Deep Learning-Enabled Statistical Model Estimation for Power Transformers with Censoring and Truncation Problems

Jiaxiang Cheng[1], Sungin Cho[2], Yap-Peng Tan[3], and Guoqiang Hu[4]

[1,3,4] *Nanyang Technological University, 639798, Singapore*
jiaxiang002@e.ntu.edu.sg
eyptan@ntu.edu.sg
gqhu@ntu.edu.sg

[2] *SP Group, 349277, Singapore*
chosungin@spgroup.com.sg

Abstract

Traditional statistical models, e.g., Weibull distributions, are popular solutions for failure modeling and degradation analysis in a variety of industries. To estimate the parameters of these statistical models, maximum likelihood estimation (MLE) is often engaged through various optimization algorithms. However, when dealing with highly reliable or new equipment, it is challenging to fit limited or unbalanced data to obtain an accurate model. In this paper, we propose a deep learning (DL)-based model for estimating the Weibull parameters with both censoring and truncation problems. Instead of using the conventional matrices such as concordance index, we propose a novel validation framework to examine the prediction accuracy of different models. We examine the performance of the proposed approach on real-world power transformer data, and the results show that our approach can improve prediction accuracy and is less susceptible to the truncation problem. Our results also suggest that deep learning techniques can help enhance traditional statistical modeling for reliability analysis.

1. Introduction

Power transformers are critical to the reliable and stable operation of power grids, delivering electricity to homes, businesses, industries, and critical utilities. However, power transformers are subject to various types of failures, which can lead to power outages, equipment damage, and even safety hazards. In order to reduce the risk of failures and improve transformer maintenance strategies, it is essential to understand the probability distribution of the time to failure.

Survival analysis is a statistical method that studies the probability of event occurrence over time. The two-parameter Weibull distribution is a popular statistical model used in many engineering applications, including power industries (Hong, Meeker, & McCalley, 2009; Chmura, Morshuis, Gulski, Smit, & Janssen, 2011), to describe the distribution of failure probability. However, traditional methods may not be able to capture complex relationships between additional covariates and the probability of failures, while recent advances in deep learning (DL) offer new solutions for modeling with high-dimensional inputs and predicting survival outcomes.

While DL models have shown promising results in various survival analysis applications (Katzman et al., 2018), the existing DL-based survival models that assume the Weibull models as underlying distributions have only considered the censoring problem (Bennis, Mouysset, & Serrurier, 2020, 2021; Nagpal, Li, & Dubrawski, 2021), which refers to incomplete observations of the time to failure. In the meanwhile, the truncation problem, which exists when there are samples with events happening before the start of the studies, e.g., transformers failed before the actual observation of failures, has not been fully addressed by DL-based methods. However, both problems have been examined and addressed by formulating the likelihood function and estimating two-parameter Weibull models with maximum likelihood estimation (MLE) (Hong et al., 2009).

In this paper, we propose a DL-based model to consider both censoring and truncation problems and estimate the two-parameter Weibull distribution for power transformers. Our model aims to capture the complex relationships between the predictors and the response variable to improve the accuracy of failure prediction. We evaluate the performance of our model on a real-world dataset of power transformers and compare it with traditional estimation methods. Our results demonstrate the effectiveness of our proposed model in predicting survival probability and the number of failures.

The remainder of the paper is organized as follows. Section 2 introduced the proposed DL-based model and the validation framework to evaluate the prediction performance. The experimental results are presented in Section 3, while the conclusion is provided in Section 4.

2. METHODOLOGY

The two-parameter Weibull distribution, which is specified by a shape parameter β and a scale parameter η, is a commonly used statistical model in failure analysis. Its probability density function ($p.d.f.$) $f(t)$ and cumulative distribution function ($c.d.f.$) $F(t)$ are defined as (Weibull, 1951),

$$f(t) = \frac{\beta}{\eta}(\frac{t}{\eta})^{\beta-1}e^{-(\frac{t}{\eta})^{\beta}}, \quad (1)$$

$$F(t) = 1 - e^{-(\frac{t}{\eta})^{\beta}}, \quad (2)$$

respectively, where t is the survival time. The Weibull distribution is favored for its ability to model both increasing and decreasing hazard rates over time by simply varying β, where the hazard rate is the probability of an event occurring in a small time interval, given that the individual has survived up to that point, and it is defined as $f(t)/(1-F(t))$.

In recent years, DL models have gained popularity in survival analysis due to their ability to capture complex relationships between covariates and the survival outcome that may be missed by traditional statistical models. In this paper, we propose a DL-based model named DeepWeibull, which models the reliability of power transformers with the two-parameter Weibull distribution as the underlying distribution, while taking into account the covariates as inputs. By using DeepWeibull, we aim to improve the accuracy of the predictions for transformer reliability, especially for complex data with a large number of covariates.

2.1. DeepWeibull

The model configuration of DeepWeibull is presented in Figure 1, which consists of multiple full-connected layers followed by *ReLU*,

$$ReLU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0, \end{cases} \quad (3)$$

as the activation function. To address the non-negativity of the shape and scale parameters of Weibull models, different from Bennis et al. (2020, 2021), we utilize *SoftPlus*,

$$SoftPlus(x) = ln(1 + e^{x}), \quad (4)$$

as the final activation function, which guarantees positive values for both β and η. Moreover, we incorporate scaling factors to adjust the shape and scale parameters before the final outputs, which can be learned during training or pre-set based on prior empirical knowledge.

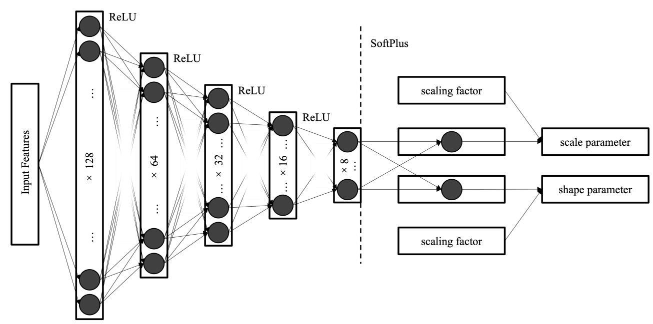

Figure 1. Model configuration of DeepWeibull.

During the inference stage, the covariates are input into the deep neural networks and the outputs are the shape and scale parameters for the two-parameter Weibull distributions. Another contribution of our work is to address both censoring and truncation problems when using DL-based survival models, which is covered in the following introduction to the loss function in detail.

2.2. Loss Function

When performing MLE with statistical models, the likelihood function is built as the basis for optimization. Considering both censoring and truncation problems, the following four cases are introduced for each subject i aged t_i:

- Non-censored ($\delta_i = 1$) and truncated ($\tau_i = 1$):

$$\frac{f(t_i)}{1 - F(t_i^L)}, \ \delta_i = 1, \ \tau_i = 1, \quad (5)$$

  where δ_i indicates if subject i is censored ($\delta_i = 0$ if censored), τ_i indicates if subject i is truncated ($\tau_i = 1$ if truncated), $f(t)$ is the $p.d.f.$ and $F(t)$ is the $c.d.f.$, t_i^L is the truncated time equal to the difference between the origin time of subject i and the corresponding time when the observation starts. Therefore, the truncation problem is considered through conditional probability by dividing $1 - F(t_i^L)$, indicating the subject i has survived up to t_i^L when the observation starts.

- Censored ($\delta_i = 0$) and truncated ($\tau_i = 1$):

$$\frac{1 - F(t_i)}{1 - F(t_i^L)}, \ \delta_i = 0, \ \tau_i = 1. \quad (6)$$

  where $1 - F(t_i)$ is used instead of $f(t_i)$ to cope with the censoring problem, indicating the subject i has survived up to t_i during observation with no failure occurance.

- Non-censored ($\delta_i = 1$) and non-truncated ($\tau_i = 0$):

$$f(t_i), \ \delta_i = 1, \ \tau_i = 0. \quad (7)$$

- Censored ($\delta_i = 0$) and non-truncated ($\tau_i = 0$):

$$1 - F(t_i),\ \delta_i = 0,\ \tau_i = 0. \tag{8}$$

With these different cases, the likelihood function can be obtained by the products of all subjects. Supposing we have a survival dataset $\mathbf{x} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, where $\mathbf{x}_i = (t_i, \delta_i, \tau_i, t_i^L)$ and N is the total number of subjects, then the likelihood function can be computed as (Hong et al., 2009),

$$\begin{aligned} L(\beta, \eta \,|\, \mathbf{x}) = \prod_{i=1}^{N} [\frac{f(t_i)}{1 - F(t_i^L)}]^{\delta_i \tau_i} [\frac{1 - F(t_i)}{1 - F(t_i^L)}]^{(1-\delta_i)\tau_i} \\ f(t_i)^{\delta_i(1-\tau_i)}[1 - F(t_i)]^{(1-\delta_i)(1-\tau_i)}. \end{aligned} \tag{9}$$

Subsequently, by replacing the *p.d.f.* and *c.d.f.* with Equations 1 and 2, the log-likelihood function $LL(\beta, \eta \,|\, \mathbf{x})$ as the log-form of Equation 9 can be formulated as,

$$\begin{aligned} LL(\beta, \eta \,|\, \mathbf{x}) &= ln(\prod_{i=1}^{N} [\frac{f(t_i)}{1 - F(t_i^L)}]^{\delta_i \tau_i} [\frac{1 - F(t_i)}{1 - F(t_i^L)}]^{(1-\delta_i)\tau_i} \\ &\quad f(t_i)^{\delta_i(1-\tau_i)}[1 - F(t_i)]^{(1-\delta_i)(1-\tau_i)}) \\ &= \sum_{i=1}^{N} \delta_i \tau_i [ln(\frac{\beta}{\eta}) + (\beta - 1) ln(\frac{t_i}{\eta}) - (\frac{t_i}{\eta})^{\beta} + (\frac{t_i^L}{\eta})^{\beta}] \\ &\quad + (1 - \delta_i)\tau_i[-(\frac{t_i}{\eta})^{\beta} + (\frac{t_i^L}{\eta})^{\beta}] \\ &\quad + \delta_i(1 - \tau_i)[ln(\frac{\beta}{\eta}) + (\beta - 1) ln(\frac{t_i}{\eta}) - (\frac{t_i}{\eta})^{\beta}] \\ &\quad + (1 - \delta_i)(1 - \tau_i)[-(\frac{t_i}{\eta})^{\beta}]. \end{aligned} \tag{10}$$

The log-likelihood function derived is the key to MLE when dealing with both censoring and truncation problems. The loss function to train DeepWeibull is then introduced with the negative log-likelihood function as,

$$\mathcal{L}\,(\beta, \eta \,|\, \mathbf{x}) = -LL(\beta, \eta \,|\, (t_i, \delta_i, \tau_i, t_i^L)_{i=1,2,...,N}). \tag{11}$$

2.3. Validation Framework

To validate the prediction accuracy of a survival model, the concordance index (C-Index or CI) is a commonly used indicator. It measures the ability of a model to correctly rank the survival times of pairs of individuals. Let y_i be the survivability of individual i, and let $\hat{y}_i$ be the predicted survivability from a survival model. The CI is defined as,

$$CI = \frac{\sum_{i<j} I(\hat{y}_i < \hat{y}_j) I(y_i < y_j) + \frac{1}{2} \sum_{i<j} I(\hat{y}_i = \hat{y}_j)}{\sum_{i<j} I(y_i < y_j)} \tag{12}$$

where $I(\cdot)$ is the indicator function and only equal to 1 when the condition is fulfilled. Intuitively, the CI can be interpreted as the probability that, given two individuals with different survival times, the model correctly predicts which individual will survive longer. A model with a CI of 0.5 is no better than random guessing, while a model with a CI of 1.0 makes perfect predictions. The CI is a useful metric for comparing model prediction accuracy, but it does not assess calibration or goodness-of-fit. To compare the accuracy of predictions on the number of failures given a base population and time horizon, we propose a validation framework.

Given a survival dataset $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_N\}$, where $\hat{\mathbf{x}}_i$, different from $\mathbf{x}$, can be expanded as $(t_i, \delta_i, \tau_i, t_i^L, T_i^S, T_i^E)$. T_i^S is the calendar time when the subject i enters the study, and T_i^E is the calendar time when the observation on subject i ends. For censored data, T_i^E is the same calendar time when the observation ends, while for non-censored data it is the same calendar time when the event is observed. To perform prediction based on existing data, we need to revert the current survival data to its past presence in a given reference year T_{ref}. The reverted survival dataset $\hat{\mathbf{x}}^{T_{ref}}$ can be generated by:

- Excluding data with $T_i^S > T_{ref}$.
- For data with $T_i^S < T_{ref} < T_i^E$, $\hat{\delta}_i = 0$, $\hat{t}_i = T_{ref} - T_i^S$.

Then, the prediction on the number of failures given a time horizon h can be calculated as

$$\hat{N}(h) = \sum_{i=1}^{N^{T_{ref}}} \frac{F(\hat{t}_i + h) - F(\hat{t}_i)}{1 - F(\hat{t}_i)} \tag{13}$$

where $N^{T_{ref}}$ is the total number of subjects in $\hat{\mathbf{x}}^{T_{ref}}$ and $F(\cdot)$ is the estimated $c.d.f.$ following Weibull distribution.

3. Experimental Results

3.1. Preparation

The transformer lifetime data we used for experiments is provided by a power company, including the distribution transformers at over 12,000 substations. As the collection of failure data only started after year of 2000, there is truncation problem existing when performing survival analysis. The event is defined as a failure happened to the transformer leading to the termination of service.

Two-parameter Weibull models were estimated using both traditional optimization with MLE and DeepWeibull. In this paper we adopted Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher, 2013) for MLE, denoted as MLE (BFGS). In the following experiments with MLE (BFGS), the data was first partitioned into 15 sets by distinct pairs of voltage classes and manufacturers, and the Weibull distributions were estimated with each set independently. When training DeepWeibull, the voltage class and manufacturer were the input features, converted into binary values with one-hot encoding. Hence, only one DeepWeibull model was trained with the entire dataset, while different

Table 1. Shape and normalized scale parameters for Weibull models estimated from Type I experiment for each group.

No.	β_{mle}	$\tilde{\eta}_{mle}$	β_{dw}	$\tilde{\eta}_{dw}$
1	2.06	0.93	2.23	0.67
2	1.82	0.89	2.06	0.63
3	1.56	0.75	1.76	0.54
4	7.58	0.31	2.57	0.77
5	5.78	0.31	2.14	0.65
6	14.36	0.32	3.30	0.98
7	11.05	0.31	2.37	0.71
8	10.91	0.31	2.47	0.74
9	12.91	0.31	2.28	0.69
10	0.93	0.79	1.49	0.46
11	3.64	0.22	1.81	0.56
12	2.15	0.63	2.14	0.65
13	2.82	0.51	2.34	0.71
14	2.16	1.01	2.19	0.66
15	15.18	0.32	2.70	0.81

shape and scale parameters can be output on condition of different inputs. For DeepWeibull, the configuration of the network is illustrated in Figure 1. As there are 3 different voltage classes and 10 different manufacturers in the experiment data, the features with dimension of 13 are input into the model.

To validate the model performance comprehensively, Type-I and Type-II experiments are introduced:

- Type-I: We split the dataset into training and testing sets in a time-independent manner. The model's parameters were estimated using the training set, and the performance was evaluated on the testing set using the CI and sum of log-likelihood, with T when *c.d.f.* $F(T) = 0.05$ as the predicted survivability $\hat{y}_i$.
- Type-II: We reverted the data to its presence in previous times by setting different reference years T_{ref} as introduced in Section 2.3. The parameters were estimated using the reverted data $\hat{\mathbf{x}}^{T_{ref}}$ and then used to predict the total number of failures in a given horizon h.

For Type-II experiments, we evaluated the performance of different models using the mean absolute error (MAE),

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|N_i(h) - \hat{N}_i(h)|, \tag{14}$$

and root mean squared error (RMSE),

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(N_i(h) - \hat{N}_i(h))^2}, \tag{15}$$

where $n = 15$ is the total number of groups, $\hat{N}_i(h)$ is the predicted number of failures for group i as in Equation 13, and $N_i(h)$ is the actual number of failures within the horizon.

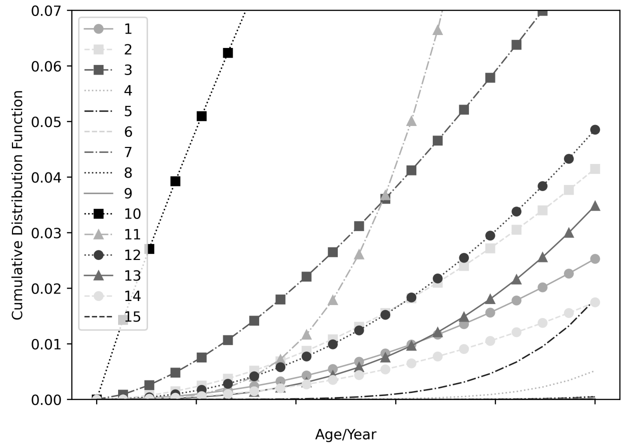

Figure 2. Cumulative distribution functions of Weibull models estimated by MLE (BFGS) $w.T.$ on Type I experiment training data for different groups.

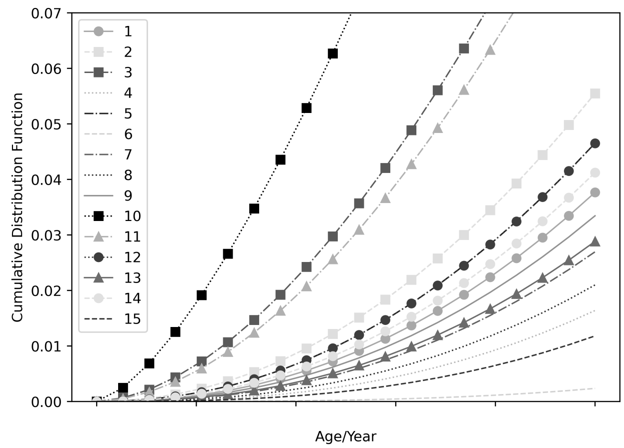

Figure 3. Cumulative distribution functions of Weibull models estimated by DeepWeibull $w.T.$ on Type I experiment training data for different groups.

3.2. Type-I Experimental Results

The abbreviations $w.o.T.$ and $w.T.$ are used to denote *without truncation* and *with truncation*, respectively, referring to whether or not the truncation problem was considered in the model, while all tests considered censoring problem. As shown in Table 1, the shape and normalized scale parameters estimated by MLE (BFGS) $w.T.$ (β_{mle} and $\tilde{\eta}_{mle}$) and DeepWeibull $w.T.$ (β_{dw} and $\tilde{\eta}_{dw}$) using Type-I experiment training data for different groups are presented, which are also visualized with *c.d.f.* in Figures 2 and 3 respectively.

To evaluate the prediction accuracy of the selected models, the CI and the sum of log-likelihood are used. The results of the testing phase are presented in Table 2. The DeepWeibull $w.T.$ model, which considers both censoring and truncation problems, shows the optimal performance on the test data, as

Table 2. Concordance index (CI) and sum of log-likelihood (LL) evaluated on the test data from Type-I experiments.

Method	CI	Sum of LL
MLE (BFGS) $w.o.T.$	0.61	-206.14
MLE (BFGS) $w.T.$	0.68	-203.59
DeepWeibull $w.o.T.$	0.73	-201.91
DeepWeibull $w.T.$	**0.73**	**-201.78**

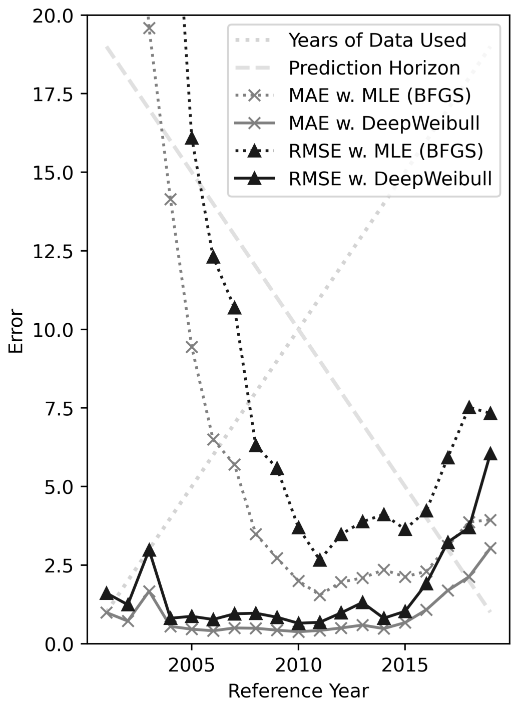

Figure 4. Testing results from Type-II experiments.

indicated by the highest CI and sum of log-likelihood. Moreover, considering the truncation problem improves the performance of both methods, as observed in the results. While MLE (BFGS) shows more significant improvement when the truncation problem is considered during estimation, DeepWeibull can still achieve less biased estimations even without considering it.

3.3. Type-II Experimental Results

For Type-II experiments, we have used 2001 to 2019 as reference years T_{ref} for testing, with the end of the prediction horizon fixed at 2020. This design allowed for a gradual increase in the number of failure data available for model training as the reference year approached 2020, while also shortening the prediction horizon, i.e., $h = 2020 - T_{ref}$.

As shown in Figure 4, the DeepWeibull model outperforms the traditional methods on failure number prediction in terms of MAE and RMSE. When the data are limited, e.g., with reference year from 2001 to 2005 where only 1 to 5 years of failure data since 2000 are available, the Weibull models estimated using MLE (BFGS) fail to reliably predict failures and show significant errors. However, as the amount of data used for modelling increases and the prediction horizon becomes narrower, i.e., after 2015, both methods shows similar performance, but DeepWeibull still outperforms the traditional approach.

4. Conclusion

In this paper, we proposed a DL-based approach for estimating Weibull distributions of transformer failure probability and demonstrated its effectiveness in prediction. Our DeepWeibull model outperformed traditional MLE using BFGS on two types of experiments, indicating its superiority in modelling and prediction accuracy.

Meanwhile, we investigated the impact of considering the truncation problem on the performance of survival models and found that it improved the performance of both traditional methods and our model. However, the DeepWeibull model is less susceptible to the truncation problem. Our results also highlighted the importance of selecting an appropriate prediction horizon and training data volume for predicting the number of failures given a fleet of assets.

In conclusion, our results indicate that DL models, particularly the DeepWeibull model, can be an effective tool for modelling and predicting failure probability. Future research could focus on improving the interpretability of the model and optimizing hyperparameters when working with datasets of different profiles.

Acknowledgment

This research is supported by SP Group, the Energy Market Authority, Singapore, and Nanyang Technological University through Project 3: Failure Mode Analysis and Mitigation Optimisation under the Energy Programme (EMA-EP010-SNJL-003).

References

Bennis, A., Mouysset, S., & Serrurier, M. (2020). Estimation of conditional mixture weibull distribution with right censored data using neural network for time-to-event analysis. In (Vol. 12084, p. 687-698).

Bennis, A., Mouysset, S., & Serrurier, M. (2021). DPWTE: A deep learning approach to survival analysis using a parsimonious mixture of weibull distributions. In *Artificial neural networks and machine learning – ICANN 2021* (p. 185-196). Springer International Publishing.

Chmura, L., Morshuis, P. H. F., Gulski, E., Smit, J. J., & Janssen, A. (2011). Statistical analysis of subcomponent failures in power transformers. In *2011 IEEE* *electrical insulation conference (EIC).*

Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.

Hong, Y., Meeker, W. Q., & McCalley, J. D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *The Annals of Applied Statistics*, *3*(2), 857–879.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, *18*(1), 1–12.

Nagpal, C., Li, X., & Dubrawski, A. (2021). Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 3163–3175.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*.