)

3.2 Dataset 1 (N=60x900)

First, the dataset used by Van der Vliet et al. (11) was used as observed data to run the previously described models. A summary of the combined posterior distributions of the learning and noise parameters across all subjects is shown in table 2. For an overview of the reliability of these results, table 3 shows a count of each data point where $\hat{R} > 1.1$ for each learning and noise parameter.

Table 3 shows that for all models, the parameter σ_η does not converge well, with $\hat{R} > 1.1$ for between 86.67% and 100% of all subjects. While the hierarchical noise models show fewer σ_ϵ sample distributions with $\hat{R} > 1.1$ when compared to the non-hierarchical noise models, these results are still not believed to be better. This is because the probability distributions for σ_η are strongly weighted towards 0.0, which is unlikely to be an accurate estimation, based on results from previous studies. (11, 17)

In order to get a better understanding of the performance of each model, the distributions of $\hat{R}$ of the parameter σ_η for each subject per model are shown in figs. figs. 5 and 6. These figures show that, while the number of data points where $\hat{R} > 1.1$ is similar, the convergence for these parameters in the models with a hierarchical design for the noise parameters is much worse. The median $\hat{R}$ for the models with a non-hierarchical approach for the noise parameters are 1.49 (Informed prior), 1.50 (Wide prior) and 1.91 (Non-hierarchical learning parameters). For the models with a hierarchical approach for the noise parameters, the median $\hat{R}$ are 2.18 (Gamma prior), 2.44 (Non-centred reparametrization) and 2.49 (Probability ratio).

Table 2: Posterior summary of group level learning and noise parameters. NH: Non-hierarchical; NC: Non-centred

Model	A Median [94% HDI]	B Median [94% HDI]	σ_ϵ Median [94% HDI]	σ_η Median [94% HDI]
Non-hierarchical noise models				
Informed	0.98 [0.97, 0.99]	0.16 [0.06, 0.31]	2.5 [1.7, 4.1]	0.19 [0.00, 0.75]
Wide	0.98 [0.97, 0.99]	0.13 [0.07, 0.21]	2.5 [1.7, 4.1]	0.28 [0.00, 0.63]
NH	0.97 [0.94, 0.99]	0.16 [0.06, 0.31]	2.4 [1.7, 4.1]	0.3 [0.00, 0.76]
Hierarchical noise models				
Gamma	0.98 [0.97, 0.99]	0.13 [0.062, 0.21]	2.5 [1.7, 4.2]	0.01 [0.00, 0.46]
NC	0.98 [0.97, 0.99]	0.13 [0.062, 0.23]	2.5 [1.8, 4.2]	0.00 [0.00, 0.27]
Ratio	0.98 [0.97, 0.99]	0.13 [0.06, 0.22]	2.5 [1.8 4.2]	0.00 [0.00. 0.48]

Table 3: Count of data points where $\hat{R} > 1.1$. NH: Non-hierarchical; NC: Non-centred

Model	A	B	σ_ϵ	σ_η
Non-hierarchical noise models				
Informed	0 (0%)	2 (3.33%)	21 (35%)	52 (86.67%)
Wide	0 (0%)	0 (0.0%)	19 (31.67%)	51 (85%)
NH	1 (1.67%)	4 (6.67%)	32 (53.33%)	53 (88.33%)
Hierarchical noise models				
Gamma	0 (0%)	0 (0%)	8 (13.33%)	53 (88.33%)
NC	1 (1.67%)	6 (10%)	7 (11.67%)	60 (100%)
Ratio	0 (0%)	2 (3.33%)	7 (11.67%)	56 (93.33%)

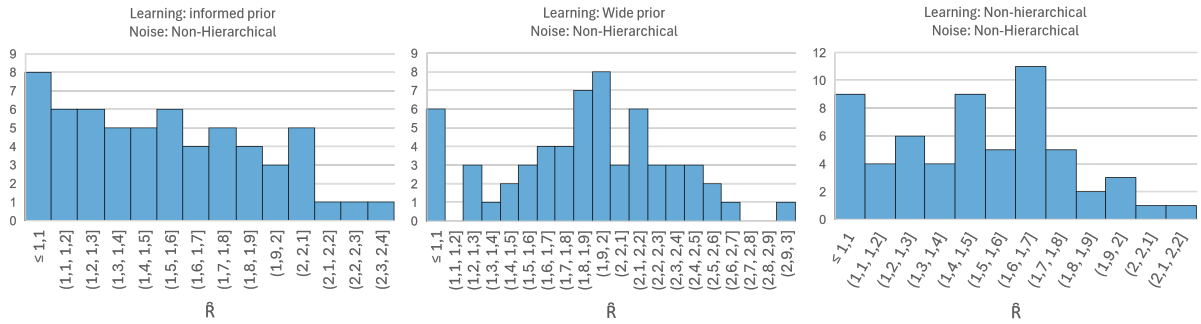

Figure 5: Distributions of Gelman-Rubin statistic for σ_η for each subject using non-hierarchical priors for the noise parameters

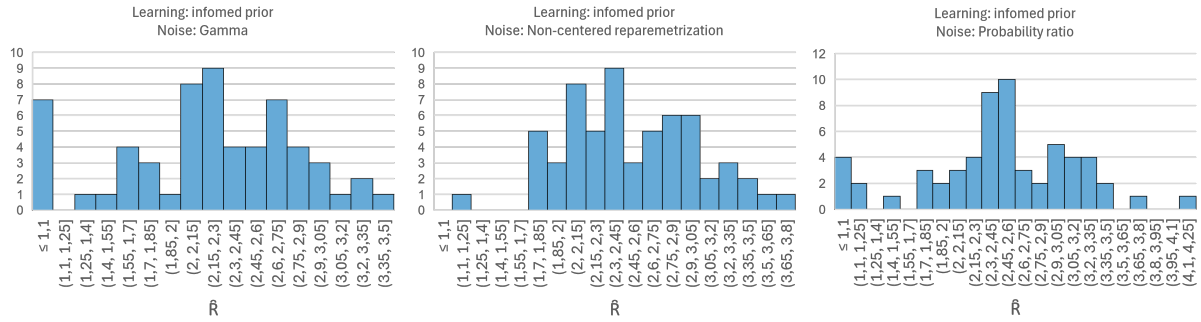

Figure 6: Distributions of Gelman-Rubin statistic for σ_η for each subject using hierarchical priors for the noise parameters

3.2.1 Model comparison

The performance of the models were compared using the ELPD following from PSIS LOO-CV. The ELPD and the corresponding standard error are shown in fig. 7. As shown in the figure, the ELPD for all models is very similar meaning that there is no significant difference in expected predictive accuracy between the models.

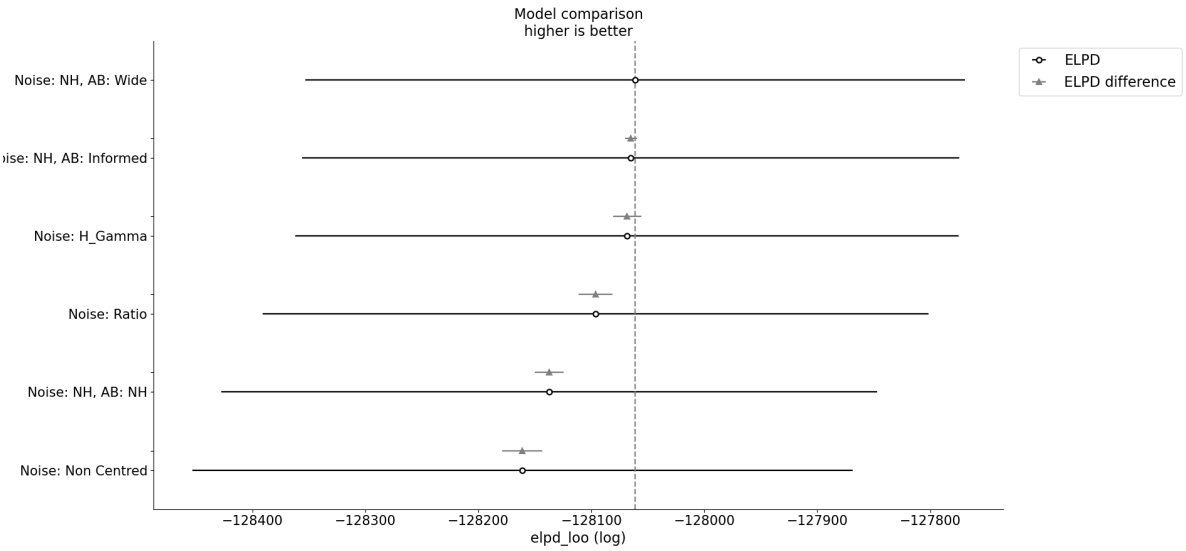

Figure 7: Model comparison using PSIS LOO-CV

3.3 Dataset 2 (2226x105)

As the results of the hierarchical noise models that used dataset 1 showed a tendency for the parameter σ_η to become infinitely small, it was decided to only continue with the non-hierarchical approach for the noise parameters. A summary of the posterior distributions of the learning- and noise parameters for all four non-hierarchical noise models is shown in table 4.

As seen in table 4, every set of priors used in the model result in a similar posterior, with credible results when compared to similar studies. This gives reason to believe that the model is implemented correctly and the posterior distributions are not biased towards the given prior distributions. Table 5 shows that none of the models perform flawlessly, as shown by the number of subjects for which the parameter σ_η has a $\hat{R} > 1.1$, ranging from 24.66% up to 45.46%. However, these results are an

improvement when compared to the results of dataset 1. Overall, the model performs the best for the design that uses the non-hierarchical approach for both the learning- and noise parameters.

Table 4: Posterior summary of group level learning and noise parameters. NH: Non-hierarchical

Model	A Median [94% HDI]	B Median [94% HDI]	σ_ϵ Median [94% HDI]	σ_η Median [94% HDI]
Non-hierarchical noise models				
Informed	0.96 [0.95, 0.96]	0.25 [0.097, 0.43]	3.4 [1.7, 5.6]	0.74 [0.14, 1.6]
Wide	0.96 [0.95, 0.96]	0.25 [0.098, 0.43]	3.4 [1.7, 5.6]	0.74 [0.14, 1.6]
NH	0.93 [0.86, 0.97]	0.19 [0.058, 0.33]	3.4 [1.7, 5.7]	0.8 [0.16, 1.7]

Table 5: Count of data points where $\hat{R} > 1.1$. NH: Non-hierarchical

Model	A	B	σ_ϵ	σ_η
Non-hierarchical noise models				
Informed	124 (5.5%)	4 (0.18%)	8 (0.36%)	665 (29.87%)
Wide	802 (36%)	3 (0.13%)	4 (0.17%)	627 (28.17%)
NH	0 (0%)	0 (0%)	3 (0.13%)	549 (24.66%)

3.3.1 Model comparison

PSIS LOO-CV was also performed for the models using the second dataset. The ELPD and the corresponding standard error are shown in fig. 8. The comparison shows that both hierarchical configurations for the learning parameters had equal performance, while the non-hierarchical approach performs slightly poorer.

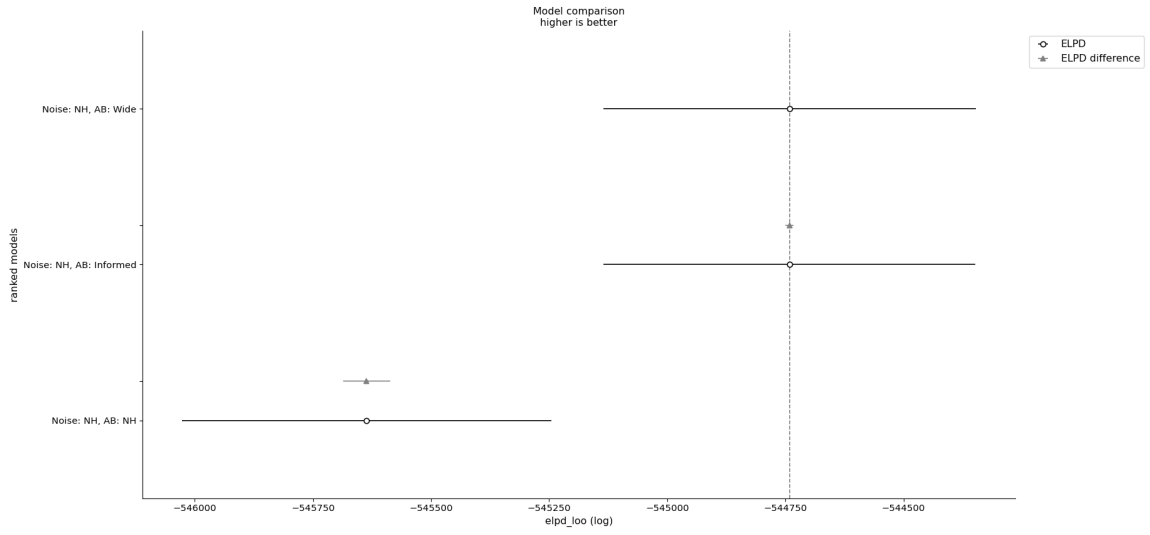

Figure 8: Model comparison using PSIS LOO-CV

3.4 Model evaluation

For the further evaluation of the model, we will focus on the prior set that showed the best performance in tables 2 to 5, which is the model using a non-hierarchical approach for the learning and noise parameters.

3.4.1 Posterior predictive checks

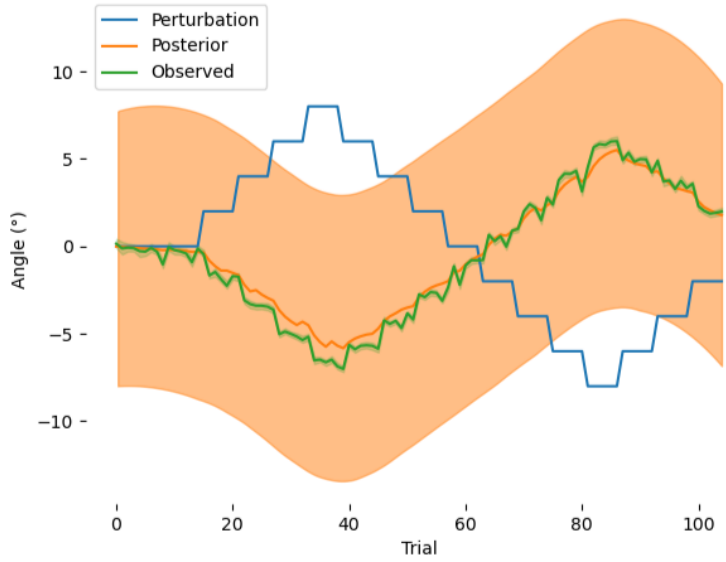

Figure 9: Median reaching angles across subjects per trial compared to the median posterior predictive distribution across subjects estimated by the model using a non-hierarchical design for learning and noise parameters

In order to get a general view of the performance of the model, fig. 9 shows the posterior predictive distribution of the non-hierarchical model when compared to the median reaching angel across subjects. This figure shows that the median predictive posterior reaching angle follows the observed median with reasonable accuracy, but with a large 94% HDI.To get a better idea of the ability of the model to determine the relevant parameters and accurately adapt to a subjects performance, posterior predictive checks were performed on an individual level.

Fig. 10 shows two examples of posterior predictive checks for subjects 1636 (fig. 10A) and 514 (fig. 10B). These are subjects with respectively a very high and very low adaptation rate B, as estimated by the model. A summary of the posterior distributions of the learning and noise parameters for both subjects is shown in table 6. The posterior distributions of these parameters are shown in figs. 10C and D. The posterior predictive checks of the reaching angles accurately follow the observed data show with most of the observed data falling within the 94% HDI of the posterior predictive distributions. However, fig. 10B shows some peaks that fall outside the 94% HDI, which could suggest that the noise factors were underestimated for subjects with high variability in planning or execution noise.

Fig. 11 shows posterior predictive checks for subjects 982 (fig. 11A) and 2110 (fig. 11B) with respectively a high or low estimated total noise (σ_{tot}). A summary of the posterior distributions of the learning and noise parameters for each subject is shown in table 7. The posterior distributions of these parameters is shown in figs. 11C and D. While the median posterior predictive reaching angle

again follows the observed data, fig. 11B again shows that, for subjects with a high noise factor, outliers can fall outside of the 94% HDI of the posterior predictive distribution.

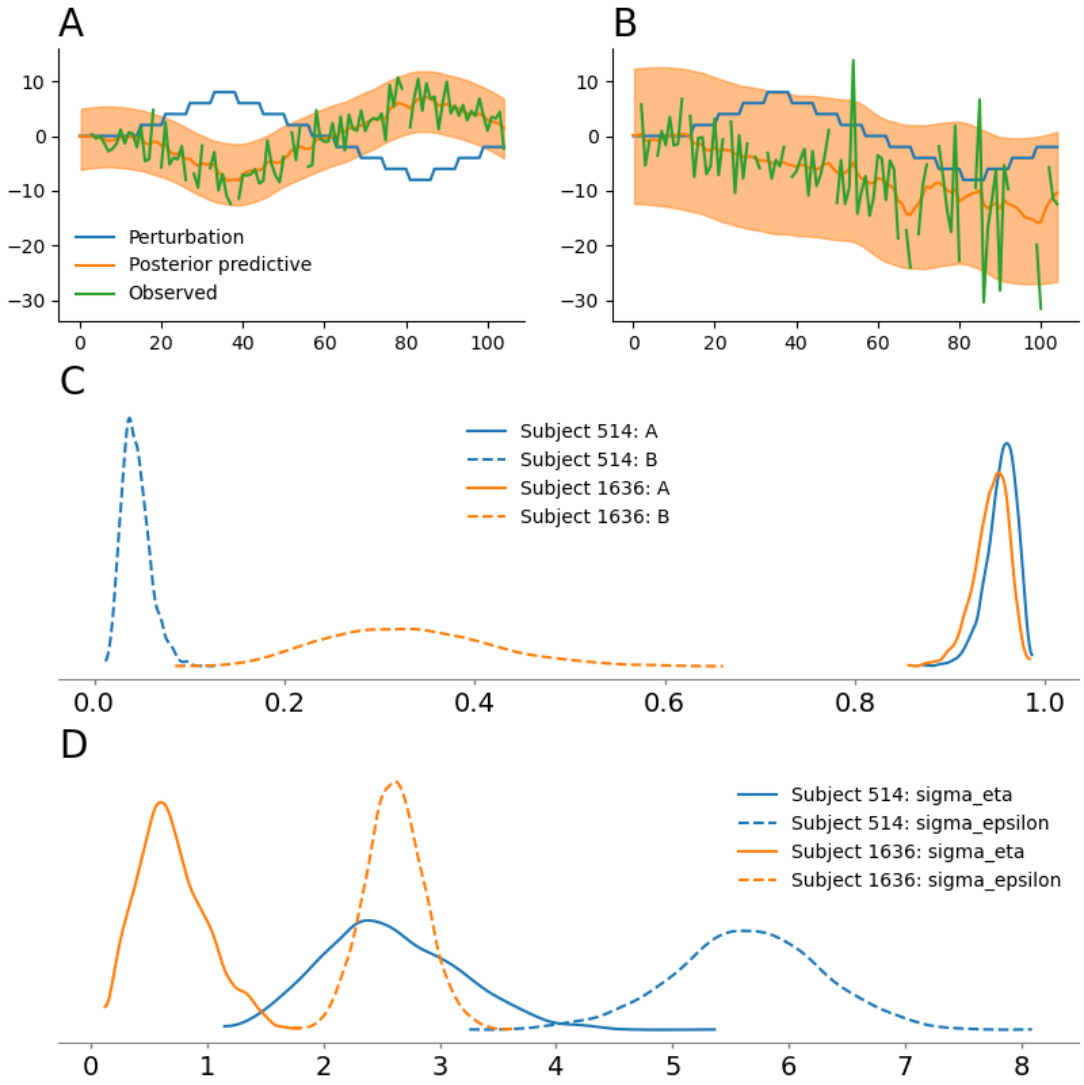

Figure 10: Posterior predictive distribution of reaching angle compared to observed data for subjects 1636 and 514 with respectively a high estimated adaptation rate (Fig. A) and a low estimated adaptation rate (Fig. B). Fig. C: Posterior distributions of learning parameters A and B for subjects 514 and 1636. Fig D: Posterior distributions of noise parameters σ_ϵ and σ_η for subjects 514 and 1636.

ID	A [94% HDI]	$\hat{R}$	B [94% HDI]	$\hat{R}$	σ_ϵ [94% HDI]	$\hat{R}$	σ_η [94% HDI]	$\hat{R}$
514	0.95 [0.92, 0.98]	1.00	0.04 [0.02, 0.07]	1.00	5.64 [4.46, 6.87]	1.01	2.57 [1.53, 3.69]	1.02
1636	0.94 [0.91, 0.98]	1.00	0.33 [0.18, 0.49]	1.00	2.61 [2.14, 3.11]	1.01	0.70 [0.18, 1.25]	1.06

Table 6: Summary of posterior distributions of learning and noise parameters for a subject with a high and low adaptation rate (B)

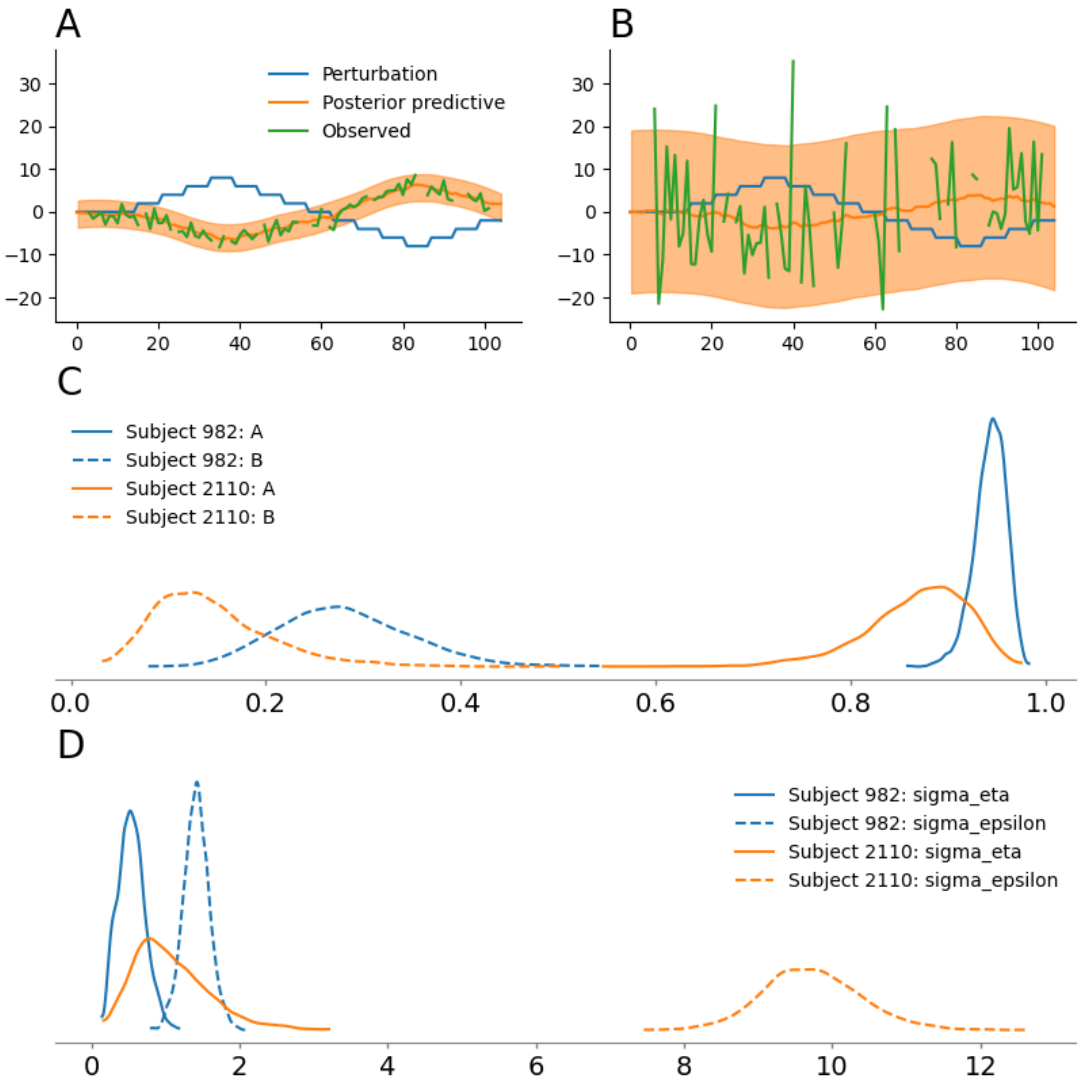

Figure 11: Posterior predictive distribution of reaching angle compared to observed data for subjects 982 and 2110 with respectively a low estimated total noise (Fig. A) and a high estimated total noise (Fig. B). Fig. C: Posterior distributions of learning parameters A and B for subjects 982 and 2110. Fig D: Posterior distributions of noise parameters σ_ϵ and σ_η for subjects 982 and 2110.

ID	A [94% HDI]	$\hat{R}$	B [94% HDI]	$\hat{R}$	σ_ϵ [94% HDI]	$\hat{R}$	σ_η [94% HDI]	$\hat{R}$
982	0.94 [0.91, 0.97]	1.00	0.28 [0.16, 0.41]	1.00	1.42 [1.09, 1.77]	1.00	0.54 [0.22, 0.87]	1.02
2110	0.87 [0.77, 0.96]	1.00	0.15 [0.05, 0.26]	1.00	9.73 [8.44, 10.92]	1.00	1.08 [0.27, 1.98]	1.1

Table 7: Summary of posterior distributions of learning and noise parameters for a subject with a high and low total noise (σ_{tot})

3.4.2 Model efficiency

The model efficiency was checked by determining the marginal energy density and energy transition density for each chain. These distributions are shown in fig. 12, with the corresponding BFMI per chain shown in the legend. When compared to the energies of an efficient Hamiltonian Markov chain shown in fig. 3, it is clear that this model has much more difficulty exploring the entire target distribution. This is further evidenced by the low BFMI for each chain. However, no divergent transitions were encountered. Therefore it is still assumed that the sampling of the Hamiltonian

Markov chain is sufficient.

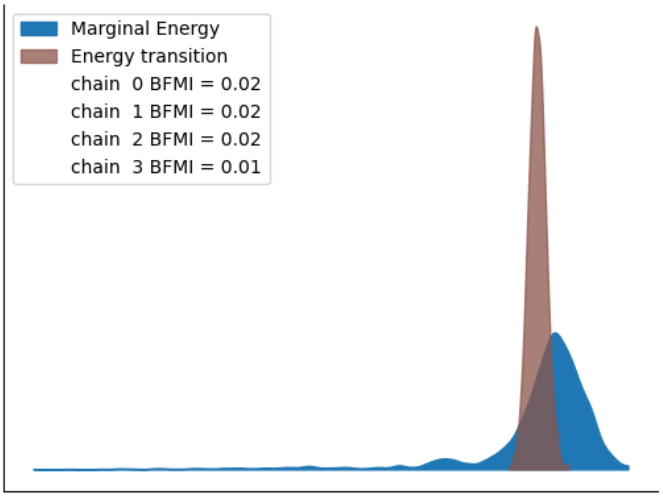

Figure 12: Marginal energy density and energy transition density distributions for each chain of the model using non-hierarchical priors for both the learning and noise parameters.

4. Discussion

In this study, a Bayesian hierarchical model was created that can use data on reaching angles from a reaching task to infer the learning rate, retention rate, planning noise and execution noise for a large number of subjects. The Python code is freely accessible at https://github.com/rickvandervliet/Bayesian-behavioral-toolbox
The model shows little difference in performance when used for a relatively small amount of trials per subject, when compared to a dataset with a more conventional number of trials. Several approaches were used to approximate noise and learning parameters using reaching task data. While none of these approaches produced perfect results, the model using a non-hierarchical approach for both the learning and noise factors showed the most promising results.

However, the results show that the model can still be improved, as none of the models show complete convergence for all parameters for every subject. This issue is most clear for the planning noise σ_η, with a lack of convergence across chains ($\hat{R} > 1.1$) for all approaches used for both datasets. It is possible the complexity of the model makes it difficult for the algorithm to sample all parameters higher in the hierarchy. As σ_η is a comparatively small part of the total noise, this could be an explanation for the difficulty the model has to sample this parameter.

The most straightforward path to resolve this, is to increase the number of samples taken by the HMC algorithm or by increasing the number of chains running simultaneously. This will, however, increase the running time and file size of the model results. For smaller datasets this will not be an issue, but for more complex models with large datasets, this will quickly become unmanageable as every added step will add a data point for every variable, if applicable for every subject and every time point in the dataset. When applied to the hierarchical noise model with Gamma priors (2.1.2) using the smaller dataset (60x900), every added sampling step adds 216.368 extra data points.
The issue of the large file size could be resolved by thinning the chains, meaning that only every ***n***-th sample is saved and discarding the rest. An advantageous side effect of thinning chains is that autocorrelation is also reduced. (32) However, thinning reduces the precision with which the results are summarized and in most cases it is not the most efficient way to reduce autocorrelation. (33, 34)
While it was expected for the hierarchical approaches to perform better, due to the added robustness, it is possible that the higher complexity of the models decreased inference speed and decreased the ability for the model to converge quickly. This effect was further amplified by the large number of data points in the GenR dataset. The increased computation time and bigger file sizes meant that the number of samples taken by the algorithm had to be reduced. However, as the posterior distributions of the hierarchical and non-hierarchical approaches for the learning parameters are similar, it is unlikely that the non-hierarchical model has suffered from decreased robustness.
As we can see the samples of the learning parameters converge more quickly than the noise parameters and show a small HDI, it could be advantageous to continue sampling for only the noise parameters, while using the median values for A and B as a constant value. This will likely increase the effective sample size of the noise parameters, while impacting the running time less than increasing the number of draws for the whole model would.
Another method to reduce dimensionality and improve efficiency is to use a pseudo-marginal Monte

Carlo method (PMMC) by combining the MCMC method used in the models in this paper with a Kalman filter. (35) A known problem with MCMC methods is that it struggles with high dimensionality and does not take advantage of the sequential structure of the observed data, making the algorithm less efficient. An ensemble Kalman filter uses an ensemble of states to approximate the mean and covariance of the posterior distribution. These can then be used for a proposed distribution the NUTS algorithm can sample from. This will improve computational efficiency and improve convergence of the model. (36–39)
The hierarchical parametrizations of the learning factors also show a problem in some of the model designs, where the learning parameters converge to a single number for all subjects. This is likely a case of funnelling, where the sampling algorithm gets stuck in an area of high curvature. (24) This effect could be reduced with a non-centred reparametrization for the learning parameters, as was used for one of the hierarchical approaches for the noise parameters.

While the model did not perform equally well for all subjects, for over 75% of the subjects in the large GenR dataset, this model has produced usable data on adaptation and retention and execution and planning noise. This data can be used for further research into factors that might influence motor adaptation, such as neurodevelopmental disorders, structural differences or possible genetic determinants. Several studies have shown that neurodevelopmental disorders like attention-deficit/hyperactivity disorder (ADHD) or autism spectrum disorder (ASD) impairs visuomotor adaptation to sudden perturbations. (40, 41) It is suggested that this impairment is related to differences in cerebellar structure and cerebellar- cortical pathways, leading to changes in the way sensory feedback is processed and used during motor adaptation. (42–44) As this study shows that the designed model is capable of assessing learning parameters on large datasets, the model can be used for further research into the influences of these disorders on motor adaptation.

Bibliography

[1] Magill RA, Anderson DI. Motor learning and control: concepts and applications. 10th ed. McGraw-Hill; 2014.

[2] Shadmehr R, Smith MA, Krakauer JW. Error correction, sensory prediction, and adaptation in motor control. Annu Rev Neurosci. 2010;33(1):89-108.

[3] Christiansen L, Thomas R, Beck MM, Pingel J, Andersen JD, Mang CS, et al. The beneficial effect of acute exercise on motor memory consolidation is modulated by dopaminergic gene profile. J Clin Med. 2019 Apr;8(5):578.

[4] Cárdenas-Morales L, Grön G, Sim EJ, Stingl JC, Kammer T. Neural activation in humans during a simple motor task differs between BDNF polymorphisms. PLoS One. 2014 May;9(5):e96722.

[5] Paus T. Population neuroscience: why and how. Hum Brain Mapp. 2010 Jun;31(6):891-903.

[6] Paus T, Brook JR, Keyes K, Pausova Z. Principles and advances in Population Neuroscience. Springer Nature; 2024.

[7] White T, Muetzel RL, El Marroun H, Blanken LME, Jansen P, Bolhuis K, et al. Paediatric population neuroimaging and the Generation R Study: the second wave. Eur J Epidemiol. 2018 Jan;33(1):99-125.

[8] Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van IJzendoorn MH, et al. The Generation R Study: design and cohort update 2017. Eur J Epidemiol. 2016 Dec;31(12):1243-64.

[9] Hofman A, Jaddoe VWV, Mackenbach JP, Moll HA, Snijders RFM, Steegers EAP, et al. Growth, development and health from early fetal life until young adulthood: the Generation R Study. Paediatr Perinat Epidemiol. 2004 Jan;18(1):61-72.

[10] Wolpert DM, Diedrichsen J, Flanagan JR. Principles of sensorimotor learning. Nat Rev Neurosci. 2011;12(12):739-51. doi:10.1038/nrn3112.

[11] Van der Vliet R, Frens MA, de Vreede L, Jonker ZD, Ribbers GM, Selles RW, et al. Individual differences in motor noise and adaptation rate are optimally related. eNeuro. 2018 Jul;5(4).

[12] Van Beers RJ, Haggard P, Wolpert DM. The role of execution noise in movement variability. J Neurophysiol. 2004 Feb;91(2):1050-63.

[13] Cheng S, Sabes PN. Modeling sensorimotor learning with linear dynamical systems. Neural Comput. 2006 Apr;18(4):760-93.

[14] Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Stat Comput. 2014 Nov;24(6):997-1016.

[15] Albert ST, Shadmehr R. Estimating properties of the fast and slow adaptive processes during sensorimotor adaptation. J Neurophysiol. 2018 Apr;119(4):1367-93.

[16] Allenby GM, Rossi PE, McCulloch RE. Hierarchical Bayes Models: A Practitioners Guide. Econometrics eJournal. 2005.

[17] Ferrea E, Franke J, Morel P, Gail A. Statistical determinants of visuomotor adaptation along different dimensions during naturalistic 3D reaches. Sci Rep. 2022 Jun;12(1):10198.

[18] Andrieu C, de Freitas N, Doucet A, Jordan MI. An Introduction to MCMC for Machine Learning. Mach Learn. 2003;50(1/2):5-43.

[19] Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014;15(47):1593-623.

[20] Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. Phys Lett B. 1987;195(2):216-22. doi:https://doi.org/10.1016/0370-2693(87)91197-X.

[21] Brooks S, Gelman A, Jones G, Meng XL. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC; 2011. doi:10.1201/b10905.

[22] Betancourt M. A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv: Methodology. 2017.

[23] Abril-Pla O, Andreani V, Carroll C, Dong L, Fonnesbeck CJ, Kochurov M, et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Comput Sci. 2023 Sep;9:e1516.

[24] Papaspiliopoulos O, Roberts G. Non-Centered Parameterisations for Hierarchical Models and Data Augmentation. Bayesian Statistics. 2003 01;7:307-26.

[25] Betancourt M, Girolami M. Hamiltonian Monte Carlo for hierarchical models. In: Current Trends in Bayesian Methodology with Applications. Chapman and Hall/CRC; 2015. p. 79-101.

[26] Kass RE, Carlin BP, Gelman A, Neal RM. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. The American Statistician. 1998;52(2):93-100. doi:10.1080/00031305.1998.10480547.

[27] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992 Nov;7(4):457-72.

[28] Vats D, Knudson C. Revisiting the Gelman–Rubin Diagnostic. Stat Sci. 2021 11;36. doi:10.1214/20-STS812.

[29] Betancourt M. Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo. arXiv: Methodology. 2016 04. doi:10.48550/arXiv.1604.00695.

[30] Rubin DB. Multiple Imputation for Nonresponse in Surveys. 99th ed. Rubin DB, editor. Probability & Mathematical Statistics S.. Nashville, TN: John Wiley & Sons; 1987.

[31] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017 Sep;27(5):1413-32.

[32] Kruschke JK. Doing Bayesian Data Analysiss : a tutorial with R, JAGS, and Stan. 2nd ed. Elsevier; 2015.

[33] Link WA, Eaton MJ. On thinning of chains in MCMC. Methods Ecol Evol. 2012;3(1):112-5. doi:https://doi.org/10.1111/j.2041-210X.2011.00131.x.

[34] Owen AB. Statistically Efficient Thinning of a Markov Chain Sampler. J Comput Graph Stat. 2017;26(3):738-44. doi:10.1080/10618600.2017.1336446.

[35] Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. J R Stat Soc Series B Stat Methodol. 2010 Jun;72(3):269-342.

[36] Katzfuss M, Stroud JR, Wikle CK. Understanding the ensemble Kalman filter. Am Stat. 2016 Oct;70(4):350-7.

[37] Katzfuss M, Stroud JR, Wikle CK. Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. J Am Stat Assoc. 2020 Apr;115(530):866-85.

[38] Drovandi C, Everitt RG, Golightly A, Prangle D. Ensemble MCMC: Accelerating pseudo-marginal MCMC for state space models using the ensemble Kalman filter. Bayesian Anal. 2022 Mar;17(1).

[39] Iglesias MA, Law KJH, Stuart AM. Ensemble Kalman methods for inverse problems. Inverse Problems. 2013 mar;29(4):045001. doi:10.1088/0266-5611/29/4/045001.

[40] Loschiavo-Alvares FQ, Benda RN, Lage GM, Nicolato R, Ugrinowitsch H. Executive functions and motor adaptation to predictable and unpredictable perturbations. Percept Mot Skills. 2023 Apr;130(2):581-606.

[41] Bo J, Shen B, Pang Y, Shen J, Lasutschinkow P, Dillahunt A. Do the enhanced errors impact visuomotor adaptation in children with autism spectrum disorder? Exp Brain Res. 2025 May;243(6):135.

[42] Bruchhage MMK, Bucci MP, Becker EBE. Chapter 4 - Cerebellar involvement in autism and ADHD. In: Manto M, Huisman TAGM, editors. The Cerebellum: Disorders and Treatment. vol. 155 of Handbook of Clinical Neurology. Elsevier; 2018. p. 61-72. doi:https://doi.org/10.1016/B978-0-444-64189-2.00004-4.

[43] McCracken HS, Murphy BA, Ambalavanar U, Glazebrook CM, Yielder PC. Sensorimotor integration and motor learning during a novel force-matching task in young adults with attention-deficit/hyperactivity disorder. Front Hum Neurosci. 2022;16:1078925.

[44] McCracken HS, Murphy B, Ambalavanar U, Zabihhosseinian M, Yielder PC. Sensorimotor integration and motor learning during a novel visuomotor tracing task in young adults with attention-deficit/hyperactivity disorder. J Neurophysiol. 2023 Jan;129(1):247-61.

A. Graphical representation of model designs

A.1 Non-hierarchical noise models

Hierarchical learning parameters

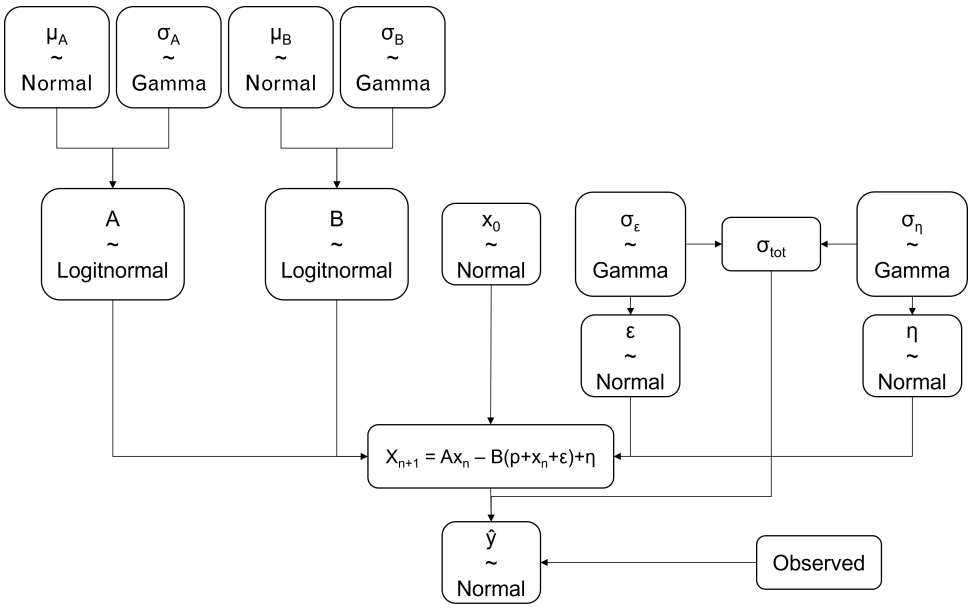

Figure 13: Graphical representation of the designed non hierarchical noise model with a hierarchical approach for the learning parameters A and B

Fully non-hierarchical model

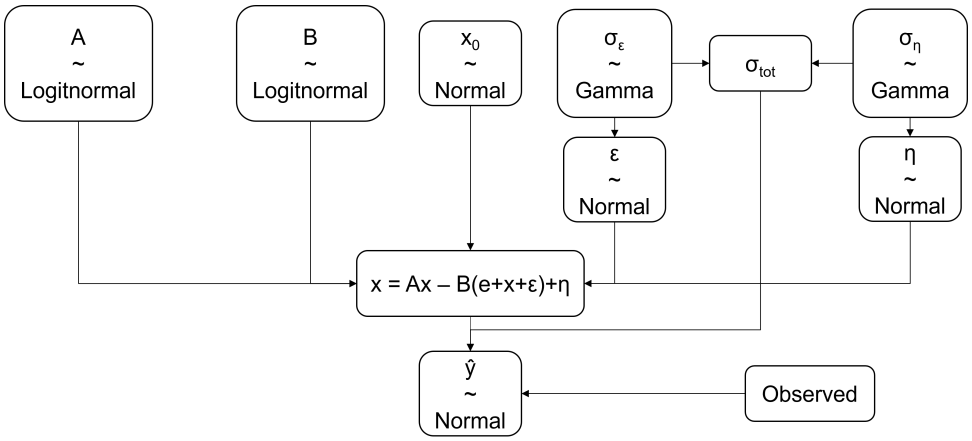

Figure 14: Graphical representation of the designed model with a non-hierarchical approach for both the learning- and noise parameters

A.2 Hierarchical noise models

Gamma

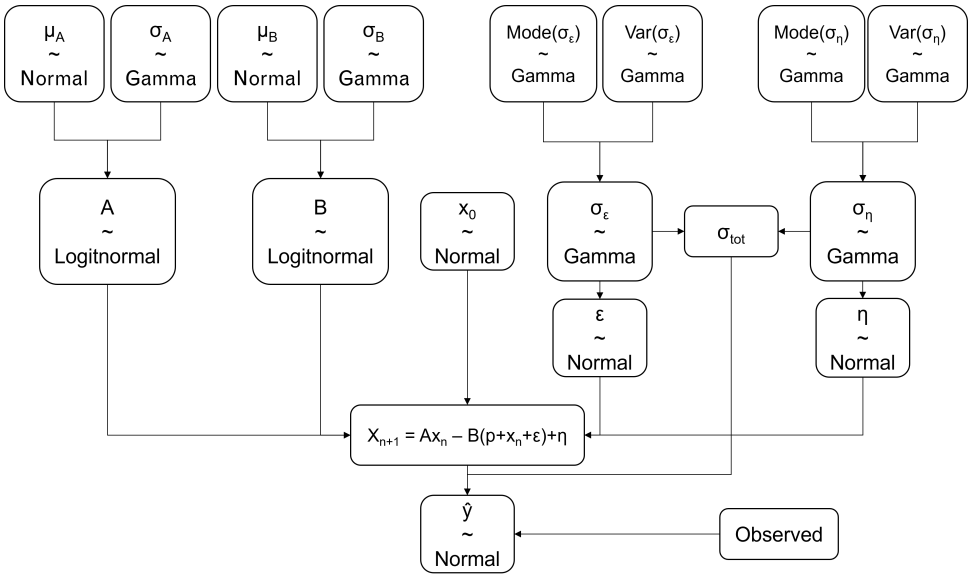

Figure 15: Graphical representation of the model design with informed priors for the learning parameters A and B and gamma priors for the noise parameters

Non-Centered

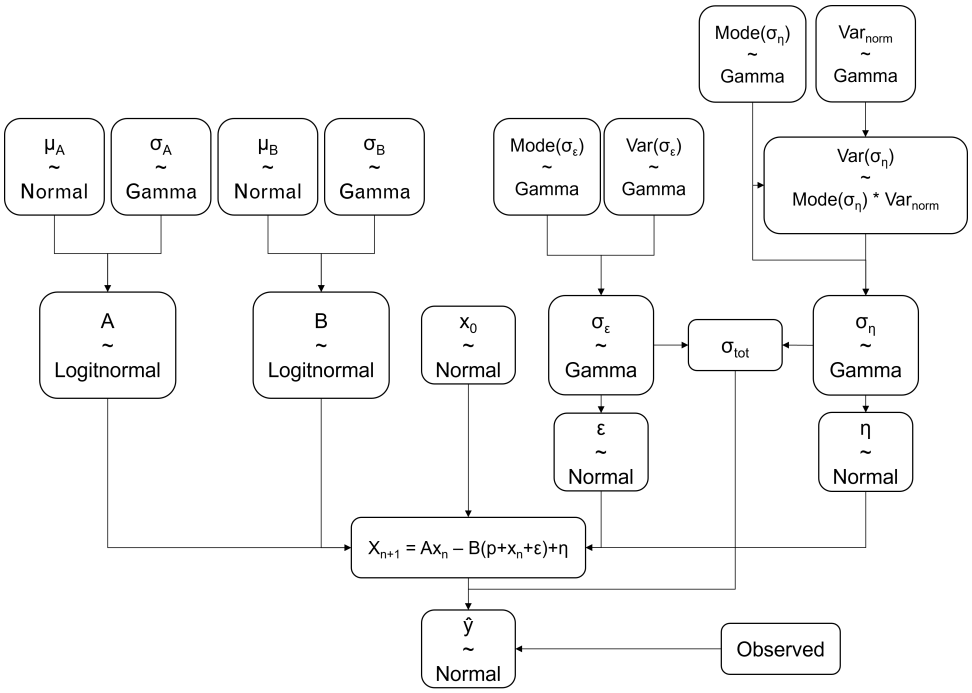

Figure 16: Graphical representation of the model design with informed priors for the learning parameters A and B and a non-centered reparametrization for the noise parameters

Probability ratio

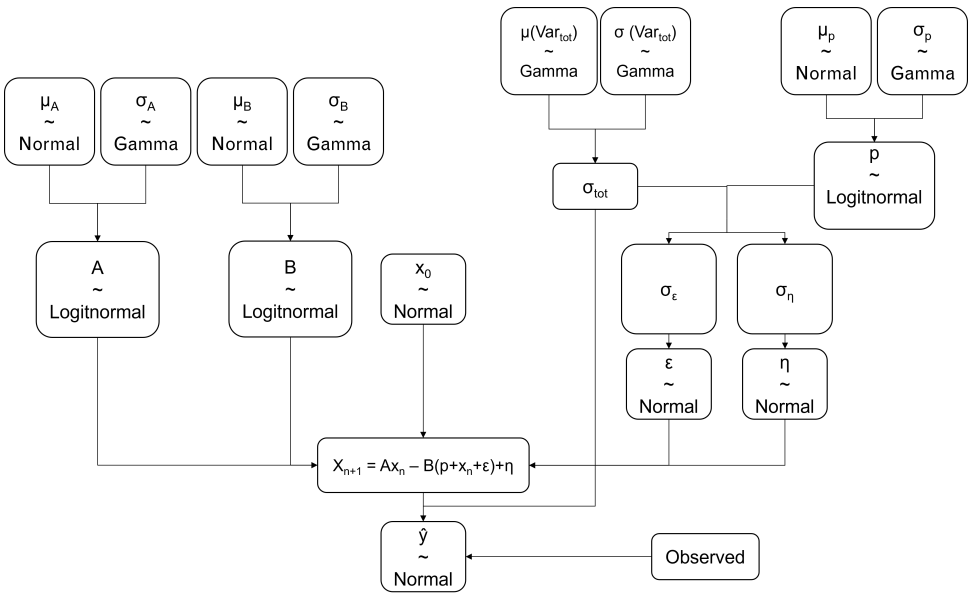

Figure 17: Graphical representation of the model design with informed priors for the learning parameters A and B and the noise parameters defined as components of the total noise