# Learning Curves for Marine Operations in the Offshore Renewable Energy Sector

Ben Hudson*†1, Edward Kay*†2, Mark Lawless†3 and Tom Bruce*4

*Industrial Doctoral Centre for Offshore Renewable Energy (IDCORE), University of Edinburgh
The Kings Building, West Mains Road, Edinburgh, EH9 3JG, United Kingdom

1 b.hudson@ed.ac.uk, ben.hudson@jbaconsulting.com

2 edward.kay@ed.ac.uk, edward.kay@jbaconsulting.com

4 tom.bruce@ed.ac.uk

†JBA Consulting Ltd.
South Barn, Broughton Hall, Skipton, North Yorkshire, BD23 3AE, United Kingdom

3 mark.lawless@jbaconsulting.com

***Abstract*—Installation, operations and maintenance costs represent one of the major barriers to the growth of wave and tidal-stream energy projects and a cost that can be significantly reduced in offshore wind. Time-domain Monte Carlo simulations used for developing and optimising offshore operation strategies can help minimise these costs. To achieve realistic and comprehensive estimates of project duration using these tools, it is essential to represent accurately the durations and variability of the input operation data. This paper attempts to quantify this stochastic nature of operation durations through the analysis of recorded data for an offshore wind farm installation project. Here we present evidence of learning for the majority of analysed operations and describe a stochastic learning curve model that is well-suited for implementation within time-domain Monte Carlo simulations. Additionally, the lognormal probability distribution is shown to be the most applicable for representing the random nature of operation durations for the recorded data-set. The proposed methods can be used during the operation phase of a project, where iterative analysis of recorded operational data is continually performed to improve predictions of project duration. Previously determined learning and distribution characteristics can also be used to obtain estimates for comparable future projects.**

***Keywords*—Learning curves; offshore operations; time-domain Monte Carlo simulations; installation, operations and maintenance (IO&M).**

## I. Introduction

The combined cost of installation, operations and maintenance (IO&M) activities are estimated to represent 35%, 46% and 30% of the lifetime cost of wave, tidal current and offshore wind energy arrays respectively [1], [2]. Previous work has proposed the use of advanced metocean planning tools that enable the time-domain Monte Carlo simulation of a marine project to help optimise offshore operation strategies and minimise these costs [3].

The three main input data requirements for these time-domain simulations are (*a*) an appropriate metocean time-series, (*b*) representative estimates of operation durations and (*c*) the metocean thresholds for each operation. Accurately incorporating the variability of operation durations is an essential requirement for producing realistic and exhaustive predictions of project duration. This paper summarises the analysis of recorded operation duration data that have been obtained for an offshore wind farm installation project. Specifically, the concept of a *learning curve* is investigated and shown to be an important phenomenon that should be incorporated in planning and optimisation studies for offshore renewable energy projects. The paper also addresses the implementation of these learning methods within the sophisticated metocean planning tool known as ForeCoast® Marine.

Section II introduces the basic learning curve theory, describing both the well-known deterministic model and the alternative stochastic representation. Additional stochastic implementations that have been derived are then discussed in Section III. A preliminary analysis of the data that demonstrates the basic LC model and confirms the presence of learning is then provided. Section V details the methodology used for the analysis, before Section VI summarises a selection of the most important results. Finally, the applications of the analysis method are discussed briefly in Section VII and the conclusions are given in Section VIII.

## II. Theory

### A. Deterministic Learning Curve Model

A learning curve (LC) is a mathematical description of workers' performance in repetitive tasks [4]. As repetitions take place, workers tend to demand less time to perform tasks due to familiarity with the operation and tools, and because shortcuts to task execution are found [4]. Crucially, the theory assumes that the task is identical for each repetition that takes place.

Two of the most common LC models currently in use are Wright's cumulative average model and Crawford's unit model [5]. As outlined in [4], both models have the common mathematical form

$$y = C_1 s^m \,, \tag{1}$$

where $C_1$ is the time to produce the first unit and $m$ is the slope of the LC on a log-scale ($-1 < m < 0$) referred to as the learning slope. In Wright's model, $y$ is the average time of all units produced up to the $s^{\text{th}}$ unit. In Crawford's model, $y$ is the processing time of the $s^{\text{th}}$ unit [5].

Figure 1 shows three representative examples of theoretical learning curves, calculated using Equation 1 for various values of $C_1$ and $m$ for the first 50 units of a hypothetical operation.

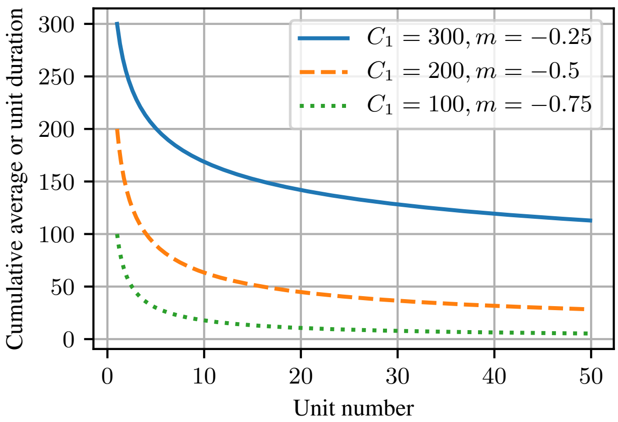


Fig. 1. Theoretical learning curves for three representative examples.

In log-linear form, Equation 1 can be written as

$$\ln y = \ln C_1 + m \ln s \,. \tag{2}$$

With some simple manipulation, the equations for the learning rate and the progress ratio can be derived. The learning rate, $\phi$, is the percentage of $y$ labour hours required to produce unit $2s$. In other words, as cumulative production doubles, the time required to produce a unit is reduced by a constant amount, known as the progress ratio, $L$.

$$\phi = 2^m \,, \tag{3}$$

$$m = \frac{\ln \phi}{\ln 2} \tag{4}$$

and

$$L = 1 - \phi \,. \tag{5}$$

For fitting the standard LC model to observed data, the method of non-linear curve fitting can be used to obtain the optimal model parameters [4].

### B. Stochastic Learning Curve Model

In direct contrast to the majority of LC models, in which learning is treated as a deterministic phenomenon, a method was proposed in [6] that addresses the stochastic nature of learning and treats it as a random process. It is argued that ignoring the randomness of the process will introduce significant errors to the model.

The stochastic model is defined by

$$y_s = y_1 s^m \,, \tag{6}$$

which is identical in form to Equation 1 but is stochastic in nature. Both $y_1$, the performance time of the first repetition, and $y_s$, the performance time of the $s^{\text{th}}$ unit, are stochastic parameters, meaning that they have an expected value, variance, coefficient of variation and probability density function (PDF). Performance time is equivalent to the previously mentioned unit duration, $y$, in Crawford's LC model. Although the learning slope $m$ is a random variable, it is assumed to be deterministic [6]. This does not eliminate the deviation of the parameter, but merely adds it to the deviation of $y_s$.

Using standard statistical rules, equations for the expected value ($\mathrm{E}[y]$), variance ($\mathrm{Var}[y]$), standard deviation ($\sigma_y$), coefficient of variation ($\mathrm{CV}_y$) and PDF ($f_y(y)$), as a function of the task repetition number $s$, are derived in [6]. These are given as;

$$\mathrm{E}[y_s] = \mathrm{E}[y_1] s^m \,, \tag{7}$$

$$\mathrm{Var}[y_s] = \mathrm{Var}[y_1](s^m)^2 \,, \tag{8}$$

$$\sigma_{y_s} = \sigma_{y_1} s^m \,, \tag{9}$$

$$\mathrm{CV}_{y_s} = \mathrm{CV}_{y_1} \tag{10}$$

and

$$f_{y_s}(y_s) = s^{-m} f_{y_1}(y_s s^{-m}) \,. \tag{11}$$

### C. Implementation for Normal Distribution

For a univariate parametric normal distribution, fully defined by its mean $\mu$ and variance $\sigma$, Equations 7 and 9 can be re-written in terms of these two parameters;

$$\mu_s = \mu_1 s^m \tag{12}$$

and

$$\sigma_s = \sigma_1 s^m \,. \tag{13}$$

Figure 2 shows the variation of the PDF, $f_{y_s}(y_s)$, as a function of the repetition number, $s$, for a normal distribution with $\mu_1 = 3.5$, $\sigma_1 = 1$ and $m = -0.5$, for the first four iterations of a task ($s = [1, 2, 3, 4]$).

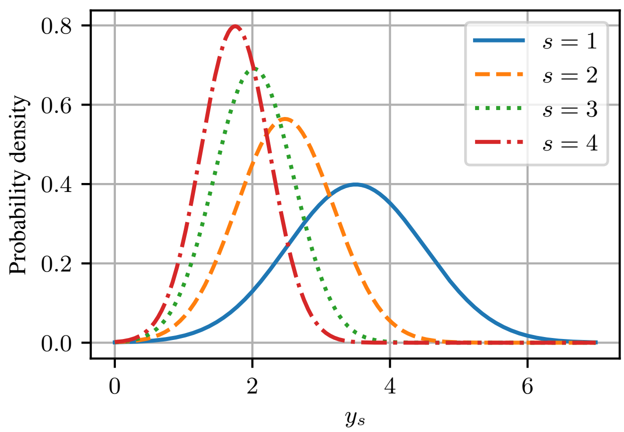


Fig. 2. Example PDFs as a function of repetition number assuming a normal distribution.

This figure shows that as the number of repetitions of the task increase, both the mean and variance of the distribution decrease. In other words, as more repetitions are completed, the average time to complete the task decreases and the probability density of that average time occurring increases — the PDF becomes "peakier" [6].

*D. Estimating Statistical Characteristics*

The stochastic LC method requires the statistical characteristics of the distribution to be estimated. Understandably, this is difficult for the typical sample of observed data because there is only one data point corresponding to each task repetition number. A method is outlined in [6] for obtaining these statistical characteristics.

The transformation equation

$$y_{n\to\kappa} = y_n \left(\frac{\kappa}{n}\right)^m , \quad n = 1, 2, \ldots, N \tag{14}$$

is used to transform the raw data to an arbitrary cycle number $\kappa$, where $y_n$ are the raw data points.

The transformation method is illustrated in Figure 3. The mathematical manipulations generate a distribution about the arbitrary $\kappa^{\text{th}}$ cycle. The distribution of the transformed data points are then used to calculate the required statistical characteristics.

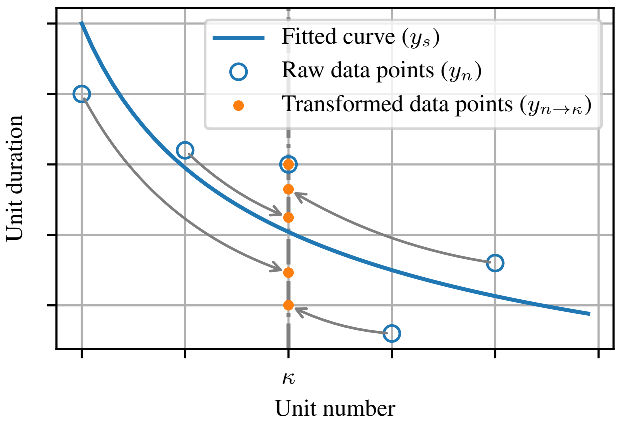


Fig. 3. Transformation method for estimating statistical characteristics.

## III. Additional Stochastic LC Implementations

The implementation of the stochastic LC model is particularly straightforward for the normal distribution because the parameters of this distribution are the expected value and the mean. For other distributions, the relationship between the distribution parameters and the statistical characteristics is not as trivial. In this case, Equations 7–11 can be used to obtain expressions for the distribution parameters as a function of the task repetition number.

The purpose of obtaining these expressions is to enable fast and straightforward random sampling for the operation duration input data in the time-domain simulation software. Although an equation for the PDF has already been derived, it can be difficult to randomly sample from the distribution using this equation. To do so requires the integration of the PDF to obtain an expression for the cumulative distribution function (CDF), as well as a method such as inverse transform sampling [7]. Integrating the PDF to obtain an analytical expression for the CDF can be a complex and, in certain cases, impossible task. Furthermore, the statistical sampling methods within the *SciPy* Python package [8], that have been implemented within the core functionality of ForeCoast® Marine, perform optimally when the distribution parameters are known.

For these reasons, expressions that relate the statistical characteristics to the distribution parameters were derived for five common parametric probability distributions;

1. triangular,
2. beta-PERT,
3. lognormal,
4. gamma and
5. Weibull.

There are several reasons for choosing these widely recognised distributions. Firstly, when defined appropriately, none of the above distributions can return a negative duration. This is one of the major limitations with the normal distribution — even if the probability of sampling a value less than 0 is negligible, the use of Monte-Carlo methods throughout the time-domain simulations, that perform several thousand iterations of an operation, increase the probability of returning a negative duration, particularly for low values of expected duration.

The triangular and beta-PERT distributions have the added advantage that their distribution parameters refer to quantifiable and comprehensible characteristics. Namely, both distributions have three parameters that refer to the minimum, maximum and most-likely (modal) values of the random variable in question. The intuitive nature of these parameters make these distributions well-suited for users without a statistical background or who only have qualitative estimates of operation durations. Finally, the lognormal, gamma and Weibull distributions are particularly suited for representing the data that is transformed as part of the stochastic LC model, as will be discussed in Section VI.

Future publications will detail the derivation of each of these additional expressions, as the mathematical methods are outside the scope of this paper. The derived expressions have, however, been fully validated and subsequently implemented within the ForeCoast® Marine software.

## IV. Preliminary Analysis

To allow convenient representation and grouping of tasks in the simulation software, an offshore project is typically divided into major *activities*, each comprised of several *operations*. For example, a wind farm installation project may be broadly categorised into loadout, transit and wind turbine generator (WTG) installation activities. As part of a preliminary investigation, the raw data was analysed on an *activity level* to determine the presence of learning.

Figure 4 shows the recorded durations for an activity corresponding to the installation of an individual WTG. The data have been normalised with respect to the time taken to complete the first unit. The figure shows both the unit duration and the cumulative average duration, along with the respective Crawford and Wright model fits. The Gauss-Newton Algorithm was used to perform the non-linear curve fitting [9]–[11].

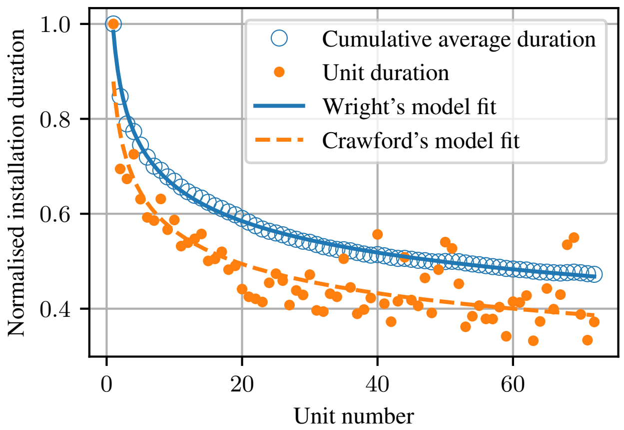


Fig. 4. Fitted learning curves for an activity in an offshore wind farm installation project

The figure clearly shows that learning is an observable phenomenon for the marine operations in the offshore wind farm construction project. There is some scatter in the unit duration data points around the best-fit curve, but the general trend is clear. The figure also serves as a visualisation of the shape of the LC model defined in Equation 1.

Grouping the individual operations that comprise an activity introduces an averaging effect and this is shown in Figure 4. The extremely good fit between the the cumulative average duration data and the best-fit Wright LC is explained by the fact that it is the cumulative average of data that has already been averaged. Still, the preliminary results demonstrate clear evidence of learning and the need for further investigation.

## V. Methodology

The operation data were provided in the format of daily progress reports (DPRs) detailing the exact times of each operation completed on a specific day for the entire installation campaign. The first step in the analysis was to collate this data and extract the *operational* components of the data. This required the removal of every instance of weather downtime and technical downtime. Technical downtime refers to any unexpected downtime that arises independently of the metocean conditions. This can include time lost due to breakdown of equipment, periods spent waiting for pilots and unplanned mechanical and electrical (M&E) works.

The analysis presented in the following sections focuses on the operations involved in the most important activity of the installation project — the WTG installation activity. There are 20 individual operations in the data-set. The methodology for the analysis of operation durations and the determination of learning is shown in Figure 5. This methodology was applied to each of the operations in the data-set.

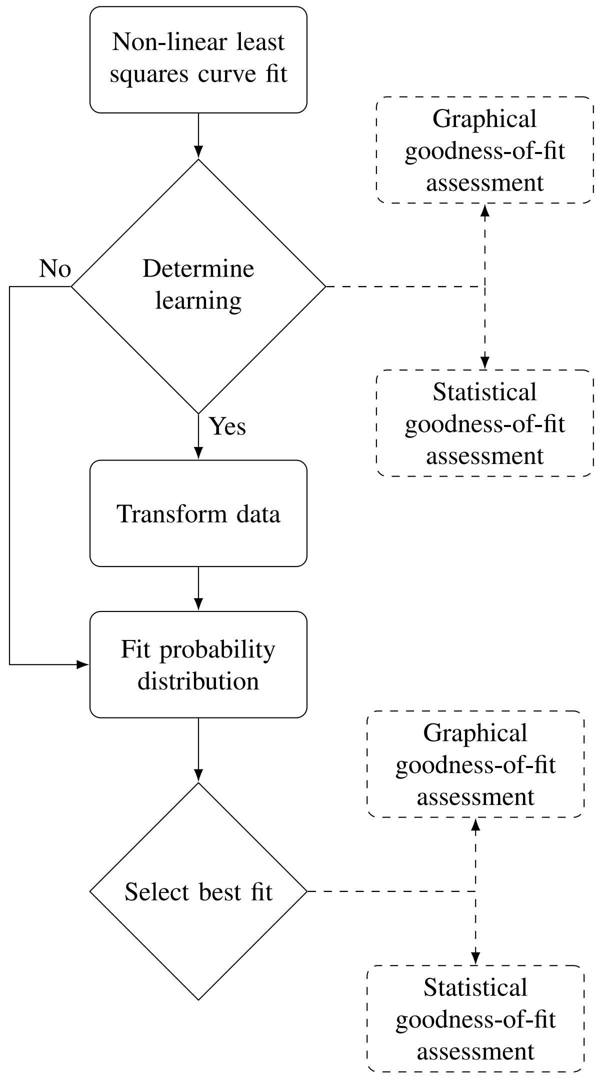


Fig. 5. Methodology for the analysis of operation durations and determination of learning.

### A. Non-linear Curve Fitting

As per the analysis on an activity level outlined in Section IV, the Gauss-Newton Algorithm was used to perform the non-linear regression curve fitting.

To evaluate the goodness-of-fit and determine the presence of learning, the methods recommended in [11] are used. The fit is assessed graphically by comparing the fitted curve to the raw data and ensuring the results are sensible. *Confidence bands* and *prediction bands* are calculated in addition to the best-fit curve and these are also shown on the plot. Confidence bands indicate where the expected LC parameters lie. Prediction bands take into account the uncertainty of the true position of the LC, as for the confidence bands, but also the random error around the curve. The methods for approximating the confidence and prediction bands outlined in [9], [10] are used in this analysis.

In addition to the graphical methods, the presence of learning is determined using common statistical tests and methods. The *t*-based 95% confidence intervals for the LC parameters are calculated as in [11]. Most importantly, the *t*-test statistics for evaluating the null hypothesis that the model parameters are equal to 0 are used to decide whether learning is present. The corresponding *p* values are calculated using a *t*-distribution as a reference distribution [11]. For the basic LC model, the learning slope parameter $m$ is of primary concern. If the hypothesis test fails to reject the null hypothesis that $m = 0$, this shows that at the assumed significance level, $\alpha = 0.05$, the test fails to show that there is learning present.

If the analysis confirms the presence of learning, Equation 14 is used to transform the data to an arbitrary cycle number. The $1^{st}$ cycle has been chosen for this analysis.

### B. Probability Distribution Fitting

The well-known method of maximum likelihood estimation (MLE) is used to fit a parametric probability distribution to sample data [12], [13]. In particular, the Nelder-Mead method, also known as the downhill simplex method, is used throughout the analysis [14].

A combination of graphical and statistical measures are used to determine the most suitable probability distribution. Graphically, four classical goodness-of-fit plots are used [13], [15];

1. a density plot comparing the histogram of the empirical distribution and the density function of the fitted distribution,
2. a cumulative distribution function (CDF) of both the empirical distribution and the fitted distribution,
3. a Q-Q plot of empirical quantiles against the theoretical quantiles and
4. a P-P plot of the empirical distribution function at each data point against the fitted distribution function.

The density plot and the CDF plot are considered the classical goodness-of-fit plots. The Q-Q and P-P plots are complementary and can be very important in some cases; the Q-Q plot emphasises a lack-of-fit at the distribution tails, while the P-P plot emphasises a lack-of-fit at the centre of the distribution [13].

Additionally, the following well-known goodness-of-fit statistics are used:

1. the Kolmogorov-Smirnov statistic [13], [16],
2. the Cramer-von Mises statistic [13], [16],
3. the Anderson-Darling statistic [13], [16],
4. Akaike's information criterion (AIC) [17] and
5. the Bayesian information criterion (BIC) [18].

The first three statistics measure the distance between the fitted parametric distribution and the empirical distribution. The AIC and the BIC are classical penalised criteria based on the log-likelihood that account for the complexity of the model. In this analysis, the distribution that results in the minimum value for the majority of these five goodness-of-fit measures is selected as the most appropriate.

Five theoretical distributions are tested; Weibull, gamma, lognormal, exponential and uniform. These distributions have been selected as they represent viable parametric distributions for the operation duration data that is being analysed. They are also included in the statistical package that has been used to perform the MLE method [13]. Also included in this statistical package are the normal and logistic distributions, but these are not considered as the support of these distributions extend to $-\infty$.

## VI. Results

This section presents the results for two specific operations. The first example, corresponding to the transit between WTG locations, resulted in the most significant evidence for *no learning*. The second example describes the installation durations for the first blade of each WTG and represents the operation with the most significant evidence for *learning*. After these two examples, a summary of the full data-set of 20 operations is given.

### A. Example 1 — Between Locations

The LC for the transit between locations operation is shown in Figure 6. The data have been normalised with respect to the mean of the raw data sample. The figure shows that even the best fit curve is quite flat, reflecting the lack of learning evident in the raw data. Furthermore, for early unit numbers, the lower curve of the 95% confidence band is sloping upwards and the 95% prediction band is predicting negative operation durations.

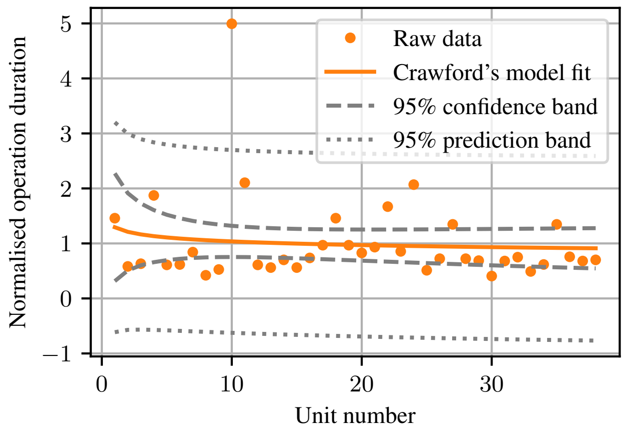


Fig. 6. Fitted learning curve for Example 1 — transit between locations.

The graphical evidence for a lack of learning is reflected in the statistical results shown in Table I. The upper value of the 95% confidence interval (*CI*) for the learning slope $m$ is positive (+0.181). As discussed in Section II, this parameter is restricted to $-1 < m < 0$ in the LC model. As such the upper 97.5% confidence interval for this parameter indicates that the operation durations increase with unit number. Most importantly, the $p$ value for the hypothesis test is 0.486,

TABLE I
SELECTED FITTING RESULTS FOR LEARNING SLOPE

| Operation | Between Locations | Installing Blade 1 |
|---|---|---|
| *t*-statistic | -0.703 | -12.892 |
| $p$ value | 0.486 | 3.68E-20 |
| 2.5% (lower *CI*) | -0.374 | -0.411 |
| 97.5% (upper *CI*) | +0.181 | -0.301 |

meaning that the null hypothesis of the learning slope being equal to zero cannot be rejected.

The fact that this transiting operation shows no evidence for learning can be seen as a validation of the analysis. For relatively short transits between WTG locations, it is plausible that not much learning is possible. This task is constricted by the physical limits of the vessel and by the required transiting distance. In fact, there is a legitimate argument that this particular operation should not have been included in the learning analysis because there are minor variations in the transit distances between each turbine. This would mean that the tasks being analysed are not identical, invalidating one of the major assumptions of the LC theory. Future work could investigate if normalising the transit duration with respect to the distance travelled affects the LC analysis.

As the analysis has shown that there is no learning for this operation, a probability distribution is fitted to the raw data. Table II shows the goodness-of-fit results. The lognormal distribution is selected as the most appropriate distribution as it results in the lowest values for each of the five statistical measures.

Figure 7 shows the comparison of the best fit lognormal distribution and the raw operation duration data. Again, the data have been normalised with respect to the mean of the raw data sample. Each of the four plots suggest that the lognormal distribution is an appropriate fit for this operation. Although the P-P plot indicates a slight deviation around the centre of the distribution, the fit is considered adequate.

### B. *Example 2 — Installing Blade 1*

The LC for the installation of the first blade is shown in Figure 8, again with the data normalised with respect to the mean of the raw data sample. Despite the scatter around the best fit LC, there is clear evidence of learning for this particular operation. Both the confidence and prediction bands are much closer than in Example 1. The first unit is over four times the sample mean and there is a rapid decrease in duration to this mean value — from approximately 10 units on, the normalised operation duration varies between 0.5 and 1.5.

As shown in Table I, the statistical fitting results for the learning slope corroborate the graphical assessment. The confidence interval is much more narrow, showing a 95% chance that $m$ is between -0.411 and -0.301. The $p$ value is essentially 0, meaning that the null hypothesis that the learning slope is equal to zero can be rejected.

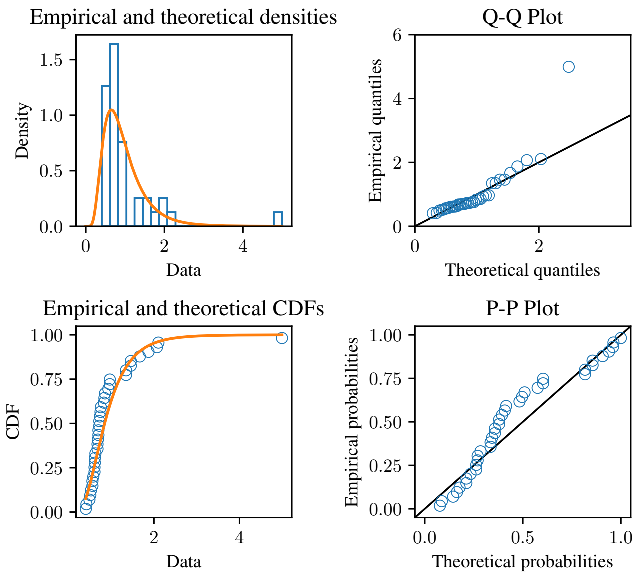


Fig. 7. Comparison of observed duration data and the fitted lognormal distribution for Example 1 — transit between locations.

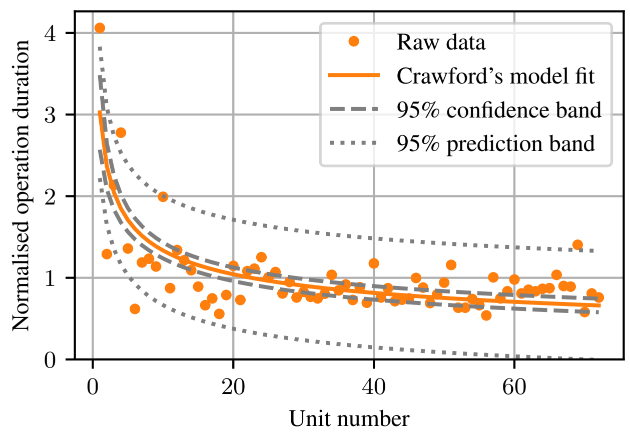


Fig. 8. Fitted learning curve for Example 2 — blade installation.

As there is significant evidence for learning, the raw data are transformed to the $1^{st}$ cycle using the fitted value for $m$ and Equation 14. A probability distribution is then fitted to this transformed data-set. The goodness-of-fit results for this operation are summarised in Table II. They show that the gamma distribution is the most suitable model for all five of the statistical measures.

The comparison of the fitted gamma distribution and the transformed data-set is shown in Figure 9. Again the data have been normalised to the mean of the raw sample. The figure shows that the chosen gamma distribution is an excellent fit for the transformed values of operation duration. There is only one outlier in the Q-Q plot which is especially encouraging considering the Q-Q plot accentuates deviations about the tails of the distribution [13].

TABLE II
GOODNESS-OF-FIT RESULTS FOR DISTRIBUTION FITTING

| | Weibull | Lognormal | Gamma | Exponential | Uniform |
|---|---|---|---|---|---|
| *Between locations* | | | | | |
| Kolmogorov-Smirnov statistic | 0.214 | 0.19 | 0.215 | 0.335 | 0.641 |
| Cramer-von Mises statistic | 0.52 | 0.289 | 0.444 | 1.006 | 6.879 |
| Anderson-Darling statistic | 2.918 | 1.507 | 2.322 | 5.163 | - |
| Akaike's Information Criterion | 95.42 | 75.449 | 86.536 | 104.959 | - |
| Bayesian Information Criterion | 98.695 | 78.724 | 89.812 | 106.597 | - |
| *Installing blade 1* | | | | | |
| Kolmogorov-Smirnov statistic | 0.08 | 0.076 | 0.064 | 0.423 | 0.328 |
| Cramer-von Mises statistic | 0.111 | 0.055 | 0.034 | 3.834 | 2.492 |
| Anderson-Darling statistic | 0.781 | 0.437 | 0.28 | 18.399 | - |
| Akaike's Information Criterion | 326.888 | 322.635 | 320.368 | 450.309 | - |
| Bayesian Information Criterion | 331.441 | 327.189 | 324.922 | 452.586 | - |

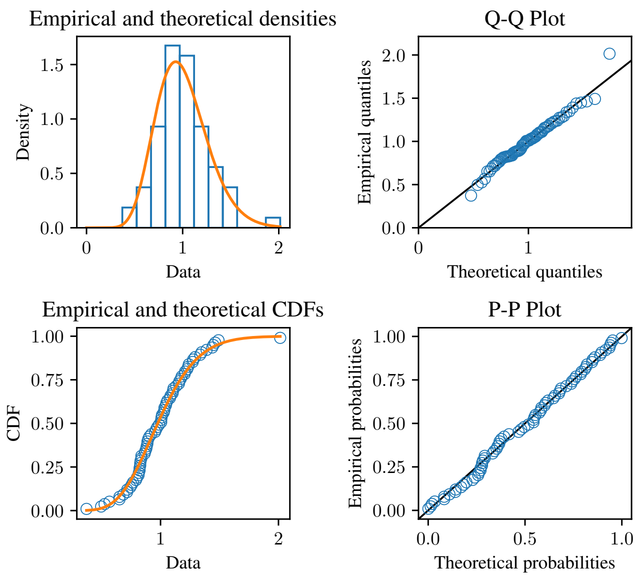


Fig. 9. Comparison of transformed duration data and the fitted gamma distribution for Example 2 — blade installation.

Using the fitted gamma distribution parameters and Equation 11, the resulting PDFs can be plotted as a function of the unit number. This is shown in two dimensions (2D) in Figure 10 and in three dimensions (3D) in Figure 11.

Figure 10 shows similar trends to that of the theoretical plot of Figure 2. As more units are completed, both the expected value and the variance decrease — the PDF becomes "peakier". Expectedly, the change in shape of the PDF is much more significant between units 1 and 35 than between units 35 and 70. For example; expressed as a percent of the mean raw data value, the mode for units 1, 35 and 70 are 0.84, 0.4 and 0.35 respectively.

Figure 11 shows the same data and trends as Figure 10 but in three dimensions and includes a PDF for each of the recorded unit numbers. This figure is particularly useful for visualising how the stochastic LC model is implemented within the simulation software. Imagining that each operation in the simulation model is associated with a similar plot and noting that the software continuously has access to the current operation and unit number, the operation duration sampling procedure can be viewed as a simple case of cycling through the appropriate 3D plot for the given operation, selecting the appropriate PDF based on the unit number and randomly sampling from that distribution.

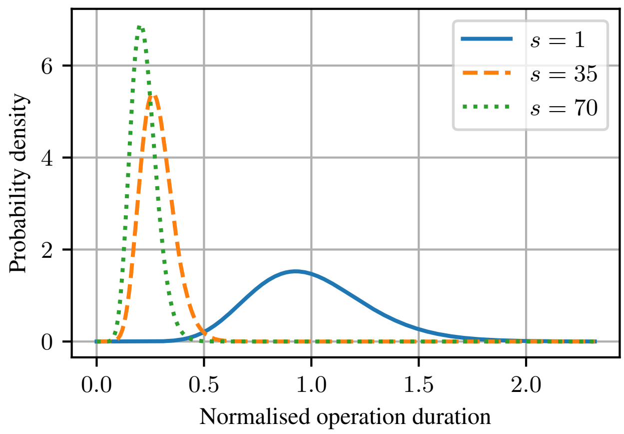


Fig. 10. PDFs plotted as a function of unit number for the fitted gamma distribution for Example 2 — 2D.

### C. Summary of Results for Full Data-set

This section summarises the results for the full data-set of operations that includes tasks such as the installation of

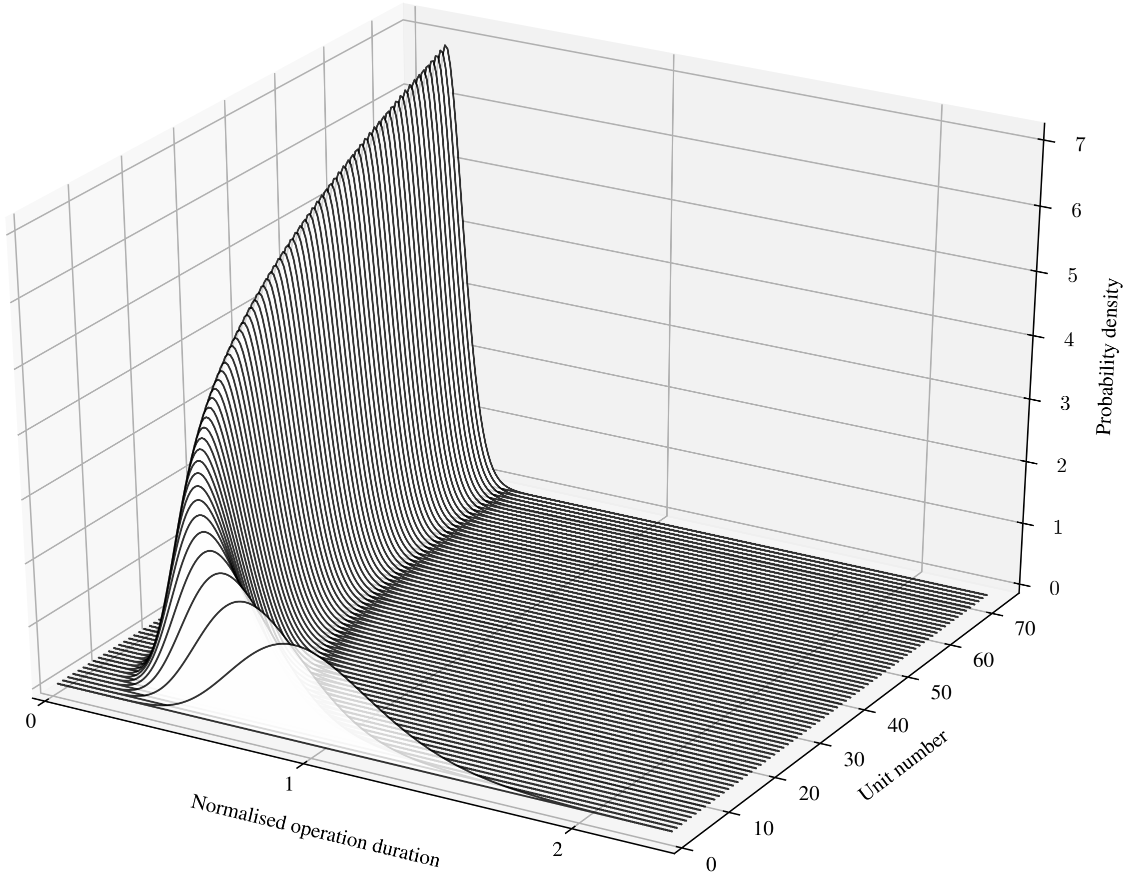


Fig. 11. PDFs plotted as a function of unit number for the fitted gamma distribution for Example 2 — 3D.

the tower and nacelle of the WTG, as well as the jack-up and jack-down operations at each WTG location. The analysis results are shown in Table III, which categorises the number of operations based on the presence of learning and on the selected probability distribution. The table shows that of the 20 operations analysed, 16 showed significant evidence for learning. This means the presence of learning was confirmed in 80% of the analysed operations, indicating that learning is an important factor that arises in the installation of offshore wind farms.

The results also show that the lognormal distribution is the most commonly selected probability distribution, regardless of the presence of learning. This indicates that the lognormal distribution is the most appropriate distribution for modelling both the raw data and the data transformed using Equation 14. The proportions of operations represented by the lognormal distribution are 56.25%, 75% and 60% for the learning, no learning and total categories respectively.

The Weibull and gamma distributions are the next most commonly selected distribution with 30% and 10% of the total operations respectively. The exponential and uniform distributions were never selected as the most suitable probability distribution.

TABLE III
NUMBER OF OPERATIONS CATEGORISED BY PRESENCE OF LEARNING AND FITTED PROBABILITY DISTRIBUTION

| | Learning | No Learning | Total |
|---|---|---|---|
| Lognormal | 9 | 3 | 12 |
| Weibull | 5 | 1 | 6 |
| Gamma | 2 | 0 | 2 |
| Total | 16 | 4 | 20 |

## VII. Implementation and Future Work

### A. Cost of Energy

The inclusion or exclusion of the learning phenomenon from time-domain simulation analysis of offshore operations will have a significant impact on project duration and the corresponding cost estimate. Future work will assess the impact of omitting learning on the projected duration and associated costs of marine operations.

### B. Software Integration

The stochastic LC model has been integrated within the technical methods of ForeCoast® Marine. With this sophisticated learning module, the software is capable of accurately and realistically representing the variation in operation duration that is so critical to the time-domain simulations. Furthermore, the stochastic LC method can be implemented regardless of the prior knowledge of the user.

In the case where no prior data exist or the user only has an intuitive knowledge of the durations and expected learning rates, the triangular or beta-PERT distribution can be selected and the minimum, most likely and maximum value of the operation duration specified. If learning is expected, the progress ratio $L$, as defined in Equation 5, can be specified. This intuitive parameter corresponds to the percentage reduction in duration that can be expected when the number of units completed doubles. Conversely, if a user with significant experience of a certain type of operation has prior knowledge or confident estimates of its statistical characteristics, then the appropriate distribution and learning slope can be chosen explicitly.

### C. Iterative Updating of Input Data

The second major application of the presented analysis method is the possibility of iteratively updating the simulation input data for ongoing projects. If progress updates are available during the operational phase of a project, similar to the DPRs mentioned previously, it is possible to continuously update the task duration input data based on the analysis of the incoming performance data.

The analysis then becomes a cyclical simulation process. As the project progresses, more operational data will be obtained and analysed, leading to more accurate estimates of the operation duration probability distributions and learning parameters. This will in turn lead to more accurate and reliable simulations and predictions of future progress.

The frequency of these iterative updates is entirely at the user's discretion. In one particular project, significant deviations between the simulations and actual performance were identified approximately halfway through the project. At this point, a thorough analysis of the operational data discovered that the original estimates of project duration slightly overestimated the time it was actually taking to complete specific tasks. Updating the input data for these tasks significantly improved progress predictions for the remainder of the campaign.

### D. Building an Operations Database

It is possible to estimate the learning curve for prospective marine operations with some certainty in the cases where comparable operations exist [19]. Similarly, prior knowledge of the expected value and variance of the duration of historic operations can be used to estimate the durations of similar tasks in the future. Of particular interest will be the comparison of operations across the offshore wind, tidal stream and wave energy sectors. At present it is impossible, due to the lack of data from comparable projects, but future work should focus on assessing the accuracy of using duration and learning statistics from one project for similar operations in a completely distinct project.

Regardless, it is worthwhile to develop a database of operational characteristics using available performance data, taken from either the literature or observed data. Correctly categorising this data will be of extreme importance. For example, an offshore WTG can be installed in myriad ways, so it is important to classify based on criteria such as the size and type of the turbines; the size and type of the installation vessel and the site conditions that can include water depth, wave climate and tidal regime. Through careful categorisation and continual expansion of this database of observed marine operation data, their representation in time-domain Monte Carlo simulations will dramatically improve.

## VIII. Conclusions

A stochastic learning curve method is described that is well-suited for representing the operation duration inputs to time-domain Monte Carlo simulations. The method allows a probability density function to be specified for each consecutive repetition of an operation, while accounting for the presence of learning at each iteration.

Using the stochastic model, expressions for the PDF as a function of the repetition number and the probability distribution parameters have been developed for five additional probability distributions; triangular, beta-PERT, lognormal, gamma and Weibull. These expressions have been successfully validated and implemented as part of a sophisticated learning module within the metocean planning tool known as ForeCoast® Marine. Future work will concentrate on publishing the derivations of these expressions.

Recorded operation duration data for the installation of an offshore wind farm have been analysed and it has been shown that a learning curve is present in 80% of the analysed operations. Furthermore, the lognormal, Weibull and gamma distributions have been identified as the distributions that most suitably model the operation duration data. The lognormal distribution is the most commonly selected distribution, irrespective of whether learning is present, and resulted in the best fit distribution for 60% of the total operations.

Two further applications of the proposed method are discussed. Firstly, leaning curves and probability distributions can be analysed during the operation phase of a project. Estimates for the expected value, variance and learning slope of the operation data can be iteratively updated as the campaign progresses, continually improving project progress predictions.

Secondly, it is possible to use previously determined learning curves and distribution characteristics to obtain estimates for comparable operations in the future. Continuing to build a database of operations characteristics is recommended and expected to improve the accuracy of time-domain simulations. Future work will assess the viability of using the statistical characteristics of a historic operation for a similar operation in a separate project.

## Acknowledgment

The authors would like to thank the Energy Technology Institute and the Research Council UK Energy Programme for funding this research as part of the IDCORE programme (EP/J500847/1). Sincere thanks are also given to the sponsoring company JBA Consulting and its staff, particularly Laurie Wilkinson and Anthony Gray, as well as additional IDCORE supervisors Helen Smith and Rafet Emek Kurt.

## References

[1] S. I. Ocean, "Ocean energy: cost of energy and cost reduction opportunities," *Strategic Initiative for Ocean Energy (SI OCEAN) May*, 2013.

[2] B. Maples, G. Saur, M. Hand, R. van de Pieterman, and T. Obdam, "Installation, operation, and maintenance strategies to reduce the cost of offshore wind energy," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2013.

[3] B. Hudson, E. Kay, M. Lawless, and T. Bruce, "Advanced metocean planning tools for the wave and tidal energy sectors," in *Twelfth European Wave and Tidal Energy Conference*, Cork, Ireland, 2017.

[4] M. J. Anzanello and F. S. Fogliatto, "Learning curve models and applications: Literature review and research directions," *International Journal of Industrial Ergonomics*, vol. 41, no. 5, pp. 573–583, 2011.

[5] J. Tilindis and V. Kleiza, "Learning curve parameter estimation beyond traditional statistics," *Applied Mathematical Modelling*, vol. 45, pp. 768–783, 2017.

[6] S. Globerson and D. Gold, "Statistical attributes of the power learning curve model," *International journal of production research*, vol. 35, no. 3, pp. 699–711, 1997.

[7] L. Devroye, *Non-Uniform Random Variate Generation*. New York, Berlin, Heidelberg, Tokyo: Springer-Verlag, 1986.

[8] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–. [Online]. Available: http://www.scipy.org/

[9] A. Ruckstuhl, "Introduction to nonlinear regression," 2010. [Online]. Available: https://pdfs.semanticscholar.org/8fa1/3fead47cc6ecf3d27de9e682dcef36c77502.pdf

[10] D. M. Bates and D. G. Watts, "Nonlinear regression analysis and its applications," *Wiley Series in Probability and Mathematical Statistics.*, 1988.

[11] F. Baty, C. Ritz, S. Charles, M. Brutsche, J.-P. Flandrois, M.-L. Delignette-Muller *et al.*, "A toolbox for nonlinear regression in r: the package nlstools," *Journal of Statistical Software*, vol. 66, no. 5, pp. 1–21, 2015.

[12] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011, vol. 100.

[13] M. L. Delignette-Muller, C. Dutang *et al.*, "fitdistrplus: An r package for fitting distributions," *Journal of Statistical Software*, vol. 64, no. 4, pp. 1–34, 2015. [Online]. Available: http://www.jstatsoft.org/v64/i04/

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965. [Online]. Available: http://dx.doi.org/10.1093/comjnl/7.4.308

[15] A. C. Cullen and H. C. Frey, *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*. Springer Science & Business Media, 1999.

[16] R. B. D'Agostino and M. A. Stephens, *Goodness-of-fit techniques*, ser. Statistics: Textbooks and Monographs. New York: Marcel Dekker, Inc., 1986, vol. 68.

[17] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec 1974.

[18] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 03 1978. [Online]. Available: https://doi.org/10.1214/aos/1176344136

[19] C. Jablonowski, "Incorporating uncertainty into learning curves: A case study in oil drilling estimates," in *American Society for Engineering Education*. American Society for Engineering Education, 2010.