
Optimal design of motor learning experiments informed by Monte-Carlo simulation

Pritesh N. Parmar* and James L. Patton

Richard and Loan Hill Department of Bioengineering,
University of Illinois at Chicago,
851 South Morgan Street, Chicago,
Illinois, 60607, USA

and

Shirley Ryan AbilityLab (Formerly RIC),
355 East Erie Street, Chicago,
Illinois, 60611, USA

Email: pparma2@uic.edu

Email: pattonj@uic.edu

*Corresponding author

Abstract: In our field of neuro-rehabilitation, it is often difficult to have patients endure a long session of training, and we seek an estimate of minimum practice trials. In such cases, motor learning measurements across trials are usually exponentially decaying transient signals. Here we employed Monte-Carlo methods to determine the minimum number of samples required from transient responses. We tracked the accuracy of recovery of synthesised data to reveal a prescription for the minimum number of samples for a robust identification of the underlying learning process, given preliminary estimates of the time constant and noise levels. Our results revealed a systematic relationship for the minimum number of samples required from transient signals that can be used as a stopping criteria for data collection. We also evaluated these results by using a past motor learning study and determining the minimum number of required samples (trials) to best estimate learning curves.

Keywords: optimal design; motor learning; exponential regression; first-order system; Monte-Carlo method; point of diminishing returns; sample size; signal-to-noise ratio; time constant; transient response.

Reference to this paper should be made as follows: Parmar, P.N. and Patton, J.L. (2021) 'Optimal design of motor learning experiments informed by Monte-Carlo simulation', *Int. J. Experimental Design and Process Optimisation*, Vol. 6, No. 4, pp.289–303.

Biographical notes: Pritesh N. Parmar is currently a researcher at the Shirley Ryan AbilityLab. He received his Bachelor's, Master's, and Doctoral degrees in Bioengineering from the University of Illinois, Chicago in 2009, 2011, and 2020, respectively. His research interests include neuro-robotic-rehabilitation, motor control and motor learning, human-machine/human-robot interactions, haptics and virtual-reality technologies, therapeutic and assistive interactive technologies, motor performance enhancement, optimisation, optimal control, machine learning, artificial intelligence, system-identification and control, data-driven modelling and simulation.

James L. Patton is a Professor of Bioengineering at the University of Illinois, Chicago and a Senior Research Scientist at the Shirley Ryan AbilityLab. He received his Bachelor's in Mechanical Engineering and Engineering Science from the University of Michigan in 1989, Master's in Theoretical Mechanics from the Michigan State in 1993, and a Doctoral in Biomedical Engineering from the Northwestern University in 1998. His research interests include neurorehabilitation through novel interface, automatic feedback control theory, dynamics and simulation and related modelling techniques, optimisation, mechatronics and haptic development, human-machine interfaces, robotic teaching, control of balance in locomotion, sensorimotor intelligence.

1 Introduction and background

In an age where a very large amount of empirical data is often employed to identify models (Ljung, 2017), there are still many situations where data is costly, time-consuming, or otherwise difficult to gather in large quantities. The initial challenge is to determine the minimum amount of data needed. This is true in a variety of application areas, including neuro-rehabilitation and motor learning, neural adaptation, neural coding (Kim et al., 2011; Lewi et al., 2009), computer adaptive testing, and gains in human performance (Ding et al., 2018; Kim et al., 2017; Suzuki and Furuta, 2012). Moreover, data is always accompanied by some level of unavoidable noise, which adversely affects model estimations. To overcome uncertainty, multiple measurements and sometimes repeated experiments are required to improve model estimations. In our field of neuro-rehabilitation, it is often difficult to have a person endure a long session of training, and we seek an estimate of minimum practice trials. In such cases, motor learning measurements across trials are usually exponentially decaying transient signals. We postulate that there is a systematic limit on the number of samples required to robustly model the learning process.

Even when measurement samples span a broad range of variation of system behaviours, each successive observation yields progressively smaller gain in the model estimation (Ljung and Söderström, 1983). Initially, each new observation provides relatively novel information, resulting in significant marginal improvement for the estimation. However, as more samples are acquired, the estimation may not continue to improve because the data observed so far have sufficiently captured system behaviours, and any additional observation provides redundant information. Furthermore, the sample acquisitions are sometimes costly, and acquiring samples long after system behaviours have plateaued often leads to overfitting of the noise. Hence the identification might suffer from too much data. Therefore, a guiding framework for the samples required can greatly benefit the model identification.

The simplest identification method employs a first-order linear model at an operating point to gain primary knowledge about important system characteristics such as dominating time constants, time delay, and stationary gain (Ljung, 2017). For linear time-invariant systems, the step and impulse functions are used to excite transience that reveals system characteristics. The initial value problems where a system responds to initial conditions also exhibit transience. The time-series of these transient responses can be fitted to an exponential function:

$$y(t) = A * \exp(-t/\tau) + B + \mathcal{N}(0, \sigma^2), \quad (1)$$

where A , B and τ are related to the characteristic system parameters and \mathcal{N} represents Gaussian noise with zero mean and σ^2 variance. Here if less than three time-series samples of the transient responses are acquired, the first-order approximation of a system is underdetermined. Only in an ideal case when there is no observation noise, the system parameters can be approximated using three samples. However, in the presence of noise, more than three samples must be acquired to approximate the system parameter values with certainty by also estimating the noise statistics from the observations. So, there must be a point of diminishing returns for the number of such time-series observations beyond which the estimation of the noise statistics and the characteristic parameter values do not significantly improve.

It is well recognised that the accuracy of parameter estimation depends on the signal-to-noise ratio (SNR), i.e., the level of the desired signal to the level of noise. Higher SNR yields better estimates. In the transient response, the desired signal is detectable above the noise level immediately after the onset of the step or impulse inputs. However, as the transient decays, the influence of noise becomes more dominant, deteriorating fit quality. Therefore, early observations should be preferred for such model estimation. Additionally, the rate at which the transient response decays depends on system characteristics, and thus it is conceivable that the time constant, τ , would also affect the minimum number of samples required (see Appendix for Fisher information analysis for this problem).

In this simulation paper, we attempt to identify the point of diminishing returns for the number of samples from transient responses beyond which the exponential regression does not significantly improve. We assumed normalised transient responses and generated data for various combinations of τ and SNR . We then tracked the improvement in accuracy of the regression as we incrementally included additional samples. We repeated the regression analysis on independent randomly sampled data using Monte-Carlo methods. The results show a systematic relationship between the time constant and SNR of transient responses and the point of diminishing returns for the number of samples. We also evaluated these results by using a past motor learning study and determining the minimum number of required samples (trials) to best estimate learning curves of a motor skill.

2 Methods

The first-order transient responses can assume various forms, but they can be normalised to exponential decay functions. Consider the following normalisation of equation (1):

$$(y(t) - B)/\sigma = (A/\sigma) * \exp(-t/\tau) + \mathcal{N}(0, \sigma^2)/\sigma, \quad (2)$$

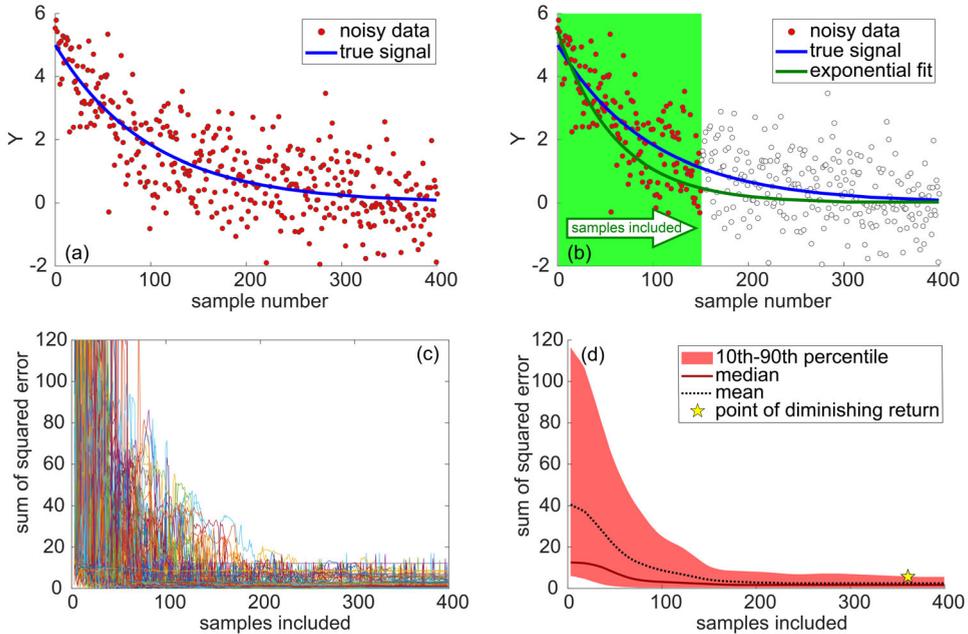
$$Y(t) = SNR * \exp(-t/\tau) + \mathcal{N}(0, 1), \quad (3)$$

where $Y(t)$ is the normalised transient response with normally distributed random noise $\mathcal{N}(0, 1)$ and A/σ ratio is SNR .

Using equation (3), we generated data using known values of τ and SNR for $t = 0$ to 400 with an increment of 1 (sample). This we denote as the *true signal* [Figure 1(a)]. τ

values ranged from 5 to 30 with increments of 1 and from 30 to 100 with increments of 5. SNR values ranged from 0.5 to 3 with increments of 0.1 and from 3 to 10 with increments of 0.5. We then added normally distributed random noise (mean 0 and standard deviation 1) to the true signal.

Figure 1 Summary of methods used to identify the point of diminishing returns for the number of samples required from the transient responses to accurately perform exponential regression (see online version for colours)



- Notes: (a) We first generated a true signal (blue curve) using equation (3) and known SNR and τ values. Next, we added normally distributed random noise to the true signal to generate data (red circles).
 (b) We repeatedly performed the exponential regression by incrementally including an additional data sample (green region).
 (c) We calculated the accuracy of the regression per each sample included using the SSE between the fitted curve and the true signal across all 401 samples. Using the Monte-Carlo method, we repeated the regression analysis on 100 independent random samplings of added noise. Each coloured curve shows SSE for a Monte-Carlo repeat.
 (d) 10th–90th percentile distribution, median, and mean of SSE across all Monte-Carlo repeats are shown as shaded red region, dark red line and dotted black line, respectively. We identified the point of diminishing returns as the number of samples included when the 90th percentile of SSE crossed 105% of its minimum (yellow star).

We began nonlinear least-squares regression using equation (3) on the first three data samples. We then incrementally included a sample and performed regression as we tracked the accuracy [Figure 1(b)]. The regression was performed using the Levenberg-Marquardt method within a three-fold simulated annealing loop. We repeated the regression analysis on 100 independent random samplings of added noise using the Monte-Carlo method. Overall, this resulted in 67,071,900 independent regressions (41 τ

levels \times 41 SNR levels \times 399 incrementally including an additional sample \times 100 Monte-Carlo repeats). We calculated the accuracy of the regression per each sample included and per each Monte-Carlo loop using the sum of squared error (SSE) between the fitted curve and the true signal across all 401 samples [Figure 1(c)].

Our objective was to measure the point of diminishing returns for the number of samples included in the regression beyond which the accuracy does not significantly improve. We calculated the 90th percentile of SSE across all Monte-Carlo repeats [Figure 1(d)]. We used the 90th percentile of SSE over mean because the distribution of SSE per each sample included was not normally distributed. Furthermore, the 90th percentile of SSE not only captured the accuracy of the regression but also accounted for the variability due to random noise sampling across Monte-Carlo loops. We filtered the 90th percentile of SSE twice using a moving average with a centred-sliding-window of 31 samples. Next, we identified the point of diminishing returns (as the number of samples included) when the 90th percentile of SSE crossed 105% of its minimum (N_{opt}). We used 105% of minimum because we noticed that oftentimes the 90th percentile of SSE had a long tail where the SSE did not decrease significantly. A similar metric known as the settling time is often used in control theory that is defined as the time required for the signal to reach within range of a certain percentage (usually 5%) of the final value (Tay et al., 1998).

We used an affine ANOVA model to analyse whether N_{opt} had significant non-zero slopes (alpha = 0.05) along SNR and τ . We found that N_{opt} had a linear trend with τ and an exponential trend with SNR. Therefore, we fitted a surface to N_{opt} with the following functional form:

$$N_{opt}(SNR, \tau) = x_1 + x_2 * \tau + (x_3 * \tau + x_4) * \exp(x_5 * SNR), \quad (4)$$

where x_1, x_2, x_3, x_4, x_5 are coefficients, and SNR and τ are known values of the normalised transient signals for which N_{opt} was measured.

We also measured the accuracy of the regression using the 90th percentile of SSE at the point of diminishing returns (A_{opt}). We used an affine ANOVA model to analyse whether A_{opt} had significant non-zero slopes (alpha = 0.05) along SNR and τ . We found linear trends along both SNR and τ . Thus, we fitted a linear surface to A_{opt} with the following functional form:

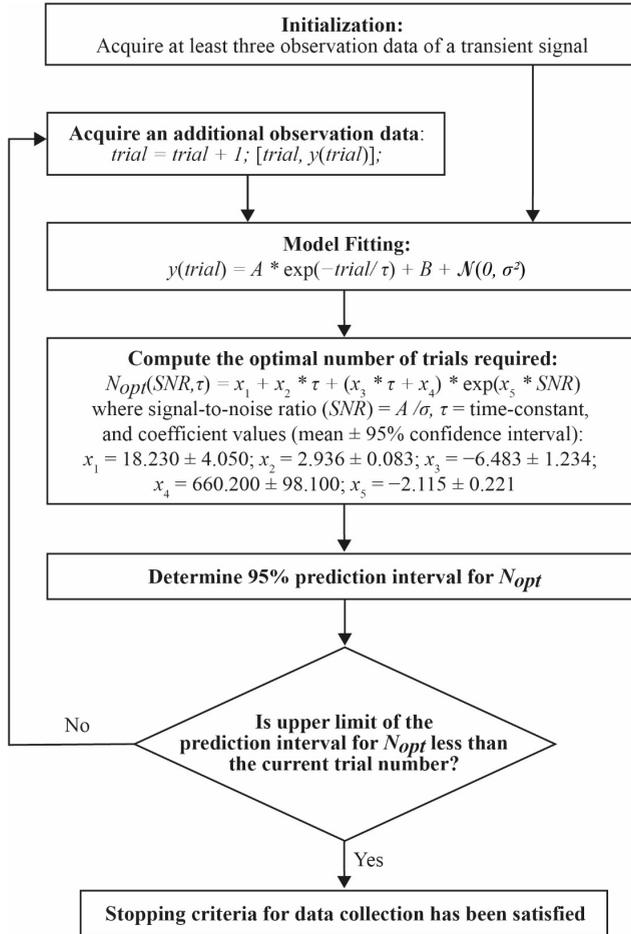
$$A_{opt}(SNR, \tau) = y_1 + y_2 * SNR + y_3 * \tau, \quad (5)$$

where y_1, y_2, y_3 are coefficients, and SNR and τ are known values of the normalised transient signals for which A_{opt} was measured. The surfaces were fitted using nonlinear least-squares regression with the Trust-Region method. All analysis was performed using MathWorks MATLAB 2017b.

Finally using dataset from a past study (Patton et al., 2013), we evaluated the results about the optimal number of required samples (trials) to best estimate learning curves of a motor skill. Briefly, 16 healthy individuals were trained to make reaching movements with 30 degrees of visuomotor rotation, and the goal of the study (Patton et al., 2013) was to test the effects of various error-augmenting feedback on the rate and extent of learning. The progression of visuomotor adaptation within the training phase was tracked by measuring the maximum perpendicular error between reaching movements and ideal straight-line (from start to goal position) at intermittent ‘catch’ trials. Such motor learning measurements across trials are usually exponentially decaying transient signals. While

the training session in that study (Patton et al., 2013) involved 406 total trials, our goal was to determine whether a lesser (minimal) amount of trials would have sufficed to fit the learning curves.

Figure 2 Flowchart for our resulting stopping criteria for data collection from transient signals



In order to determine the optimal number of samples (trials) from a transient signal using equation (4), prior knowledge about underlying time constant and *SNR* is required. However, these can be sequentially calculated from the so-far observed data as the data is being acquired. For our evaluation of a past motor learning dataset (Patton et al., 2013), we started with the first three movement error data from the training phase and performed regression using equation (1) for each subject. Next, we estimated *SNR* and time constant from the fitted parameters as in equations (2) and (3). We applied a lowpass Butterworth filter (first-order, zero-phase, normalised cut-off frequency of 0.02) to *SNR* and time constant along the number of trials included in the regression. Next, using equation (4), we estimated a 95% prediction interval for the optimal number of trials required (N_{opt}). Next, we incrementally allowed additional trial observation data in the regression while the upper limit of the prediction interval was greater than the number of trials included in

the regression. We used the upper limit of the prediction interval because it provided the largest statistical estimate for the optimal number of trials required. We set the stopping criteria for further inclusion of data in the regression when the upper limit of the prediction interval was less than the number of trials included in the regression (N_{opt}^*). Note that for this particular dataset (Patton et al., 2013), movement errors were intermittently sampled across trials, and thus regression estimates did not change during the trials gaps when observation data was not available. A flowchart in Figure 2 summarises this method for determining stopping criteria for data collection from transient signals. Finally, we computed percent error using difference between the estimated model parameter values at the stopping criteria (N_{opt}^*) and the best possible model parameter values. The best possible model parameter values were estimated using the regression on all 406 trial observations.

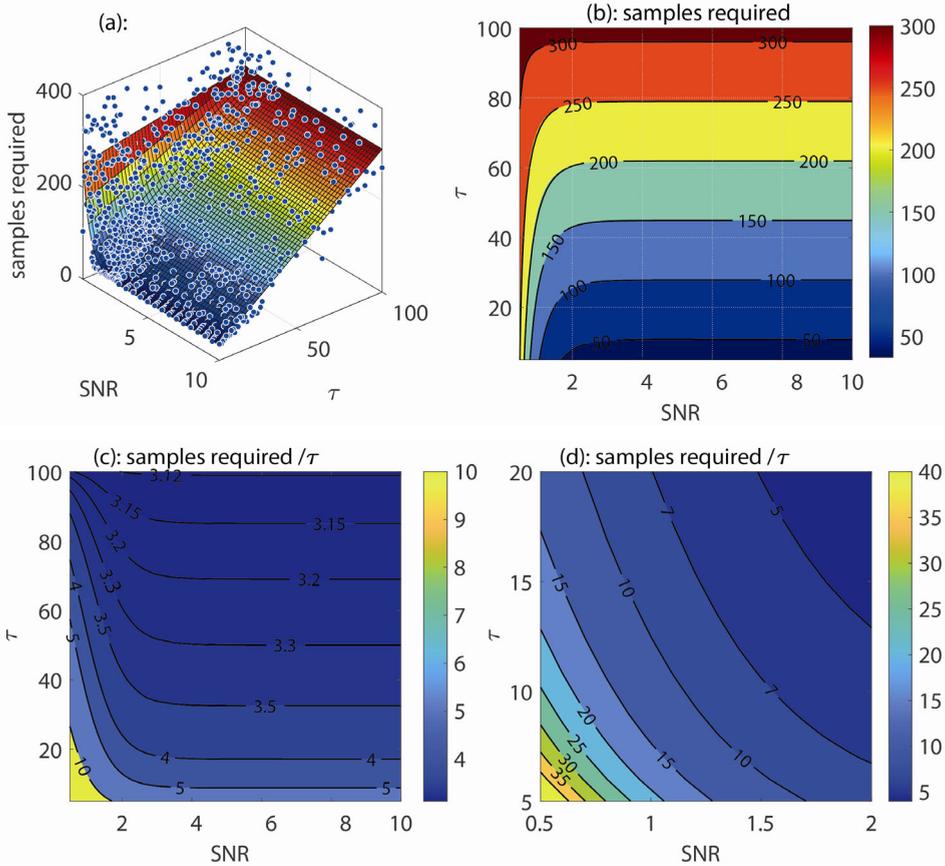
3 Results

We incrementally included samples of data and performed regressions to identify the point of diminishing returns beyond which the accuracy of estimating the true transient signal does not significantly improve. The accuracy as measured by SSE between the fitted signal and the true signal improved as more sample data were included in the regression. However, as seen in an example in Figure 1(c), SSE did not monotonically decrease. We also observed variation in SSE due to random samplings of the true signal across the Monte-Carlo repeats. This variation in SSE across Monte-Carlo repeats decreased as more sample data were included in the regression.

We identified the point of diminishing returns (as the number of samples included) when the 90th percentile of SSE crossed 105% of its minimum. The results of the point of diminishing returns for various combinations of the true signal's SNR and τ are shown in Figure 3(a). We performed ANOVA on the point of diminishing returns results and found that there were significant non-zero slopes along SNR ($p = 6.3e-48$) and along τ ($p = 0$). Furthermore, we found an exponential trend along SNR through visual inspection. Using equation (4), we fitted the results to a surface ($R^2 = 81.46\%$) as shown in Figures 3(a) and 3(b). The coefficient values were $x_1 = 18.230 \pm 4.050$ (mean \pm 95% confidence interval), $x_2 = 2.936 \pm 0.083$, $x_3 = -6.483 \pm 1.234$, $x_4 = 660.200 \pm 98.100$, $x_5 = -2.115 \pm 0.221$. The samples required at the point of diminishing returns increased linearly with τ and decreased exponentially with SNR . We also calculated the ratio of the samples required to time constant [Figures 3(c) and 3(d)]. This ratio remained reasonably constant (about 3.12–4) when the time constant and SNR were greater than 20 trials and 2, respectively.

Accuracy of the regression using the 90th percentile of SSE at the point of diminishing returns is shown in Figure 4(a). We Performed ANOVA on the accuracy results and found that there were significant non-zero slopes along SNR ($p = 2.8e-10$) and along τ ($p = 6.5e-4$). Using equation (5), we fitted the accuracy results to a surface ($R^2 = 3.15\%$) as shown in Figure 4. The coefficient values were $y_1 = 5.100 \pm 0.065$, $y_2 = -0.037 \pm 0.011$, $y_3 = -0.002 \pm 0.001$. The 90th percentile of SSE at the point of diminishing returns decreased (increase in accuracy) linearly with time constant and SNR .

Figure 3 The point of diminishing returns for the number of samples required from the transient responses to accurately perform exponential regression (see online version for colours)

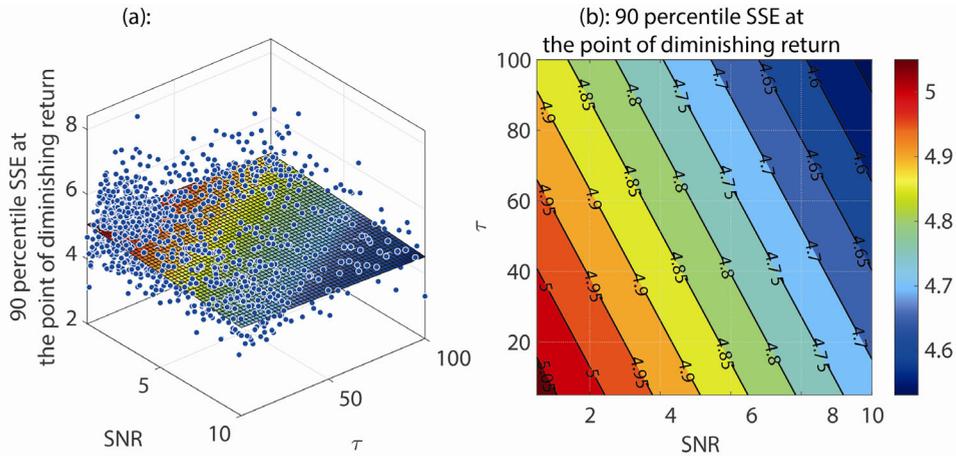


Notes: (a) We fitted the results (blue circles) to a surface ($R^2 = 81.46\%$) using equation (4). The samples required at the point of diminishing returns increased linearly with time constant, τ , and decreased exponentially with SNR.
 (b) Contours of the surface fit.
 (c–d) We also calculated the ratio of the samples required to the time constant. This ratio remained reasonably constant (about 3.12–4) when the time constant and SNR were greater than 20 trials and 2, respectively.

We evaluated dataset from a past study (Patton et al., 2013) to determine minimal trial data to best estimate learning curves of a motor skill using the above results about the optimal number of required samples (trials). This analysis indicated that about the first 195 (median with IQR 147 to 235 across subjects) out of 406 total trials would have sufficed to fit the learning curves (Figures 5 and A1). The optimal number of trials (N_{opt}^*), however, varied greatly across subjects and needed to be determined for each subject as per methods described in Figure 2. The stopping criteria for two subjects were not met within 406 trials, indicating that more trials should have been collected for these individuals. Thus, these two subjects were excluded from the following percent error analysis. At the stopping criteria (N_{opt}^*), the percent error in the model parameter A estimation was 3.33% (median with IQR -0.25 to 6.68 across subjects), in the model

parameter τ estimation was -9.45% (-55.66 to 12.09), and in the model parameter B estimation was 6.46% (-8.87 to 20.31) [Figure 5(f)]. We did not find any significant difference for the percent error when compared against zero for each model parameter (sign test, $p > 0.05$). Furthermore, magnitude of the percent error was negatively correlated with SNR [Spearman correlation coefficient = -0.56 ($p = 0.04$) for parameter A, -0.43 ($p = 0.12$) for parameter τ , -0.13 ($p = 0.65$) for parameter B]. Overall, regression using trial observations less than N_{opt}^* always yielded worse model fitting accuracy [Figure 5(e)].

Figure 4 Accuracy of the regression at the point of diminishing returns, (a) we fitted the accuracy results (blue circles) to a surface ($R^2 = 3.15\%$) using equation (5) (b) contours of the surface fit (see online version for colours)

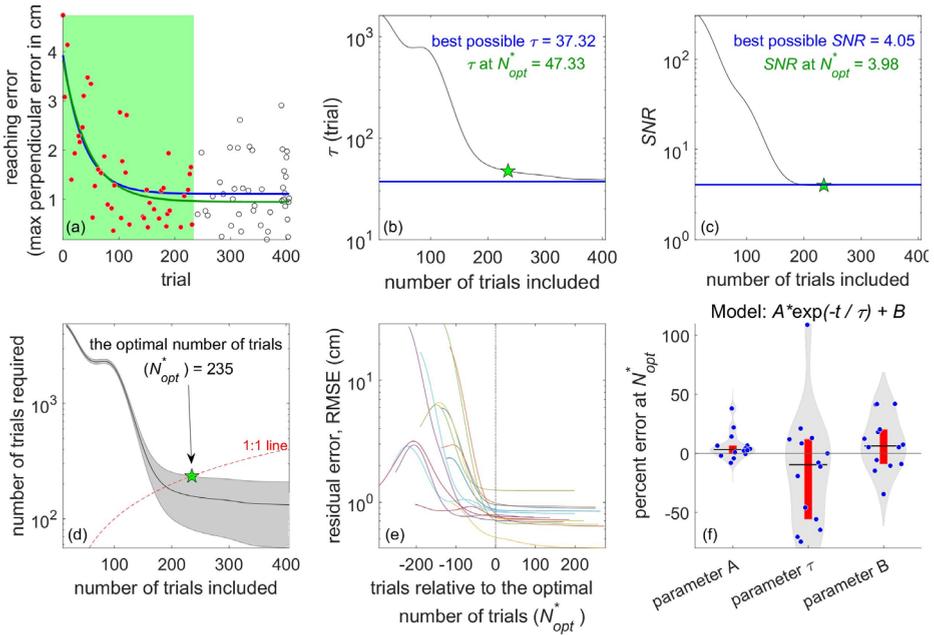


Notes: The accuracy was measured using the 90th percentile of error difference (SSE) between the true transient signal and the regression fit. The accuracy increased (decrease in SSE) linearly with both τ and SNR .

4 Discussion

Our goal was to identify the point of diminishing returns for the number of samples needed to identify the parameter values associated with a first-order transient response. By progressively allowing an additional sample into the regression, we tracked improvements in the accuracy of the regression curve to match the original synthesised signal. Our results reveal a systematic relationship between the number of samples required and the underlying time constant and noise in the data. We found that the number of samples required increased linearly with time constant and decreased exponentially with SNR (Figure 3). Furthermore, when the time constant and SNR were greater than 20 trials and 2, respectively, the number of samples required was consistently about 4 times the time constant. The accuracy of these regressions (at the point of minimum samples required) improved linearly with SNR and time constant (Figure 4). We also evaluated these results by using dataset from a past study (Patton et al., 2013) and determining the optimal number of required samples (trials) to best estimate learning curves of a motor skill (Figure 5).

Figure 5 Evaluation of dataset from a past study (Patton et al., 2013) to determine the minimal number of required samples (trials) to best estimate learning curves of a motor skill (see online version for colours)



Notes: (a) An example of learning curve for a representative subject who practiced reaching with visuomotor rotation. Circles represent movement error during training phase at intermittent catch trials. The blue curve represents exponential regression using data from all trials ($R^2 = 41.07\%$, calculated over all 406 trials), while the green curve represents exponential regression using data from the determined minimal number of trials ($N_{opt}^* = 235$ trials, green shaded regions, red circles; $R^2 = 38.94\%$, calculated over all 406 trials).

(b–c) Time constant and SNR estimates for each incrementally added trial to the exponential regression (black curve). Blue line represents the best possible values from exponential regression using data from all trials.

(d) Number of trials required and its 95% prediction interval using equation (4) (black curve and grey shaded region). N_{opt}^* was determined when upper limit of the prediction interval was less than the current trial number (1:1 line crossing).

Figures 5(a)–5(d) analysis for the same representative subject. Green stars represent the values at N_{opt}^* .

(e) Accuracy with respect to the observed data [e.g., residual error, RMSE between green curve and circles in Figure 5(a)]. Each curve represents a different subject, and each curve has been shifted to show trials relative to N_{opt}^* . Note that Figures 5(b)–5(e) are plotted with logarithmic scales.

(f) Percent error using difference between the estimated model parameter values at the stopping criteria (N_{opt}^*) and the best possible model parameter values from the regression on all 406 trial observations. Each circle represents a subject. Red bars and black lines represent interquartile range and median values of the percent error, respectively. The shaded grey regions represent kernel density estimates of the percent error. We did not find any significant difference for the percent error when compared against zero for each model parameter (sign test, $p > 0.05$).

Overall, this analysis indicated that about the first 195 trials (median with IQR 147 to 235 across subjects) would have sufficed to fit the learning curves, and additional trial data did not improve fit accuracy as seen in Figure 5(e).

The accuracy of the parameter estimation is fundamentally limited by the Fisher information, i.e., the amount of observable information from the sampled data (Dette and Neugebauer, 1997; Efron and Hinkley, 1978; Ljung, 2017) (see Appendix for Fisher information analysis). While higher accuracy was expected with higher *SNR*, it was interesting that the accuracy linearly improved with time constant (Figure 4). This is because of the number of samples (information) available in the earlier, rapidly decaying epochs. After three time constants, an exponential decays to 5% of its initial value, and therefore signal carries poor information about system behaviour. Our results also show that for larger *SNR* (greater than 2) the required number of sample observations should be about four times the time constant, while the required number of samples for lower *SNR* dramatically increases (Figure 3). In cases when signals have low *SNR*, weighted or robust regression and outlier rejections (e.g., Cook's distance) can improve the model fitting. Furthermore, bootstrapping (Jackknife method) can provide better parameter estimates. In any case, data collection less than N_{opt}^* can be detrimental to model fitting, while data collection beyond N_{opt}^* can safeguard against bad model fitting.

Some applications may require a higher accuracy for the parameter estimation at the stopping criteria for collecting further data samples. While the above results for N_{opt} were calculated from the incremental model fitting and identifying the point of diminishing returns (as the number of samples included) when the error across Monte-Carlo Simulations reached within 5% of its minimum value (i.e., when the 90th percentile of SSE crossed 105% of its minimum), we repeated the calculation for N_{opt} with the error reaching within 1% of its minimum value. Using equation (4), we fitted these results to a surface ($R^2 = 74.36\%$), and the coefficient values were $x_1 = 35.47 \pm 5.84$, $x_2 = 3.291 \pm 0.117$, $x_3 = -7.302 \pm 1.398$, $x_4 = 682 \pm 103.1$, $x_5 = -1.826 \pm 0.217$. Using equation (5), we fitted results of the accuracy of the regression at the point of diminishing returns to a surface ($R^2 = 3.20\%$), and the coefficient values were $y_1 = 4.917 \pm 0.062$, $y_2 = -0.03514 \pm 0.01101$, $y_3 = -0.00207 \pm 0.0011$. These set of coefficient values can be used to improve the accuracy at the stopping criteria (N_{opt}^* , Figure 2).

The results presented in this paper are limited to exponential decay processes that are the output of first-order linear systems with additive Gaussian noise. However, more complex systems can be modelled as a combination of such systems, and in many situations, a system can be linearised about an operating point that captures local behaviour. It remains to be seen how much the sample size calculations are affected by signal-dependent noise. Another issue is the possibility of second-order effects that cannot be accounted for from this first-order approximation. Second-order models with imaginary poles can exhibit a rich set of oscillatory behaviour. However, with such behaviour, it may still be possible to extract its dominant first-order characteristics from the envelope of oscillations. Moreover, these methods applied to a very simple model with a very simple input – a step. Such approaches can be applied in the same fashion to more complicated systems.

This sample size fitting problem is due to the discrete nature of data. Such problems often arise when the data acquisition must operate near the limits of the capability of their sampling rate, must monitor transient responses, and must minimise the cost associated with data sample collection. One example is in the study of motor adaptation or learning, where the 'student' makes discrete attempts that lead to a learning curve. Here collecting many trials is costly, and each successive trial offers less information. Our approach for determining the sample size is similar to statistical power analyses. While power is about

estimating the minimum necessary samples to robustly detect (without a type I error) a significant change, the present paper determines the minimum necessary samples to identify a model. Power analyses often will employ similar approaches using synthetic data and Monte-Carlo methods.

5 Conclusions

Some adaptive control applications require rapid identification of system parameters. For example, the so-called ‘human in the loop’ tuning of assistive controllers requires system identification at many different settings (Ding et al., 2018; Kim et al., 2017; Suzuki and Furuta, 2012). Thus, knowing a lower limit on the number of observation samples can suggest a stopping criterion for data acquisition. The outcomes of the current study can also potentially improve the design of rehabilitation protocols by determining the optimal length of training sessions for patients. Our results provide scientists and engineers with valuable knowledge and improve the efficiency of system identification and data-driven approaches.

In any case, our Monte-Carlo method provides a clear set of requirements for the minimum dataset needed to best characterise a transient response (impulse or step) to a first-order linear time-invariant system. This should provide empirical sample size estimates in the application areas such as radioactive decay, heat transfer phenomena, structural and mechanical analysis, fluid dynamics, population dynamics, pharmacokinetics, metabolite production kinetics, drug dose-response or washout analysis, toxicology, learning, and neural adaptation.

Acknowledgements

The research reported in this publication was supported by the National Institutes of Health Awards No. F31-NS100520 and 2R01-NS053606.

Pritesh N. Parmar and James L. Patton conceived and designed the experiments/simulations and wrote the paper. Pritesh N. Parmar performed the experiments/simulations and analysed the data.

We thank Felix Huang and the Robotics Lab at the Shirley Ryan AbilityLab.

References

- Dette, H. and Neugebauer, H-M. (1997) ‘Bayesian D-optimal designs for exponential regression models’, *Journal of Statistical Planning and Inference*, Vol. 60, No. 2, pp.331–349.
- Ding, Y., Kim, M., Kuindersma, S. and Walsh, C.J. (2018) ‘Human-in-the-loop optimization of hip assistance with a soft exosuit during walking’, *Science Robotics*, Vol. 3, No. 15, p.eaar5438.
- Efron, B. and Hinkley, D.V. (1978) ‘Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information’, *Biometrika*, Vol. 65, No. 3, pp.457–483.
- Kim, A.J., Lazar, A.A. and Slutskiy, Y.B. (2011) ‘System identification of drosophila olfactory sensory neurons’, *Journal of Computational Neuroscience*, Vol. 30, No. 1, pp.143–161.
- Kim, M., Ding, Y., Malcolm, P., Speeckaert, J., Siviyy, C.J., Walsh, C.J. and Kuindersma, S. (2017) ‘Human-in-the-loop Bayesian optimization of wearable device parameters’, *PloS One*, Vol. 12, No. 9, p.e0184054.

- Lewi, J., Butera, R. and Paninski, L. (2009) ‘Sequential optimal design of neurophysiology experiments’, *Neural Computation*, Vol. 21, No. 3, pp.619–687.
- Ljung, L. (2017) ‘System identification’, in Webster, J.G. (Ed.): *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp.1–19, <https://doi.org/10.1002/047134608X.W1046.pub2>.
- Ljung, L. and Söderström, T. (1983) *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
- Patton, J.L., Wei, Y.J., Bajaj, P. and Scheidt, R.A. (2013) ‘Visuomotor learning enhanced by augmenting instantaneous trajectory error feedback during reaching’, *PloS One*, Vol. 8, No. 1, p.e46466.
- Suzuki, S. and Furuta, K. (2012) ‘Adaptive impedance control to enhance human skill on a haptic interface system’, *Journal of Control Science and Engineering*, Vol. 2012, pp.1–10, <https://doi.org/10.1155/2012/365067>.
- Tay, T-T., Mareels, I. and Moore, J.B. (1998) *High Performance Control*, Birkhaeuser Boston, Cambridge, MA.

Appendix

Fisher information analysis

Nonlinear model fitting using a signal that drops off exponentially and its consequent loss of valuable information from additional observation data can be studied via Fisher information. If we observe data Y_1, \dots, Y_n at times t_1, \dots, t_n , then the likelihood for the model [equation (3)] is

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\frac{1}{2}\right)\left(Y_i - SNR * \exp\left(-\frac{t_i}{\tau}\right)\right)^2\right) \quad (6)$$

where $\theta = (SNR, \tau)^T$. Then the log-likelihood is

$$\ell(\theta) = c - \frac{1}{2} \sum_{i=1}^n \left(Y_i - SNR * \exp\left(-\frac{t_i}{\tau}\right)\right)^2 \quad (7)$$

for a constant c not involving θ . The variance of the maximum likelihood estimate $\hat{\theta}$ is well described by the Fisher information. The Fisher information is

$$I = -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right)^{-1} \quad (8)$$

and to a good approximation $\text{var}(\theta) \approx I^{-1}$. So, large I means good identification and conversely.

The partial derivatives of this log-likelihood are

$$\frac{\partial \ell}{\partial SNR} = \sum_{i=1}^n \left(Y_i - SNR * \exp\left(-\frac{t_i}{\tau}\right)\right) \exp\left(-\frac{t_i}{\tau}\right) \quad (9)$$

$$\frac{\partial \ell}{\partial \tau} = \frac{1}{\tau^2} \sum_{i=1}^n t_i \left(Y_i - SNR * \exp\left(-\frac{t_i}{\tau}\right)\right) SNR * \exp\left(-\frac{t_i}{\tau}\right) \quad (10)$$

$$\frac{\partial^2 \ell}{\partial SNR^2} = -\sum_{i=1}^n \exp\left(-\frac{2t_i}{\tau}\right) \tag{11}$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \tau^2} &= -\frac{2}{\tau^3} \sum_{i=1}^n t_i \left(Y_i - SNR * \exp\left(-\frac{t_i}{\tau}\right) \right) SNR * \exp\left(\frac{t_i}{\tau}\right) \\ &\quad + \frac{2}{\tau^3} \sum_{i=1}^n t_i^2 SNR^2 * \exp\left(-\frac{2t_i}{\tau}\right) \end{aligned} \tag{12}$$

$$\frac{\partial^2 \ell}{\partial \tau \partial SNR} = -\frac{2SNR^3}{\tau^2} \sum_{i=1}^n t_i \exp\left(-\frac{t_i}{\tau}\right). \tag{13}$$

The expected value of minus the Hessian is then

$$\sum_{i=1}^n \begin{bmatrix} \exp\left(-\frac{2t_i}{\tau}\right) & -\frac{2SNR^3}{\tau^2} t_i \exp\left(-\frac{t_i}{\tau}\right) \\ -\frac{2SNR^3}{\tau^2} t_i \exp\left(-\frac{t_i}{\tau}\right) & \frac{2}{\tau^3} t_i^2 SNR^2 * \exp\left(-\frac{2t_i}{\tau}\right) \end{bmatrix} \tag{14}$$

which is going to be small when t_i/τ are mostly large. As a simple case, suppose we already knew τ . Then likelihood would only involve SNR and the Fisher information would be

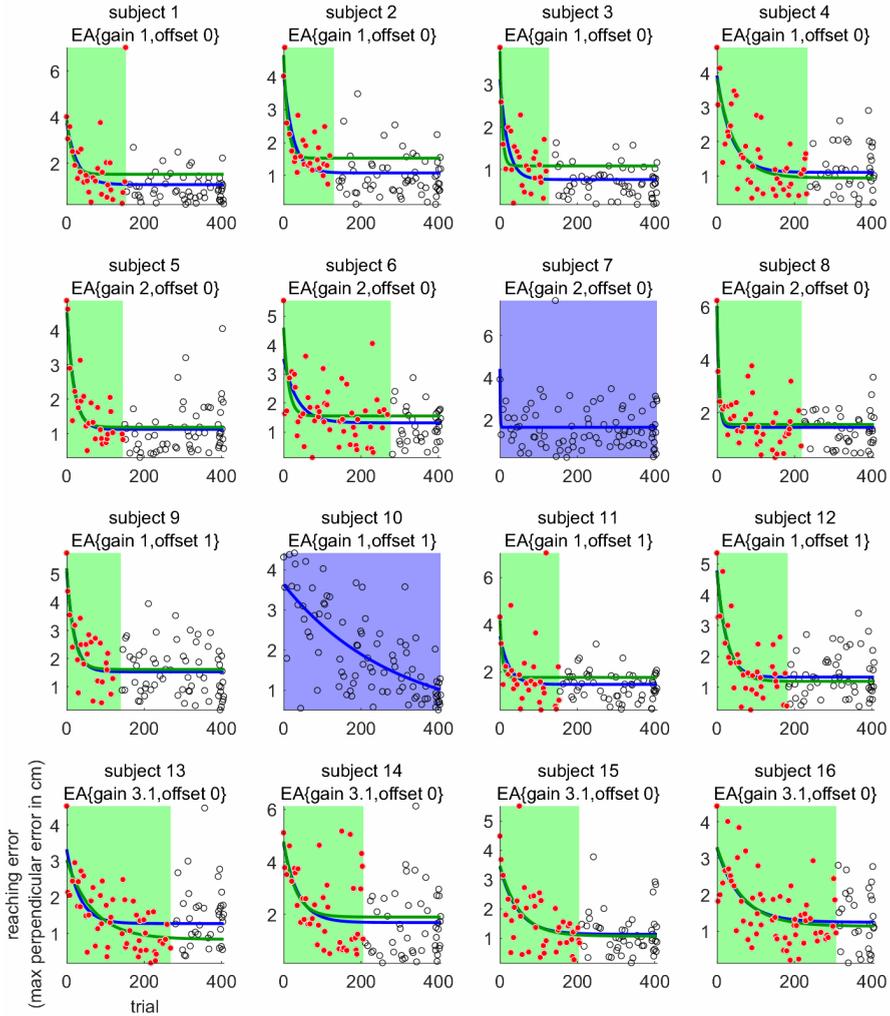
$$\sum_{i=1}^n \exp\left(-\frac{2t_i}{\tau}\right) \tag{15}$$

so the variance of \widehat{SNR} would then be about

$$\left(\sum_{i=1}^n \exp\left(-\frac{2t_i}{\tau}\right) \right)^{-1} \tag{16}$$

and of course, not knowing τ would only make things worse. Furthermore, knowing a minimum number of samples required (n) can inform the optimal design of experiments and provide best-case stopping criteria for a variety of applications when one needs to know how much data might be needed.

Figure A1 Learning curves from a past motor learning study (Patton et al., 2013) for human subjects who practiced reaching with a visuomotor rotation (see online version for colours)



Notes: Here, we evaluated this dataset to determine the minimal number of required samples (N_{opt}^* trials as per method summarised in Figure 2) to best estimate the learning curves. Circles represent movement error during training phase at intermittent catch trials. The blue curve represents exponential regression using data from all trials. The green curve represents exponential regression using data from the determined minimal number of trials (green shaded regions, red circles). The stopping criteria for determining N_{opt}^* trials for the subject 7 and subject 10 was not met within the total 406 available trials, indicating that more trials should have been collected for these individuals. Therefore, only the exponential regression using data from all trials are shown with blue curves and blue shaded regions for the subject 7 and subject 10. EA {gain, offset} represents the error-augmenting feedback condition that was used to test its effect on the rate and extent of learning (see Patton et al., 2013).