frontiers in Psychology

HYPOTHESIS AND THEORY
published: 20 May 2020
doi: 10.3389/fpsyg.2020.00911

Time Series Analysis in Forecasting Mental Addition and Summation Performance

*Anmar Abdul-Rahman**

Department of Ophthalmology, Counties Manukau DHB, Auckland, New Zealand

An ideal performance evaluation metric would be predictive, objective, easy to administer, estimate the variance in performance, and provide a confidence interval for the level of uncertainty. Time series forecasting may provide objective metrics for predictive performance in mental arithmetic. Addition and summation (addition combined with subtraction) using the Japanese Soroban computation system was undertaken over 60 days. The median calculation time in seconds for adding 10 sequential six digit numbers [CT_{Add}) was 63 s (interquartile range (IQR) = 12, range 48–127 s], while that for summation (CT_{Sum}) was 70 s (IQR = 14, range 53–108 s), and the difference between these times was statistically significant $p < 0.0001$. Using the mean absolute percentage error (MAPE) to measure forecast accuracy, the autoregressive integrated moving average (ARIMA) model predicted a further reduction in both CT_{Add} to a mean of 51.51 ± 13.21 s (AIC = 5403.13) with an error of 6.32%, and CT_{Sum} to a mean of 54.57 ± 15.37 s (AIC = 3852.61) with an error of 8.02% over an additional 100 forecasted trials. When the testing was repeated, the actual mean performance differed by 1.35 and 4.41 s for each of the tasks, respectively, from the ARIMA point forecast value. There was no difference between the ARIMA model and actual performance values (p-value CT_{Add} = 1.0, CT_{Sum}=0.054). This is in contrast to both Wright's model and linear regression (p-value < 0.0001). By accounting for both variability in performance over time and task difficulty, forecasting mental arithmetic performance may be possible using an ARIMA model, with an accuracy exceeding that of both Wright's model and univariate linear regression.

Keywords: ARIMA model, time series, mathematics, forecasting–methodology, cognitive performance

OPEN ACCESS

Edited by:
Jason C. Immekus,
University of Louisville, United States

Reviewed by:
Raydonal Ospina,
Federal University of Pernambuco, Brazil
Alen Hajnal,
University of Southern Mississippi, United States

***Correspondence:**
Anmar Abdul-Rahman
anmar_rahman@hotmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *16 January 2020*
Accepted: *14 April 2020*
Published: *20 May 2020*

Citation:
Abdul-Rahman A (2020) Time Series Analysis in Forecasting Mental Addition and Summation Performance. Front. Psychol. 11:911. doi: 10.3389/fpsyg.2020.00911

1. INTRODUCTION

Learning curves aim to model the gain in efficiency (increase in productivity, decrease in activity time, or both) of a repetitive task with increasing experience. The mathematical representation of the learning process is of particular interest across several disciplines including psychology (Mazur and Hastie, 1978; Balkenius and Morén, 1998; Glautier, 2013), medicine (Sutton et al., 1998; Ramsay et al., 2000; Dinçler et al., 2003; Hopper et al., 2007; Harrysson et al., 2014; Blehar et al., 2015), economics/industry (Cunningham, 1980; Lieberman, 1984; Badiru, 1991; Smunt and Watts, 2003) and more recently, artificial intelligence (Schmajuk and Zanutto, 1997; Perlich et al., 2003; Li et al., 2015).

Learning occurs most rapidly early in training, with equal increments in performance requiring a longer practice time in the later stages of the learning process. The classical understanding is that these diminishing returns result in learning curves that are smooth, decelerating functions (Mazur and Hastie, 1978; Jaber and Maurice, 2016). In 1880 Hermann Ebbinghaus first described the learning curve as a forgetting function; in a series of rigorous experiments he approximated the parameter as a negative exponential equation (Murre and Dros, 2015). In 1936 TP Wright investigated direct labor costs of assembling a particular aircraft and noted that the cost decreased with worker experience, a theory subsequently confirmed by other aircraft manufacturers (Wright, 1936). Analogous to Ebbinghaus's forgetting curve, he predicted the acquisition of skill followed a negative power function currently referred to as Wright's Model:

$$y_t = a \cdot x^b \tag{1}$$

Where y_t = the cumulative average time per unit, x = the cumulative number of units produced, a = the time to produce the first unit and b = learning coefficient (the slope of the function) ranging from −1 to 0; values close to −1 indicate a high learning rate and fast adaptation to task execution. Subsequently, JR Crawford described an incremental unit time model aimed at improving time representation in the algorithm, by substituting (x) in Wrights' model with the algebraic midpoint of the time required to produce a batch of units; this modification was a consequence of an observation that the time to complete a task decreased by a constant percentage, whenever the sum of the units doubled (Crawford, 1944; Jones, 2018). Three-parameter, two-parameter and the constant time exponential models were described to improve longterm predictions (Knecht, 1974; Mazur and Hastie, 1978). These algorithms were outperformed by multi-parameter hyperbolic models, where neutral, positive and negative learning episodes are represented through corresponding variable slope smooth curve profiles (Mazur and Hastie, 1978; Nembhard and Uzumeri, 2000; Shafer et al., 2001; Anzanello and Fogliatto, 2007). While the conventional univariate learning curves express a quantitative dependent variable in terms of an independent variable, multivariate models were eventually formulated to encode both qualitative and quantitative factors influencing the learning process (Badiru, 1992).

The smooth curves generated by these formal models provide an estimate at the average level of a set of observations. However, variation in performance demands a more rigorous representation of the learning process. This variation can be represented in a time series through two stochastic terms. an autoregressive term (AR), calculated as a weighted value from another point in the series, and a moving average (MA), which is estimated from the error terms in the series (Hyndman and Athanasopoulos, 2018). Characterization of time series using either an AR, MA, or combined ARMA processes was suggested independently in the 1920s by the Russian statistician and economist Eugen Slutsky (Slutzky, 1937), and the British statistician George Yule (Yule, 1921, 1926, 1927). It was not until the 1970s when Box and Jenkins described the autoregressive integrated moving average (ARIMA) model, which uses differencing of successive observations to render the series stationary, which is an essential property of the series for statistical validity (Milgate and Newman, 1990; Manuca and Savit, 1996). This study aims to investigate whether accounting for variance in the mental arithmetic using an ARIMA model is more accurate at forecasting performance, compared to Wright's model and univariate linear regression.

2. MATERIALS AND METHODS

2.1. Test Description

The learning period duration was 60 days, followed by 8 test days to assess the model forecasts. Tests were conducted between 7:00 and 7:30 a.m. Test sheets were randomly generated from the Soroban exam website (www.sorobanexam.org). Each sheet lays out both the questions and answers of a set of 10 columns of numbers (called a trial in this study). A sheet was composed of a mixture of six addition and four summation trials. The test difficulty was set to what is known in the Japanese Soroban exam system as difficulty level 3rd kyu, which consists of numbers ranging between 100,000 and 999,999. At the end of the test, a trial outcome was compared with the printed result and recorded. The last cell of the trial column was color coded with either a blue or red color, to indicate a successful or unsuccessful trial, respectively. Also, the time to complete a set of additions or summations was recorded in seconds. An example test sheet is provided in **Figure 1**. The in-built iOS voice over app (High Sierra 10.13.6) was used to vocalize the list of numbers from a .pdf list from the test sheet. Cumulative calculation time was defined as calculation time in seconds for adding 10 sequential 6-digit numbers, which either represented the addition task only (CT_{Add}) or a combination of addition and subtraction (CT_{Sum}). The author had limited prior experience with the Soroban (self-taught in 2017). Refer to the **Supplementary Material** section for the learning and test phase of the dataset.

2.2. The Soroban

The Soroban is a mechanical calculator, of which the origins are traced back to Mesopotamia, 2,500 years BC. Basic mathematical operations (addition, subtraction, multiplication, and division) can be performed using the device. There are two principles of operation: all calculations are performed as a number is pronounced, i.e., from the left to right. In addition, it reduces the complex mental mathematics to a simpler task, by using an algebraic principle of the method of complements, being in this case, either five or ten (Association, 1989; Schumer, 1999). Number representation and an example calculation is demonstrated in **Figure 1**. For clarification colored arrows are the next move in an operation (yellow = up, red = down). All computations are performed from left to right. Beads in contact with the central horizontal beam are considered in the final calculation. In this example of an addition operation (522+398), computation is started by representing the number 522 on the Soroban (**Figure 1C**i). The number (300) is then added to the hundreds rod (**Figure 1C**ii). Direct representation can take place with this step as there are an adequate number of beads not

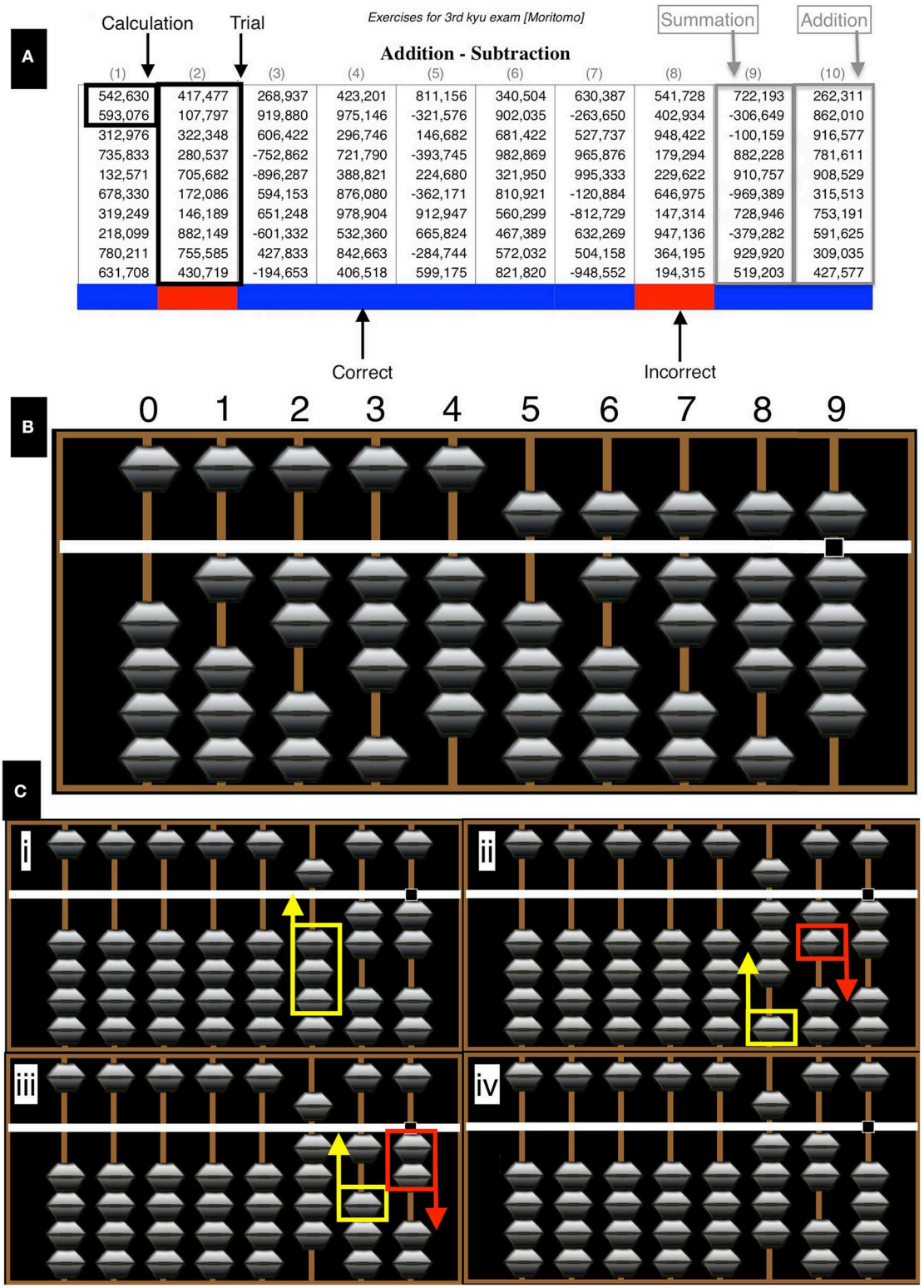

FIGURE 1 | Test example. **(A)** The test consisted of 10 columns (trials) of 6-digit numbers labeled 1–10. There were six addition and four summation trials per test sheet. To provide a visual indicator of performance in each sheet, color coding at the last cell of each column, where a blue or red color was used to indicate a correct or incorrect result, respectively. **(B)** Digits from 0 to 9 are represented on each rod by adding the numerical value of all beads contacting the central horizontal beam. The lower beads have a numerical value of 1, whereas the single upper bead has a value of 5. **(C)** An example of an addition operation (522+398) showing the principles of number representation and complementary number calculations (details provided in the text). Images generated with abacus software http://www.komodousa.com.

contacting the central beam. Adding 90 to the tens rod is not possible directly therefore, the complementary technique is used, in this method $100 - 10 = 90$, a bead is added to the hundreds rod and another subtracted from the tens rod (**Figure 1C**iii). To add 8 to the ones rod the complementary technique again, where $10 - 2 = 8$, a bead is added to the tens rod and 2 subtracted from the ones rod giving a result of 920 (**Figure 1C**iv).

2.2.1. Time Series Model Description

An ARIMA time series model is mainly defined by three terms (p,d,q), which represent the autoregressive (p), integrative (d), and the moving average (q) parameters of the model, respectively. The general mathematical description of the model is provided below (Box et al., 2015):

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \quad (2)$$

where

$$\phi(B) = 1 - \phi_1(B) - \phi_2(B)^2 .... - \phi_p(B)^p \quad (3)$$

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B)^2 .... - \theta_q(B)^q \quad (4)$$

1. (B) is the backward shift (lag) operator, which is defined by $B^k z_t = z_{t-k}$. This operator is convenient for describing the process of shifting between successive points in the series. That is to say B, operating on z_t, has the effect of shifting the data back one period.
2. $\phi(B)$ is the autoregressive polynomial operator in B of degree (p); it is assumed to be stationary, that is, the roots of $\phi(B) = 0$ lie outside the unit circle.
3. $\varphi(B) = \theta(B)\nabla^d$ is the generalized autoregressive (backward difference ∇z_t) operator; which is a non-stationary operator with d of the roots of $\varphi(B) = 0$ equal to unity, that is, d unit roots. The backwards difference operator is defined as $\nabla z_t = z_t - z_{(t-1)} = (1 - B)z_t$. Differencing is used to stabilize the series when the stationarity assumption is not met.
4. $\theta(B)$ is the moving average polynomial operator in B of degree (q); it is assumed to be invertible, that is, the roots of $\theta(B) = 0$ lie outside the unit circle.
5. The error term (a_t), which is assumed to have a Gaussian distribution, with a mean (μ) of zero and a constant variance of (σ_ϵ^2).

In practical terms, fitting the ARIMA model requires defining the model order (p,d,q). The autoregressive (ar) term, determines the value of (p), which is a datapoint in the series weighted by the value of proceeding data points. The term is given a number (ar_n); this represents the lag value in the series from where the correlation was calculated. The moving average (ma_n) corrects future forecasts based on errors made on recent forecasts; this term determines the (q) of the model order calculated from the partial autocorrelation function, which is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed. The integrated (d) portion of ARIMA models does not add predictors to the forecasting equation, rather, it indicates the order of differencing that has been applied to the time series to remove any trend in the data and render it stationary.

2.2.2. Statistical Analysis

Data was analyzed in R. The distributions of CT_{Add} and CT_{Sum} were modeled using the fitdistplus() package (Delignette-Muller and Dutang, 2015). Results are expressed as the median, range and interquartile range (IQR). Pearson's Chi-squared test (χ^2) with Yates' continuity correction was used to assess the differences in the accuracy of the calculated result between the addition and summation tasks. The Wilcoxon ranked sum test was used to assess the differences in CT_{Add} and CT_{Sum}.

Parameters of Wright's model were estimated using the learningCurve() package in R. This package uses Equation (1) in its calculations (Boehmke and Freels, 2017). The learning natural slope estimate (b) was calculated using the equation:

$$b < -\frac{Log_{10}T - Log_{10}t}{Log_{10}n - 1} \quad (5)$$

where T = total time (or cost) required to produce the first n units, t = time of all trials, n = total trials. The learning rate estimate (s) is calculated from the natural slope estimate by applying the following equation:

$$s = \frac{10^{b*log10(2)+2}}{100} \quad (6)$$

To forecast the 100th additional trial direct substitution in Equation (1) of (x) was done, where x = time for the 947th and 663rd attempt for the addition and summation tasks, respectively (a = the time for the first attempt in each of these tasks).

Univariate linear regression was utilized to assess the correlation between the time to perform the task and the number of trials and the equation of the best line fit was derived. The adjusted correlation coefficient (R^2) was used to represent the proportion of the variance explained by the model fits. The residual standard error (RSE) was used to assess model fit to the residuals.

The autoregressive integrated moving average model (ARIMA) was used for forecasting. The time series was plotted together with autocorrelation (acf) and partial autocorrelation functions (pacf). Although automated fitting of the time series (auto.arima) from the forecast package was initially used, acf and pacf graphs were used to confirm the order (p,q,d) of the series. After visual inspection of the time series plot suggested stationarity (mean, variance, and auto-covariance being independent from time), this assumption was confirmed by applying two statistical tests: the augmented Dickey-Fuller test (ADF), which is unit root test for stationarity, and The Kwiatkowski–Phillips–Schmidt–Shin test (KPSS). Unit roots (difference stationary process, i.e., a stochastic trend in a time series, sometimes called a "random walk with drift"), which exist in a time series if the value of α=1 in the general time series equation:

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon \quad (7)$$

The lag length (k) was chosen for this test (CT_{Add} k = 6, and CT_{Sum} k = 5) to avoid serial correlation of the residuals by choosing the last statistically significant lag, as determined by the partial autocorrelation function (pacf). The KPSS test was then applied, which is used for testing the null hypothesis that an observable time series is stationary around a deterministic trend (mean) or is non-stationary due to a unit root. Selection of the ARIMA model order (p,d,q) was chosen using the automated R auto.arima() command, which combines unit root tests, minimization of the corrected Akaike's Information Criterion (AICc) and Maximum likelihood estimation (MLE) to obtain an ARIMA model (Hyndman and Athanasopoulos, 2018). Validity of the model parameter choice was confirmed by plotting the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the stationary data to determine a possible model candidate as suggested by the minimal AICc. Model fitting diagnostics also considered the lowest root mean square error (RMSE) and mean absolute percentage error (MAPE). A plot of the ACF of the residuals was done to confirm if the residuals appeared to be white noise. Once these criteria were met the forecast equations were calculated. The characteristic roots of both time series equations were plotted to assess whether the model is close to invertibility or stationarity in relation to the complex unit circle. Any roots close to the unit circle may be numerically unstable, and the corresponding model will not be suitable for forecasting. This possibility is mitigated through the auto.arima() function, which avoids selecting a model with roots close to the unit circle (Hyndman and Khandakar, 2008). Plotting the fitted model against the time series plot was performed. The models were tested for autoregressive conditional heteroscedasticity using the McLeod-Li test. Plotting the acf of the residuals and the Ljung-Box test were performed to assess for autocorrelations. In order to assess the model performance, the mean point forecast was reported from each model. In addition, forecasted data was generated from the model parameters and compared with actual test performance for an additional 100 trials using a pairwise-Wilcoxon test with Bonferroni correction. A p-value of <0.05 was considered statistically significant for all tests.

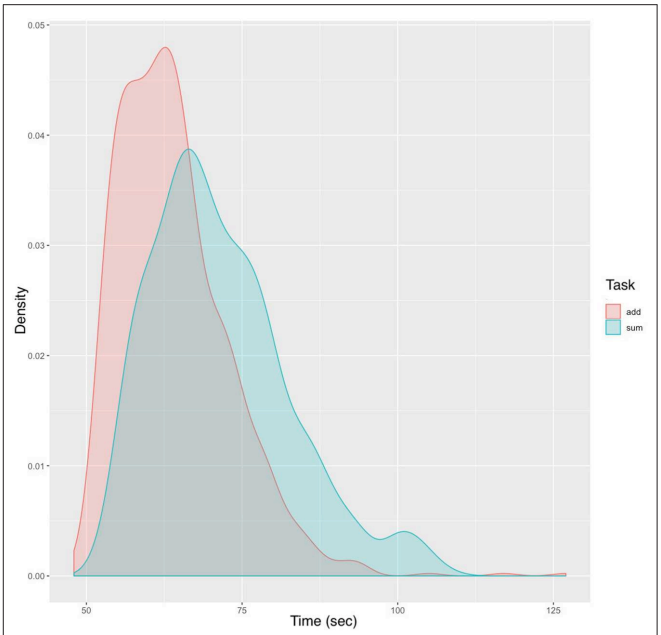

FIGURE 2 | Density plot of performance subset by task. The density plot uses kernel smoothing to reduce noise in the data, which generates a more defined distribution. The peaks of a density plot display represents values that are concentrated over the test interval. Noted in the plot are the narrower range of values and the shift to the left of the peak density for CT_{Add} compared to CT_{Sum}.

3. RESULTS

Over 60 days a total of 1,410 trials were conducted. The actual test time was 26.28 h during which a total of 847 addition and 563 summation trials were conducted. A variable number of trials, ranging from 0 to 70 trials per day were carried out. The distribution of both CT_{Add} and CT_{Sum} was non-normal and best fit a skewed exponential power type 4 distribution (model coefficients fit $p < 0.0001$). The skew and kurtosis of CT_{Add} were 1.38 and 7.46, respectively, whereas the corresponding values for CT_{Sum} were 0.80 and 3.40. The density distribution plot is demonstrated in **Figure 2**.

Addition tasks, being the simpler of the two, were more likely to yield an accurate result, and this difference compared to the outcome of the summation task was statistically significant ($\chi^2 = 9.33$, df = 1, $p < 0.002$). There was an increasing trend of total successful trials, as demonstrated in **Figure 3**. As expected, there was an improved performance with training, there were correct outcomes were recorded for 449/660 (68%) of trials in the first half compared to 598/750 (79.7%) in the second half of the learning period. The outcome of all tests classified by task type are summarized in **Table 1**.

Table 2 Summarizes the performance timing characteristics. From this table it can be noted that the median CT_{Add} was shorter compared to the median CT_{Sum}, this difference persisted throughout the learning period duration. Trials that concluded with an accurate calculated result (median time = 64 s, IQR = 13, range 48–117 s) were quicker compared to those where the result was incorrect (median time = 70 s, IQR = 15, range 51–127 s). These differences between addition, summation and performance times at the mid-learning interval were statistically significant ($p < 0.0001$). These results are shown in **Figure 3**.

3.1. Mathematical Models

3.1.1. Wright's Model

The time for the first event for CT_{Add} was 76 s, the total time 54,302 s and the number of trials 847, whereas for CT_{Sum} the first event was 74 s, the total time 40,297 s and the number of trials was 563. The learning rate was 0.98 and 0.99 for the addition and the summation tasks, respectively, this was calculated as a ratio of learning time at each doubling of the event i.e., time to event 1/time for event 2, time for event 2/time for event 4, time for event 4/time for event 8.....etc. The natural slope is calculated by dividing the log of the learning rate by log2, this was further refined by calculating the natural slope estimate when the total number of trials, total time of all trials and the time for the first trial is known. Substituting in Equation (5), the natural slope

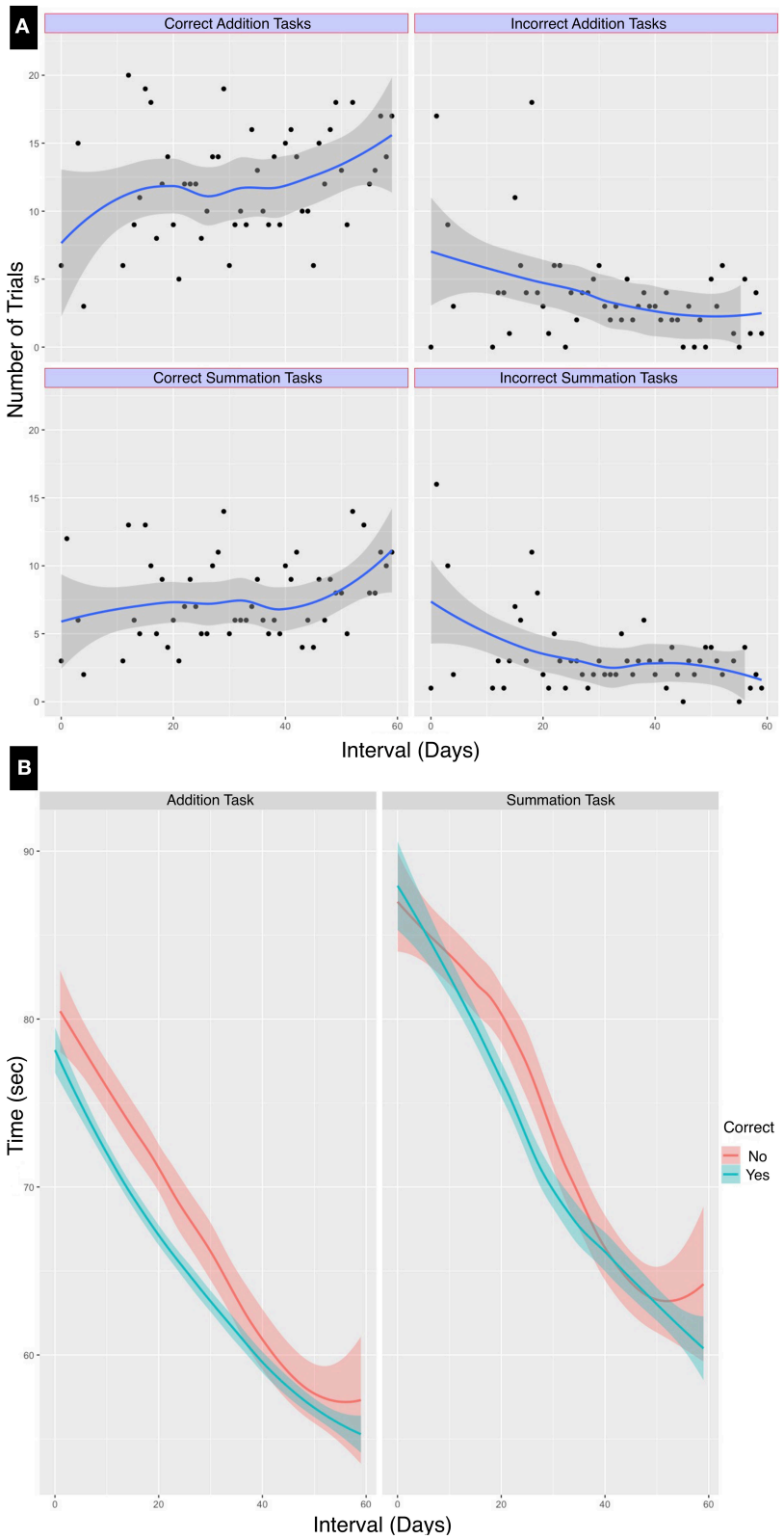

FIGURE 3 | Line graph of performance classified by task type and duration. Loess smoothed line graphs with 95% confidence intervals show the progress in test performance classified by the calculation result accuracy over the test interval of 60 days. There was both **(A)** an increase in the number of correct calculations and **(B)** a reduction of test time for addition and summation tasks over the learning period.

TABLE 1 | Learning outcome performance.

Trial	Correct	Incorrect	Total
LEARNING PERIOD (DAY 0–59)			
Addition	654 (r = 77%, c = 62%)	193 (r = 23%, c = 53%)	847 (60%)
Summation	393 (r = 70%, c = 38%)	170(r = 30%, c = 47%)	563 (40%)
Total	1047 (74%)	363 (26%)	1410
TESTING PERIOD (DAY 0–7)			
Addition	82 (r = 82%, c = 51.9%)	18 (r = 18.0%, c = 42.9%)	100 (50%)
Summation	76 (r = 76%, c = 48.1%)	24 (r = 24%, c = 57.1%)	100 (50%)
Total	158 (79%)	42 (21%)	

A comparison of the outcome of the trials during the learning and test periods. The learning period (60 days). The testing period (8 days) was to verify the accuracy of the model predictions, r = percentage by row, c = percentage by column.

TABLE 2 | The time to complete the tasks showed an expected reduction with learning.

			Range	
	Median	IQR	Min	Max
TOTAL LEARNING PERIOD				
CT_{Add}	63	12	48	127
CT_{Sum}	70	14	53	108
INITIAL LEARNING PERIOD (DAY 0–29)				
CT_{Add}	68	10	57	127
CT_{Sum}	78	14	59	108
LATE LEARNING PERIOD (DAY 30–59)				
CT_{Add}	58	7	48	88
CT_{Sum}	64	9	53	87
TEST PERIOD				
CT_{Add}	53	4	45	69
CT_{Sum}	58	6.25	47	75

The learning period was 60 days during which 1,410 trials were conducted, the performance modeled, and forecasting equations were used to generated expected timing data. The test period consisted of 100 additional trials.

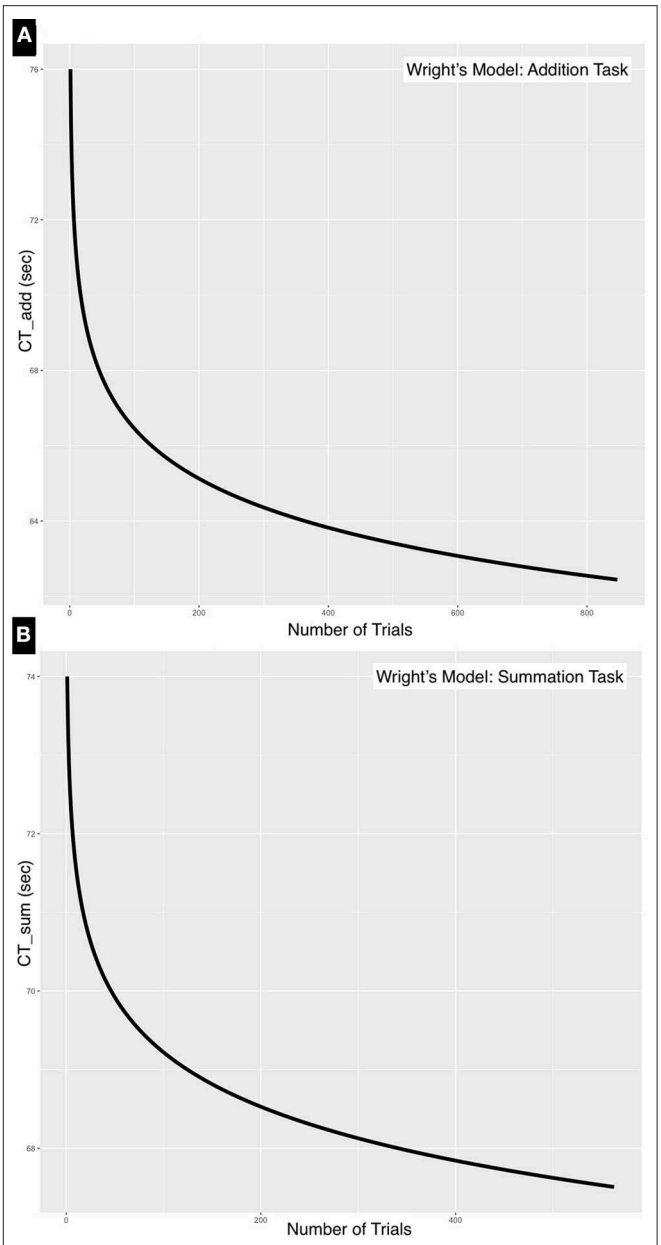

FIGURE 4 | Negative exponential curves for **(A)** addition and **(B)** summation tasks. These learning curves were generated using Wright's model. The natural slope estimate is steeper for CT_{Add} (−0.025) compared to that of CT_{Sum} (−0.005). This is a consequence of the simpler addition task reaching a plateau quicker than the more complex summation task.

estimate was −0.025 and −0.005 for the addition and summation tasks, respectively. The learning rate was further refined by taking into account the natural slope estimate by applying Equation (7), therefore the learning rate was estimated at 0.983 for the addition task and 0.996 for the summation task.

Substituting in Equation (1) where (x) is the forecasted performance time at the end of the 100th trial, CT_{Add} would be 62.24 s and CT_{Sum} would be 67.34 s. The plot of the model parameters is outlined in **Figure 4**.

3.1.2. Univariate Linear Regression

In the simplest case, when the variance in performance was disregarded, univariate linear regression was used to estimate the correlation between the cumulative time to perform both tasks and number of trials conducted. Consequently, the following equations were derived, where (x) is a variable representing the number of trials:

$$CT_{Add} = -0.027(\pm 0.00085) \cdot x + 75.63(\pm 0.41) \quad (8)$$

$$CT_{Sum} = -0.051(\pm 0.0021) \cdot x + 85.63(\pm 0.64) \quad (9)$$

The intercepts of Equations (8) and (9), +75.63(±0.41) and +85.63(±0.64), respectively, indicate the values of CT_{Add} and CT_{Sum} at baseline (day 0) when commencing the test, and therefore represent the level of prior expertise with the task.

Both the negative slope of the regression line (**Figure 5**) and the negative (x) variable coefficient demonstrate a reduction of performance time with learning. The summation task resulted in a higher RSE (7.55, df = 561) compared to the addition task (6.02, df = 845) due to the lower deviation from the regression line as shown by the median of the residuals for summation (−1.46) compared to (−0.6) for the addition task. The model's predictor (number of trials) explained about half of the variance in the dependent variable (CT_{Add} and CT_{Sum}) as indicated by an adjusted R^2 of 0.55 and 0.54 for Equations (8) and (9), respectively. Statistical significance was achieved for all model coefficients ($p < 0.0001$). Substituting in these equations, the forecast for the for the 100th additional forecasted trial yields a mean time of $CT_{Add} = 50.06$ s and $CT_{Sum} = 51.82$ s.

3.1.3. Autoregressive Integrated Moving Average Model (ARIMA)

The time series was of sub-daily frequency ranging from 10 to 70 trials (median 30) per day. There were three discontinuities in testing: interval 2, intervals 5–10, and interval 53. The interval was calculated from day 0 at the commencement of the test. As shown in **Figure 6** there was an overall declining average for both the addition and the summation trials. The ADF test confirmed stationarity of the series for both CT_{Add} (ADF value = −6.86, $p < 0.01$) and CT_{Sum} (ADF value = −6.86, $p < 0.01$). However the KPSS test was statistically significant for both CT_{Add} (KPSS value = 10.23, $p < 0.01$) and CT_{Sum} (KPSS value = 6.81, $p < 0.01$), this result indicated that the time series had stationary autoregressive terms (ar) and non-stationary moving average terms (ma), which was consistent with the declining trend in the performance time for both tasks as shown in **Figure 3**, this analysis confirmed that the series was weakly stationary and that differencing using the ARIMA model was required to render the series stationary for further analysis.

The correlograms in **Figure 6** show, for both CT_{Add} and CT_{Sum} that the acf is highly correlated at all lag values up to lag 30; therefore the suggested q would be of order 1. The pacf plot is used to select the order of the p term. For the addition task the highest significant value was at lag 6, whereas the value for the summation task was at lag 5. Therefore, a custom ARIMA model would be (6, 1, 1), AIC = 5409.09 for addition and (5, 1, 1), AIC = 3863.65 for the summation task.

Software packages (like R) provide the option of an automated ARIMA model order approximation, when this was trialed for the series, an 18 and a 23-order permutation was tested for addition and summation, from both these approaches the model order (2, 1, 3) with drift for addition, where the AIC was 5403.13 and the order for summation was (5, 1, 4) with drift, where the AIC was 3852.61. Hence the automated approximation provided more favorable model parameters. The coefficients and accuracy criterion of the model are listed in **Table 3**.

Autoregressive conditional heteroscedasticity (ARCH) among the lags was considered, the Mcleod-Li test for the addition model ARCH effects are were absent, for the summation task, from a total of 30 lags there was minimal (13%) ARCH effects in lags 3–7.

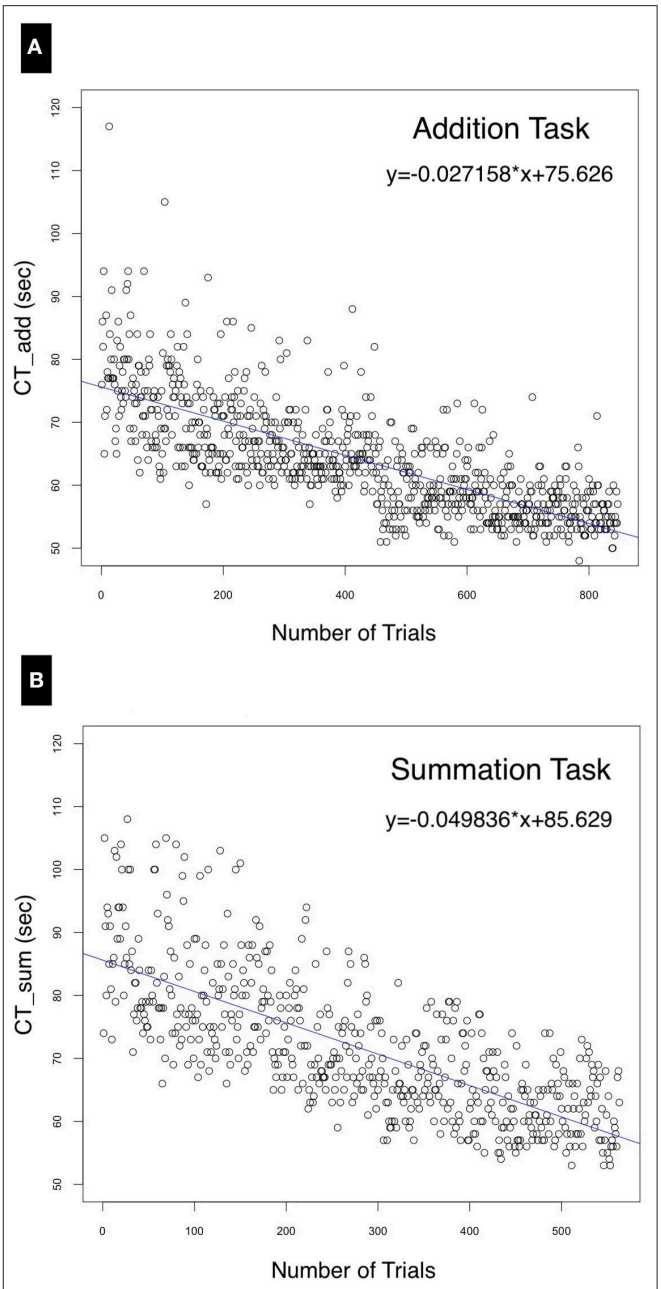

FIGURE 5 | Univariate linear regression model. correlating the performance time (seconds) with the number of trials for **(A)** the addition and **(B)** the summation trials. The distribution of residuals from the regression line for **(B)** is further from the regression line compared to **(A)** due to the wider variance in performance time for the summation task.

The following equations can be used to describe the time series fitted in **Figure 6** derived in standard notation:

$$CT_{Add}(Y_t) = -1.39Y_{(t-1)} - 0.86Y_{(t-2)} + 0.46e_{(t-1)} - 0.49e_{(t-2)} - 0.79e_{(t-3)} - 0.032e_t \quad (10)$$

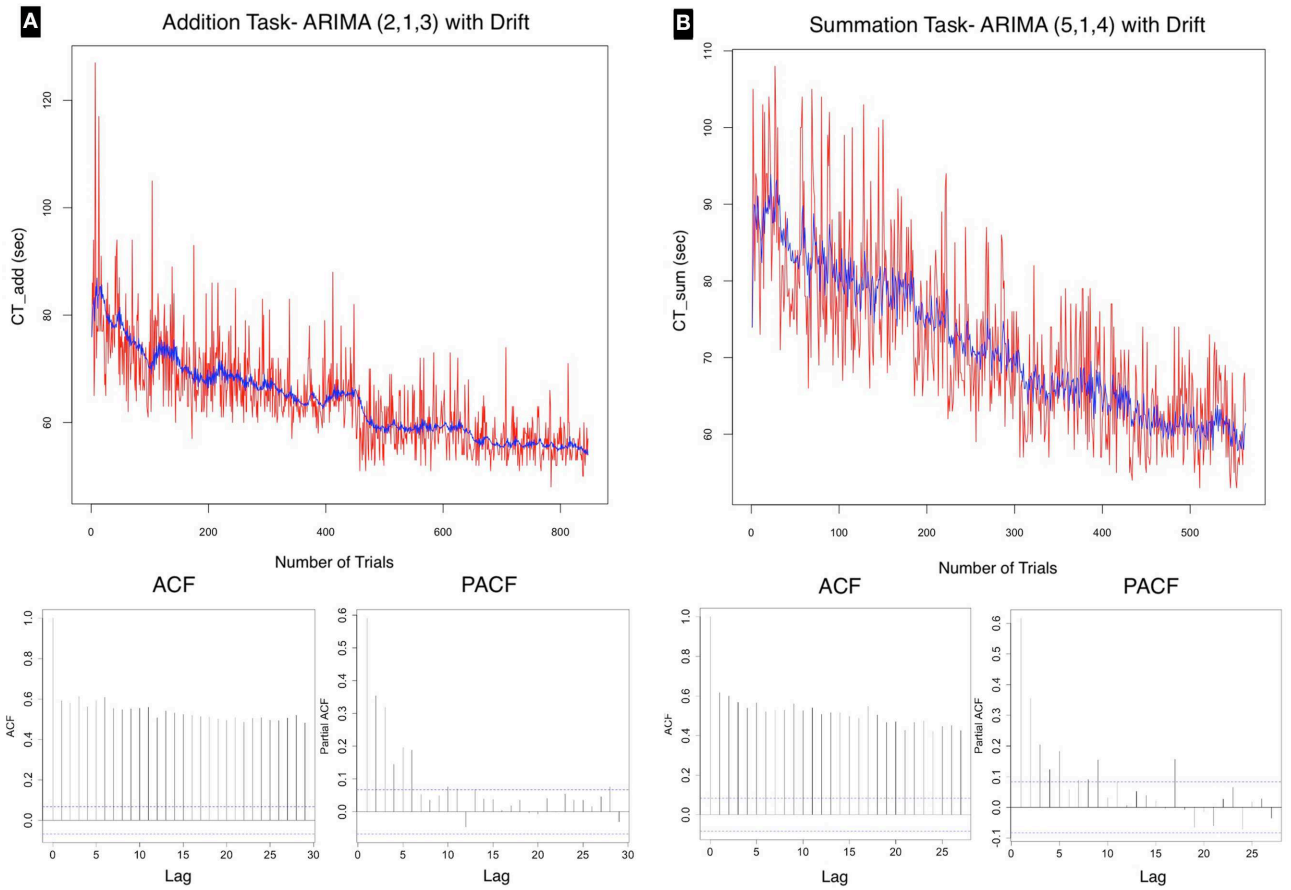

FIGURE 6 | Time series plot for **(A)** addition and **(B)** summation tasks with autocorrelation (acf) and partial autocorrelation (pacf) correlograms, which allow visual interpretation of the correlation of present data points with past points in the series. The time series for both tasks is demonstrating a declining trend. The acf plots for both tasks show a geometric pattern, with strong correlation of the sequential points up to lag 30. The pacf showed significant correlation at the 95% confidence interval up to lag 6 for addition, and lag 5 for the summation task. These tests confirm a gradual improvement in performance over the test time with a <20% correlation of test scores at each 6th trial in the series for addition and 5th trial for summation at the 95% confidence interval. A favorable model fit in the time series is demonstrated as a blue line in each of the series plots.

$$CT_{Sum}(Y_t) = -1.81Y_{(t-1)} - 1.53Y_{(t-2)} - 0.41Y_{(t-3)} + 0.31Y_t(t-4) + 0.092Y_{(t-5)} + 0.97e_{(t-1)} - 0.01e_{(t-2)} - 0.97e_{(t-3)} - 0.87e_{(t-4)} - 0.051e_t \quad (11)$$

Where (Y) is the autoregressive term, (e) is the moving average term, (e_t) is the error term and (t-n) is the lag (time interval between two data points).

As shown in **Figure 7**. The mean forecasted performance for the 100th trial for CT_{Add} was 51.50 ± 13.21 and for CT_{Sum} was 54.57 ± 15.37. From **Table 3** using (MAPE) to measure forecast accuracy, the model was able to forecast with an error of 6.42% for the addition task and 8.02% for the summation task.

Independence of the residuals for both ARIMA models was evaluated using the acf plot of the residuals, which showed the absence of autocorrelation. This was confirmed by the Ljung-Box test. Parameters for the addition task (χ^2 = 759.99, df = 800, p-value = 0.84) and the summation task (χ^2 = 480.93, df = 500, p-value = 0.72) failed to achieve statistical significance, therefore an absence of serial autocorrelation in both series, thereby confirming an appropriate model fit. The three model comparisons on predicting the actual means on repeating the tests for a further 100 trials for each of the addition and the summation tasks are listed in **Table 4**.

The forecasted mean ARIMA model values offered a closer match with actual test performance (p-value CT_{Add} = 1.0, CT_{Sum} = 0.054), this in contrast to both Wright's model and univariate linear regression, for which mean values differed from these test (p-value < 0.0001) for both tasks. Simulated data for the three models for the forecasted period were compared using the paired Wilcoxon rank sum test, which showed no difference for the ARIMA model values from the actual test values (p-value CT_{Add} 1.0, CT_{Sum}=0.054), this is in contrast to both Wright's and the linear regression models which forecasted statistically significant (<0.0001) values for both tasks.

TABLE 3 | ARIMA model parameters.

	ar1	ar2	ma1	ma2	ma3	Drift
ADDITION TRIALS ARIMA (2,1,3) MODEL PARAMETERS						
Coefficients	−1.3976	−0.864	0.4623	−0.4916	−0.7876	−0.0321
se	0.0922	0.0763	0.1077	0.0936	0.0766	0.0115
p-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.005
RMSE	5.84					
MAPE	6.32					
AIC = 5403.13, AICc = 5403.27, BIC = 5436.32						

	ar1	ar2	ar3	ar4	ar5	ma1	ma2	ma3	ma4	Drift
SUMMATION TRIALS ARIMA (5,1,4) MODEL PARAMETERS										
Coefficients	−1.8077	−1.5259	−0.4061	0.2994	0.0916	0.9703	−0.0115	−0.9706	−0.8742	−0.0495
se	0.0859	0.1229	0.1278	0.099	0.0483	0.0749	0.0426	0.0518	0.0591	0.0089
p-value	< 0.0001	< 0.0001	< 0.001	< 0.002	< 0.06	< 0.0001	< 0.8	< 0.0001	< 0.0001	< 0.0001
RMSE	7.28									
MAPE	8.02									
AIC = 3852.61, AICc = 3853.09, BIC = 3900.26										

ARIMA model parameters for the addition and the summation tasks. ar, autoregressive coefficients; ma, moving average coefficients; se, standard error. RMSE, root mean square error; MAPE, mean absolute percentage error; AIC, akaike information criterion; AICc, corrected akaike information criterion; BIC, Bayseian information criterion.

4. DISCUSSION

The ARIMA model provided a more accurate approximation to actual performance after 100 additional trials, compared to both univariate linear regression and Wright's model. Considering the model means, in their predictions, the former overestimated and the latter underestimated the actual performance (**Table 1**). Many formal models of learning generate smooth learning curves, which are seldom observed except at the level of average data (Glautier, 2013). In this example both Wright's model and simple linear regression hide important information regarding performance variance. The ARIMA model predicted CT_{Add} more accurately than that of CT_{Sum}, and this may be accounted for by the larger number of addition trials of which the test trials constituted 60% compared to the more complex summation task, where the test trials constituted approximately 40% of the total learning dataset. As shown in **Figure 2** the distribution of calculation times for both tasks were non-linear. In addition, the decline in both CT_{Add} and CT_{Sum} followed a non-linear trajectory over time (**Figure 3**). These patterns are consistent with the three phases of learning theory, which predict a three phase life cycle: the incipient phase during which a familiarization with the task occurs, which is characterized by a slow improvement; the learning phase, is where most of the improvement takes place; and the final phase, where the performance levels off (Carlson and Rowe, 1976). Whereas, prior knowledge of the task would have masked the incipient phase, the limitations of the univariate linear model become apparent by concealing the different performance phases altogether due to the constant slope of its regression line.

The neurophysiological basis of the Soroban remains unclear, however, it is known that computations using the Soroban involves a higher level of visual imagery (Tanaka et al., 2012). A longitudinal functional magnetic resonance imaging study of a patient with abacus-based acalculia suggested an important role in the parietal cortex and the dorsal premotor cortex in arithmetic ability of abacus users (Tanaka et al., 2012). Several cognitive processes required for mental arithmetic take place in these regions including retrieval, computation, reasoning, and decision making about arithmetic relations in addition to resolving interference between multiple competing solutions (interference resolution) (Menon, 2010). These factors may have played a role in the differences in variance in the performance of tasks as shown in **Figures 4**–**6**.

Calculation time for both addition and summation tasks as shown in **Figure 6** demonstrate a predictable downward trend and a slightly higher learning rate. The slope in the univariate linear regression was more negative for CT_{Sum} than CT_{Add}, although the former was a more complex task. There may be some influence of the difference in the scale of comparison, as the number of trials for the summation tasks were less than the addition tasks by about 20%. A comparison of the ARIMA model pacf plot in **Figure 6** also suggests a slightly higher learning rate with summation compared to the addition task as indicated by the loss of correlation over shorter lags with the former task. In Wright's model the steeper natural slope estimate was approximately twice that for addition compared to summation, perhaps reflecting the higher complexity of the latter and a more gradual departure from the learning phase. As multiple neural systems and pathways involved in mathematical information processing mainly the parietal cortex, prefrontal cortex with several functional dissociations between brain regions differentially involved in specific operations such as addition, subtraction, and multiplication have been suggested in literature. Menon (2010), it is therefore difficult to speculate on the underlying structural reason behind the detected difference

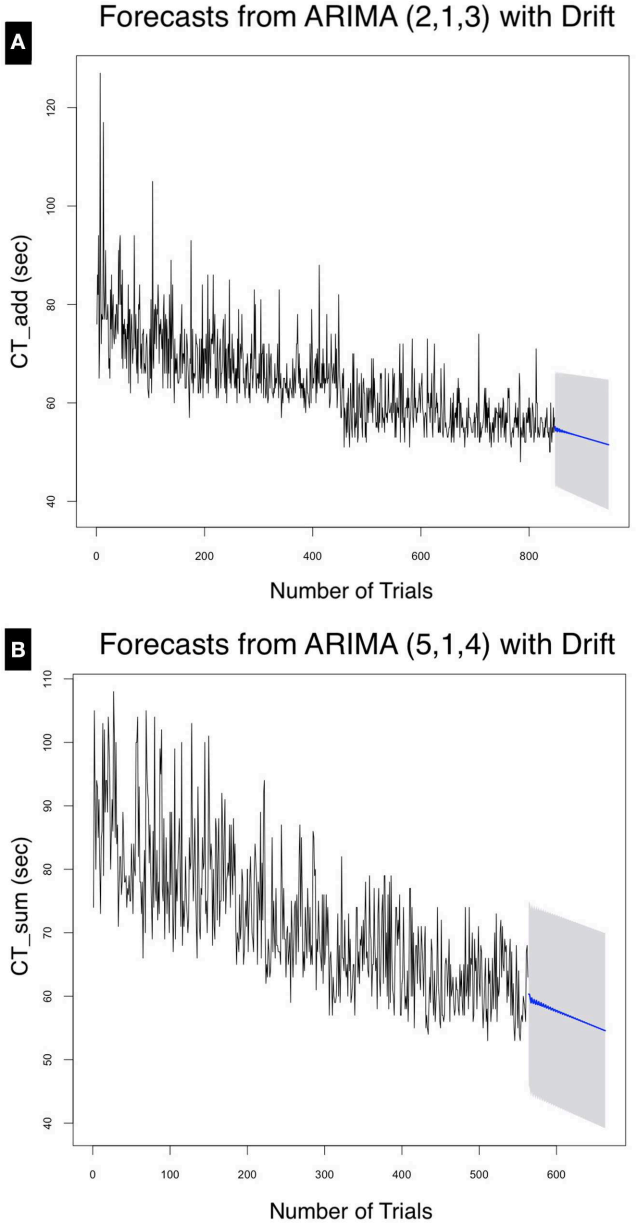

FIGURE 7 | Forecasts for **(A)** CT_{Add} and **(B)** CT_{Sum} at the 95% confidence interval for the next 100 trials. The point forecast for the 947th trial was calculated at 51.51 ± 13.21 s, whereas that for CT_{Sum} at the 663rd trial was calculated at 54.57 ± 15.36 s.

in learning speed. The other possibility behind the large variance seen in summation tasks is the measured difference in the number of complementary calculations per computation in each trial which was not considered in this experiment.

Inherent to the mathematical property of a times series analysis, is the capability of the model to capturing both linear and nonlinear relationships of the variables in the model. This property distinguishes it from other analysis methodologies which are either linear or non-linear (Yanovitzky and VanLear, 2008). In addition to describing the learning process in rigorous mathematical terms at an individual level, a psychological benefit is conferred to the test subject through accurate feedback of the improvement in performance. The limitations of this technique include the amount of data required to perform the analysis and the mathematical skill required to interpret the results. The protracted nature of the data collection requires a commitment in the testing process and may hinder some practicality as a routine test of learning performance. Although with the current experiment, the model fit was appropriate and delivered a high level of forecasting accuracy, most time series model predictions falter with extended forecast times due to non-stationarity, cohort effects, time-in-sample bias, and other challenges of longitudinal analyses (Taris, 2000; Yanovitzky and VanLear, 2008). Therefore, it is not clear from the current analysis how far into the future the forecast would be able to extend and retain its predictive accuracy. While the Soroban is still widely taught in Asian schools and therefore time series modeling may be beneficial for a more directed approach to teaching this skill, it may not apply to a wider population in other regions of the world where the use of the Soroban is less common. Future studies involving simultaneously recording an encephalogram may uncover wave activity associated with performance and the neurological basis of calculation errors in this task.

TABLE 4 | Mean point forecast.

	CT_{Add}	CT_{Sum}
POINT FORECAST COMPARISON (SEC)*		
Wright's model	62.24 (+9.38)	67.34 (+8.36)
Linear regression	50.06 (−2.74)	51.82 (−7.16)
ARIMA	51.50 (−1.36)	54.57 (−4.41)
Actual mean	52.86	58.98

*Comparison of the mean point forecasted value at the 100th trial for CT_{Add} and CT_{Sum}. *The number in brackets is the difference from the actual mean in seconds. The ARIMA model provided the closest prediction to actual performance.*

5. CONCLUSION

Time series analysis, by capturing the variance in performance may offer a more accurate mathematical representation of the learning process than classical learning theory models. The additional advantage of the ARIMA model to accurately forecast cognitive performance, with an accuracy exceeding that of both Wright's model and univariate linear regression, offers a potential for a wider applications for evaluation of cognitive function.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided

their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AA-R designed the study, analyzed the data, and created the current version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00911/full#supplementary-material

Data Sheet 1 | Learning period addition and summation times.

Data Sheet 2 | Testing period addition and summation times. For both **Supplementary Data Sheets**: rate, iOS accessibility program voice over speed; kyu, Calculation difficulty rate; Correct, correct answer calculated; time, test time in seconds; interval, Duration in days from start to the end of the learning period.

REFERENCES

Anzanello, M. J., and Fogliatto, F. S. (2007). Learning curve modelling of work assignment in mass customized assembly lines. *Int. J. Prod. Res.* 45, 2919–2938. doi: 10.1080/00207540600725010

Badiru, A. B. (1991). Manufacturing cost estimation: a multivariate learning curve approach. *J. Manuf. Syst.* 10, 431–441. doi: 10.1016/0278-6125(91)90001-I

Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Trans. Eng. Manage.* 39, 176–188. doi: 10.1109/17.141275

Balkenius, C., and Morén, J. (1998). "A computational model of emotional conditioning in the brain," in *Proceedings of Workshop on Grounding Emotions in Adaptive Systems* (Zurich).

Blehar, D. J., Barton, B., and Gaspari, R. J. (2015). Learning curves in emergency ultrasound education. *Acad. Emerg. Med.* 22, 574–582. doi: 10.1111/acem.12653

Boehmke, B., and Freels, J. (2017). learningcurve: an implementation of crawford's and wright's learning curve production functions. *J. Open Source Softw.* 2:202. doi: 10.21105/joss.00202

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control.* 5th Edn. Hoboken, NJ: John Wiley & Sons.

Carlson, J. G., and Rowe, A. J. (1976). How much does forgetting cost. *Ind. Eng.* 8, 40–47.

Crawford, J. R. (1944). *Learning Curve, Ship Curve, Ratios, Related Data.* Fort Worth, TX: Lockheed Aircraft Corporation.

Cunningham, J. A. (1980). Management: using the learning curve as a management tool: the learning curve can help in preparing cost reduction programs, pricing forecasts, and product development goals. *IEEE Spectr.* 17, 45–48. doi: 10.1109/MSPEC.1980.6330359

Delignette-Muller, M. L., and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *J. Stat. Softw.* 64, 1–34. doi: 10.18637/jss.v064.i04

Dinçler, S., Koller, M. T., Steurer, J., Bachmann, L. M., Christen, D., and Buchmann, P. (2003). Multidimensional analysis of learning curves in laparoscopic sigmoid resection. *Dis. Colon Rectum* 46, 1371–1378. doi: 10.1007/s10350-004-6752-5

Glautier, S. (2013). Revisiting the learning curve (once again). *Front. Psychol.* 4:982. doi: 10.3389/fpsyg.2013.00982

Harrysson, I. J., Cook, J., Sirimanna, P., Feldman, L. S., Darzi, A., and Aggarwal, R. (2014). Systematic review of learning curves for minimally invasive abdominal surgery: a review of the methodology of data collection, depiction of outcomes, and statistical analysis. *Ann. Surg.* 260, 37–45. doi: 10.1097/SLA.0000000000000596

Hopper, A., Jamison, M., and Lewis, W. (2007). Learning curves in surgical practice. *Postgrad. Med. J.* 83, 777–779. doi: 10.1136/pgmj.2007.057190

Hyndman, R., and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *J. Stat. Softw.* 27, 1–22. doi: 10.18637/jss.v027.i03

Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Melbourne, VIC: OTexts.

Japan Abacus Association (1989). *Soroban*: *The Japan Chamber of Commerce and Industry, 2nd Edn*. Tokyo: Japan Abacus Association.

Jones, A. R. (2018). *Learning, Unlearning and Re-learning Curves*. 1st Edn. London: Routledge; Taylor and Francis Group.

Jaber, M. Y., and Maurice B. (2016). "Chapter 14: The lot sizing problem and the learning curve: a review," in *Learning Curves: Theory, Models, and Applications*, ed M. Y. Jaber (London: CRC Press), 265–291. doi: 10.1201/b10957-17

Knecht, G. (1974). Costing, technological growth and generalized learning curves. *J. Oper. Res. Soc.* 25, 487–491. doi: 10.1057/jors.1974.82

Li, N., Matsuda, N., Cohen, W. W., and Koedinger, K. R. (2015). Integrating representation learning and skill learning in a human-like intelligent agent. *Artif. Intell.* 219, 67–91. doi: 10.1016/j.artint.2014.11.002

Lieberman, M. B. (1984). The learning curve and pricing in the chemical processing industries. *RAND J. Econ.* 15, 213–228. doi: 10.2307/2555676

Manuca, R., and Savit, R. (1996). Stationarity and nonstationarity in time series analysis. *Phys. D* 99, 134–161. doi: 10.1016/S0167-2789(96)00139-X

Mazur, J. E., and Hastie, R. (1978). Learning as accumulation: a reexamination of the learning curve. *Psychol. Bull.* 85:1256. doi: 10.1037/0033-2909.85.6.1256

Menon, V. (2010). Developmental cognitive neuroscience of arithmetic: implications for learning and education. *ZDM: Int. J. Math. Educ.* 42, 515–525. doi: 10.1007/s11858-010-0242-0

Milgate, M., and Newman, P. (1990). "Chapter 3: Arima models," in *Time Series and Statistics*, ed J. Eatwell (London: Palgrave Macmillan), 22–35.

Murre, J. M., and Dros, J. (2015). Replication and analysis of ebbinghaus forgetting curve. *PLoS ONE* 10:e0120644. doi: 10.1371/journal.pone.0120644

Nembhard, D. A., and Uzumeri, M. V. (2000). Experiential learning and forgetting for manual and cognitive tasks. *Int. J. Ind. Ergon.* 25, 315–326. doi: 10.1016/S0169-8141(99)00021-9

Perlich, C., Provost, F., and Simonoff, J. S. (2003). Tree induction vs. logistic regression: a learning-curve analysis. *J. Mach. Learn. Res.* 4, 211–255. doi: 10.1162/153244304322972694

Ramsay, C. R., Grant, A. M., Wallace, S. A., Garthwaite, P. H., Monk, A. F., and Russell, I. T. (2000). Assessment of the learning curve in health technologies: a systematic review. *Int. J. Technol. Assess. Health Care* 16, 1095–1108. doi: 10.1017/S0266462300103149

Schmajuk, N. A., and Zanutto, B. S. (1997). Escape, avoidance, and imitation: a neural network approach. *Adapt. Behav.* 6, 63–129. doi: 10.1177/105971239700600103

Schumer, G. (1999). Mathematics education in japan. *J. Curric. Stud.* 31, 399–427. doi: 10.1080/002202799183061

Shafer, S. M., Nembhard, D. A., and Uzumeri, M. V. (2001). The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Manage. Sci.* 47, 1639–1653. doi: 10.1287/mnsc.47.12.1639.10236

Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica* 5, 105–146. doi: 10.2307/1907241

Smunt, T. L., and Watts, C. A. (2003). Improving operations planning with learning curves: overcoming the pitfalls of messyshop floor data. *J. Oper. Manage.* 21, 93–107. doi: 10.1016/S0272-6963(02)00088-8

Sutton, D., Wayman, J., and Griffin, S. (1998). Learning curve for oesophageal cancer surgery. *Br. J. Surg.* 85, 1399–1402. doi: 10.1046/j.1365-2168.1998.00962.x

Tanaka, S., Seki, K., Hanakawa, T., Harada, M., Sugawara, S. K., Sadato, N., et al.(2012). Abacus in the brain: a longitudinal functional mri study of a skilled abacus user with a right hemispheric lesion. *Front. Psychol.* 3:315. doi: 10.3389/fpsyg.2012.00315

Taris, T. W. (2000). *A Primer in Longitudinal Data Analysis*. 1st Edn. London: SAGE Publications Ltd.

Wright, T. P. (1936). Factors affecting the cost of airplanes. *J. Aeronaut. Sci.* 3, 122–128. doi: 10.2514/8.155

Yanovitzky, I., and VanLear, A. (2008). "Time series analysis: traditional and contemporary approaches," in *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research,* eds A. Hayes, M. Slater, and L. Snyder (London: SAGE Publications Ltd) 89–124. doi: 10.4135/9781452272054.n4

Yule, G. U. (1921). On the time-correlation problem, with especial reference to the variate-difference correlation method. *J. R. Stat. Soc.* 84, 497–537. doi: 10.2307/2341101

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. *J. R. Stat. Soc.* 89, 1–63. doi: 10.2307/2341482

Yule, G. U. (1927). VII. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. *Philos. Trans. R. Soc. Lond. Ser. A* 226, 267–298. doi: 10.1098/rsta.1927.0007

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Abdul-Rahman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.