Hierarchical spline for time series forecasting: An application to Naval ship engine failure rate

Hyunji Moon[1] and Jinwoo Choi[1]

[1]Affiliation not available

September 10, 2020

Abstract

Predicting equipment failure is important because it could improve availability and cut down the operating budget. Previous literature has attempted to model failure rate with bathtub-formed function, Weibull distribution, Bayesian network, or AHP. But these models perform well with a sufficient amount of data and could not incorporate the two salient characteristics; unbalanced category and sharing structure. Hierarchical model has the advantage of partial pooling. The proposed model is based on Bayesian hierarchical B-spline. Time series of the failure rate of 98 Republic of Korea Naval ships have been modeled as hierarchical model, where each layer corresponds to ship engine, Engine type, and Engine archetype. As a result of the analysis, the suggested model predicted the failure rate of an entire lifetime accurately in multiple situational conditions, including the amount of prior knowledge of the engine.

Introduction

Forecasting failure rate is important as it serves as a standard for preventive measure, inventory management. Both over and underestimation of failure are detrimental to the system. Underestimation can lead to mission failure due to failure, overestimation can lead to wasted budget and reduced operational efficiency due to excessive spare part purchases. Therefore taking account of features of failure data into the model is important. Two characteristics of failure rate data, unbalanced category and sharing structure, are the main motivation for this paper and we propose hierarchical spline model for improvement. First, unbalanced category refers to the fact that collected data corresponding to each age are unbalanced for each category; product type, for example. Second, is sharing structure. In our case of predicting the failure rate of an engine of each ship, as engines are shared among ships, ships with the same type of engine display similar failure rate patterns. The underlying process also supports the empirical results, as the same engine types share design patterns and are made from the same factory. Hierarchical model provides a systemic structure to improve both unbalanced category and sharing characteristic of data. In our problem setting, even the failure rate of an age period where data of a certain ship engine is unavailable could be forecasted as parameters could be borrowed from other types of ships and engines. For this purpose, we have constructed our three-layer model as the following: a root layer that accounts for the core characteristics of an engine, i.e. engine archetype, a second layer which corresponds to each type of an engine, and lastly, the final layer that explains the specific characteristics of each ship. The proposed model has additional advantages in terms of forecasting the failure of new engine types. Republic of Korea (ROK) Navy battle ship evlove continuously; for example, FF (Fate Frigate) class have been replaced by FFG (Fast Frigate Guided-missile). Forecasting the failure rates of a new battle ship is clueless, but necessary. Most existing time series models such as ARIMA or ETS(exponential smoothing) model struggles in situation where no quantitative data exist. However, in hierachical model it is possible to constuct the outline of the failure function based on the prior qualitative information. For instance, as we will elaborate in section 5, engines constructed in similar era show similar patterns. Therefore, information on which era the unforeseen engine was made could be utilized to forecast its failure rates. The proposed model has additional advantages in terms of forecasting the failure of new engine types. ROK Navy battleship evolve continuously; for example, FF (Fate Frigate) class has been replaced by FFG (Fast Frigate Guided-missile). Forecasting the failure rates of a new battleship is clueless, but necessary. Most existing time series models such as ARIMA or ETS (exponential smoothing) model struggles in situations where no quantitative data exist. However, in hierachical model it is possible to construct the outline of the failure function based on the prior qualitative information. For instance, as we will elaborate in section 5, engines constructed in a similar era show similar patterns. Therefore, information on which era the unforeseen engine was made could be utilized to forecast its failure rates. The main contribution of this paper lies in applying hierarchical spline (HS) model to address unbalanced category and sharing structure of failure data from ROK Navy. Compared to the previous models, the proposed model not only improves overall forecast accuracy but also is capable of forecasting failure rates for categories with scarce data robustly. Moreover, the hypothetical similarity between each category can be tested and proved using our model; this enables users to utilize the qualitative knowledge on the unforeseen, ships with new engines for example, for forecasting. These results, when used as a reference for maintenance policy and budget allocation, could contribute greatly to the Navy's operating system. However, this model is not limited to Naval domain. When it comes to forecasting failure rates, the circumstances where data are hierarchical, unbalanced, or insufficient are common and therefore, our model is widely applicable. For example, mechanical equipment consists of a number of parts. The generator, which is a part of the wind turbine, is composed of parts such as a motor and a transformer (Scheu, M. N. et al., 2019) in a hierarchical structure. Using the HS model, it is also possible to predict the failure of equipment components in a hierarchical structure. The remainder of this paper consists of five sections. Section 2 introduces the background behind failure forecasting in marine as well as the key concepts upon which the HS model is based. Section 3 introduces HS model and explains the advantage especially in terms of characteristics of the real data. In section 4, data obtained from the ROK Naval ships is introduced and experimental models are described and compared. Section 5 contains an analysis of the experimental models, and lastly, conclusions are presented in section 6.

Literature Review

2.1 Failure rate in Naval ship setting

Before the mission, each naval ship is equipped with a forecasted amount of spare engines. An underestimated forecast has a risk of mission failure as spares parts cannot be resupplied during mission times. An overestimated forecast may lead to reduced operating efficiency due to a load of unnecessary spare parts. Moreover, from a system point of view, overestimation induces unnecessary use of budget and even lead to inventory shortage for other ships. So, defining the optimal set of spare parts is crucial for mission success (Zammori et al., 2020). For accurate prediction, several special features resulting from Navy's system should be noted. First of all, unbalances are observed in two categories of the data: age period and engine types. There was only a short period of failure rate data compared to the entire lifetime. In our case, for example, an early age has less data than the rest of the age period; this might be problematic as the failure rate of young ships is needed for operation. Also, the distribution of ships for each engine type category is not balanced. In our dataset with 98 ships, there are 5, 27, 43, 17, 6 ships for each engine type category. In this case, while a satisfactory model could be obtained from engine with a large amount of data, other models might suffer lack of data problems. Moreover, the similarity between ships and engines should also be noted as they undergo are expected as all the engines are under the same maintenance process; planned maintenance is performed by ROK Navy regardless of the engine type (Yoo, J. M. et al., 2019). Based on these circumstances, where ships as well as engines share certain qualities, the model with layered parameter structure is needed; it should be able to learn the specific structure between and within each layer from the data.

2.2 Failure forecasting models

Several models exist such as ARIMA, exponential smoothing, and seasonal trend decomposition using Loess (Hyndman and Athanasopoulos, 2018) that could model time series characteristics of failure rate. Among the existing time series models, Prophet, which adopts Bayesian generalized additive model shows high accuracy. Moreover, it decomposes time series into trend, seasonal, other regressor factors which enhances both its application and interpretability (Taylor, S.J., & Letham, B., 2018). More specific models concentrating on the characteristics of failure have been suggested. A bathtub is typical shape pattern of failure rate. Also, Weibull or Poisson distribution are often used as a distribution of failure rate. Wang and Yin (2019) performed failure rate forecasting through stochastic ARIMA model and Weibull distribution. Time series has been decomposed into bathtub-shape assumed trend and stochastic factors. Parameters of the Weibull distribution were separately learned for the increase, decrease, and flat period of the bathtub. The stochastic element was obtained using ARIMA, and the time series failure rate was calculated as the sum of the trend and stochastic elements. Sherbrooke (2006) proposed Pareto-optimal algorithms, named constructive algorithms, based on Poisson distribution. However, it had limits in determining the parameter. Zammori et al. (2020) tried to solve the problem of parameter estimation of Sherbrooke's (2006) model by applying time-series Weibull distribution. Other attempts such as Pareto-optimal, Monte-Carlo(Sherbrooke, 2006), ARMA, and least-squares logarithm (Wang & Yin, 2019) have been made to add the effect of stochastic factors to this distribution. Attempts have been made to integrate time series models with information about system architecture. In the risk analysis of deepwater drilling riser fracture (Chang, y. et al., 2019), Bayesian network was used to predict the fracture failure rate. Bayesian network could also used to analyze and prevent the cause of a ship's potential accidents (Afenyo, M. et al., 2017). Time series forecasting based on Bayesian network (Dikis, K., & Lazakis, I. , 2019) and Analytic Hierarchy Process (AHP) (Yoo, J. M., Yoon, S. W., & Lee, S. H., 2019) illustrate these approaches. They are based on the assumption that equipment, engines for example, within the same group follow similar failure patterns.

2.3. Hierarchical model

Hierarchical model has an edge in representing the features of Navy data introduced in 2.1; unbalanced category and sharing structure, by information pooling. Gelman et al. (2005) explained that hierarchical models are highly predictive because of pooling (Gelman et al., 2013). When hierarchical model is used, there is almost always an improvement, but to different degrees that depends on the heterogeneity of the observed data (Gelman, 2006a). When updating the model parameters, such as prior parameters, the relationship between the part of the data being used and the whole population should always be considered. Pooled effects between subclusters are partial as they are implemented through shared hyperparameters, not parameters. In a Bayesian hierarchy, the balance of fit can be learned by using hyperpriors. By properly setting the hyperprior structure, we can find a reasonable balance between over-fitting and under-fitting, as hyperpriors are known to serve as a regularizing factor. Many examples of applying hierarchical structure in cross-sectional data exist in diverse domain, such as ecology, education, business, and epidemiology (McElreath, 2020). The structure of cross-sectional data where the whole population is divided into multiple and nested subcategories provides an excellent environment for a hierarchical model. Previous literature on comparing the education effects of multiple schools has shown that incorporating the nested structure of the state, school, and class in the model had substantial improvement in terms of accuracy and interpretability (Rubin, 1981). 2.4. Model evaluation measures Time series cross-validation and k-fold cross-validation, along with the expanding forecast method, can be used to measure forecast accuracy in time series (Hyndman and Athanasopoulos, 2018). Several sets of training and test data are created in a walk-forward mode, and forecast accuracy is computed by averaging over the test sets. Various measures of forecast error exist, including the mean absolute, root mean squared and mean absolute percentage error. When a large difference of scale exists in the data, using a scaled error measure is recommended. The mean absolute scaled error is recommended for comparing forecast accuracy across multiple time series (Hyndman & Koehler, 2006). Information criteria can be used to measure the fit of a model in Bayesian models include widely applicable information criterion (WAIC) and the leave-one-out cross-validation (LOOCV); they are preferred to other criteria such as Akaike information criterion (AIC) and deviance information criterion (DIC) (Vehtari and

Lampinen, 2002). For Bayesian models, where the estimation of parameters is based on sampled results, it is essential to check whether chains have reached their convergence before comparing models. For these purposes, trace plots and numerical summaries such as the potential scale reduction factor, Rhat (Stan Development Team, 2017b) are used. Rhat lower than 1.1, for each parameter, is recommended.

Model and Data

3.1. Data Dataset consists of 98 ship engines that could be categorized with five types of engines. Therefore, our hierarchical model has a 1-5-98 structure; 1 engine archetype, 5 engine types, and 98 ship types. The numbers of ships in the five categories are also different as in section 2.1. Fig. 1 shows the age, type of engine, and ship of existing data. As can be seen from the figure, the amount of data for each category is highly unbalanced. Moreover, the similarity between data under the same category could be inferred; for example, data with the same type of engine display a similar age period.

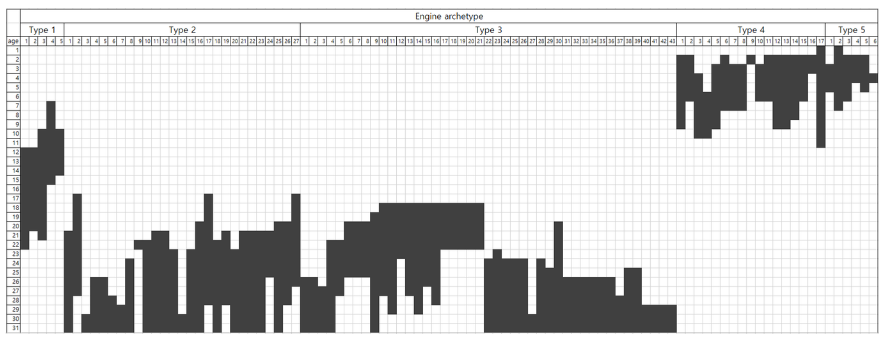


Fig. 1 Existing data by age and engine type

By arranging the failure data of 98 propulsion ship engine categorized into 5 types according to their lifetimes, we got the failure rate data for the approximate total life cycle of 31 years (It is expressed as approximate because it is not based on the total life data of each ship, but is the data of the total life created by combining parts). The data only include maintenance records from the military direct maintenance workshop. After the construction, the initial repair of the shipyard(warranty repair) was not included in the data. Also, data may not be entered into the management system for other reasons. Therefore, it can be said that the failure data of the part corresponding to the initial part after constructing is reflected less than actual.

3.2. Model and process

Naval ship engines applied in the proposed model are classified as Fig. 2. The ship engine (layer 3) of each ship belongs to the same engine type (layer 2), and 5 types belong to the entire engine archetype (layer 1).

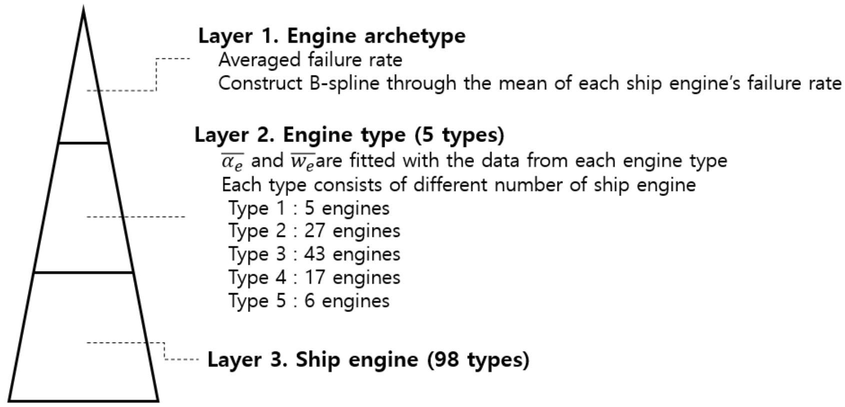


Fig. 2 Hierarchical structure of ship's engine failure Since the naval ship data is nonlinear time series data, polynomial and spline regression are considered. In polynomial regression, to achieve flexibility, a degree should be increased; however, the risk of overfitting becomes higher with its degree. To prevent this and to endow the model a form of locality, the B-Spline model is suggested: overall life span, or age period, is first divided into several sections. Then low-dimensional polynomial is fitted for each section to form a piecewise polynomial spline. The third layer of ROK naval ship hierarchy, representing the ship engine, is modeled with B-Spline. As can be seen from Equation 1, parameters corresponding to each layer are modeled in a way that could enable the pooling possible; though hyperparameter sharing structures.

$$Y_s \sim \text{Normal}(\mu_s,\ \sigma_y)$$

$$\mu_s = \alpha_s + \Sigma_{k=1}^{K} w_{k,\ s} B_k$$

$$\alpha_s \sim \text{Normal}(\ \alpha_e,\ \sigma_\alpha)$$

$$w_s \sim \text{Normal}(\ w_e,\ \sigma_w)$$

$$\alpha_e \sim \text{Normal}(\ \alpha_0,\ \sigma_{\alpha})$$

$$w_e \sim \text{Normal}(\ w_0,\ \sigma_{w})$$

$$\sigma_\alpha \sim \text{Gamma}(10,\ 10)$$

$$\sigma_w \sim \text{Gamma}(10,\ 10)$$

$$\sigma_{\alpha} \sim \text{Exponential}(1)$$

$$\sigma_{w} \sim \text{Exponential}(1)$$

$\sigma_y \sim$ Exponential(1) Equation 1

To be more specific, based on the overall average failure rate of 98 ships, B-spline is pre-fitted to obtain the hyperparameters.α_0, $w_0$are set as the posterior sample mean with their priors given as Equation 2. We used the intercept from linear regression fit as the prior mean ofα_0. Next, with these layer1 parameters, engine-specific parameters,α_e andw_e are learned. Note that weight, w, is a vector whose length is determined by the number of knots.

$$\alpha_0 \sim \text{Normal}(\ I,\ 1)$$

$w_0 \ \sim$ Normal(0, 1)Equation 2

We used Stan as a probalistic programming language which has the advantage of faster estimation of Bayesian models in ecology in MCMC. Stan HMC is effective as model complexity is large and correlated variables are large. Stan was used in consideration of future developments (complexity and variables to be added) of the proposed model. Stan code of the proposed model is included in Appendix. The working process is organized as shown in Fig. 3. From failure rate data, a rough trend of the failure rate over a lifetime is deduced. This trend is the basis for the number of B-spline knot and hyperprior estimation in the construct hierarchical model. Apply the estimated b-spline and hyperprior to the model, and parameter fit by learning the data of each layer through MCMC sampling.

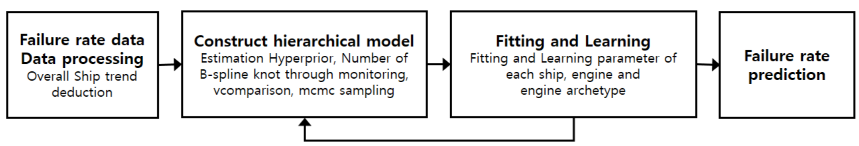


Fig. 3. Working process**5. Results and Discussion**5.1. Accuracy comparison Fig. 4 is the result of HS model predicting the total lifetime failure rate of 98 ship engines. Spots are points of failure with 98 data. Considering the missing data in the early period due to warranty repair (as in section 3.1), they are bathtub-formed, as some previous literature has suggested (as in section 2.2). Among the models introduced in section 2.2, Prophet and Arima are selected as baseline models each for its high accuracy and popularity, respectively. Accuracy comparison was performed in two ways considering the model's application situation. In general, when it is necessary to introduce a new ship engine or predict the ship engine in use, refer to the data of the same engine type. The prediction accuracy of the ship engine is used as a reference to predict the spare parts of the ship engine in use, and the prediction accuracy between the engine type and the ship engine is used as a reference when introducing a new ship engine. Therefore, first (Table 1), the accuracy of ship engine predicted and ship engine actual values were compared, and second (Table 2), the accuracy of engine type predicted and ship engine actual values were compared. RMSE (Root Mean Square Error) was used for the error measure.

Table 1. RMSE (Ship engine predicted VS Ship engine actual)

HS	Prophet	ARIMA
1.0349	1.0875	1.2587

Table 2. RMSE (Engine type predicted VS Ship engine actual)

Data (number of data)	HS	Prophet	ARIMA
Type 1 (5)	1.1273	1.1294	1.0705
Type 2 (27)	0.9531	1.0417	1.1817

Data (number of data)	HS	Prophet	ARIMA
Type 3 (43)	1. 1191	1.2958	1.2741
Type 4 (17)	0.8466	1.0007	0.8903
Type 5 (6)	1.1305	0.903	0.7811
mean	1.0252	1.1421	1.1899

The first was to compare the average by obtaining the prediction accuracy of each of the 98 ship engines. To model the 3-layer dataset, only one option exists for hierarchical model. This is because the hierarchical model predicts 3-layer data using information of all layers. For Prophet and ARIMA, which are unable to represent the hierarchical structure, input data should be preprocessed, by averaging, to learn the parameters. As can be seen from Table 1, RMSE of HS model was the lowest. Since the HS model pools the information of the layers, the overall average RMSE is low. In Table 2, the HS model has a lower RMSE than the comparative model in the engine type of type 2 ˜ type 4, which has a relatively large number of data. Type 1 was less accurate than ARIMA, and type 5 had the worst performance. Since ARIMA or Prophet has no pooling effect, the smaller the number of data, the more noise is reflected in the prediction. Therefore, when the number of data is small, the prediction accuracy of ARIMA and Prophet can be high. This will be further explained in section 5.2. 5.2. Forecasting a new type of ship or engine When we fit the hierarchical model with failure rates of 98 ships, the learned results are stored in the model in the form of each parameter's distribution, i.e. posterior. For example, whose prior had exponential form would evolve into a posterior distribution. Bayes formula explains this mechanism. As discussed in the introduction, engine failure rate of a new type of engine or ship is frequently needed. Depending on its engine type, the way by which the hierarchical model should be applied differs. If its engine type is present among the data, the posterior of parameters corresponding to layer 2 could be used for the forecast (5.2.1). On the other hand, if the engine type is new as well, the only information we could borrow from the previous data are posteriors of layer1 parameters (5.2.2). The test set data is shown in Fig. 5. Types 1 ˜ 5 are the same as the 5 engine types included in train set data. One engine data was obtained for each engine type and prepared as a test set. Types 6 ˜ 10 are new engine types not included in train set data. Ship engine data corresponding to 5 new engine types were prepared as a test set for each type. 5.2.1. New ship type Previously learned posterior of α_e and w_e , could be directly used for predicting engine failure of a new ship engine, but whose engine type is not new; in other words not among the 5 trained engine types. As in section 5.1, Prophet and Arima were used as comparative models. The results are shown in Table 3.

Table 3. Test set data RMSE (New ship type which engine type was trained)

Engine type	HS	Prophet	ARIMA
Type 1	1.0953	1.4494	1.4872
Type 2	0.9672	0.9995	1.1351
Type 3	1.1955	1.2088	1.3278
Type 4	0.9996	0.9417	1.5695
Type 5	1.0862	1.3248	1.4601
mean	1.0688	1.1849	1.328

HS model had the lowest mean RMSE. In Section 5.1, HS model showed lower accuracy than type 1 and type 5 Prophet and Arima, but the effect of hierarchical information pooling of HS model was significant when applying new data that was not learned. HS model showed lower RMSE than Arima in all types. Prophet showed lower RMSE in type 4 than HS model, but did not show much difference. 5.2.2. New engine type Robust estimation is possible even when forecasting the failure rate of ship with unforseen engine type; information on engine archetype is stored in hyperparameters of layer1 with which forecast can be made. In other words, the resulting α_0 andw_0 value of a prefit, B-spline fit on averaged failure rate, are used forα_s andw_s from equation 1. Fig. 6, 7, and 8 show the prediction results of HS and comparative models. Commonly, the black dots are the location of the test set data. Small red dots are the location of the train data. The green lines of HS model are the prediction line of each ship engine, the red lines are the prediction line of each engine type, and the blue line is the engine archetype prediction line. The blue line of Prophet and ARIMA is the prediction line of the whole life cycle and corresponds to the archetype prediction line of the HS model. In the figures, y-axis is expressed as scaled values due to data privacy. The test for new engine type data has the following meaning. As technology evolves, the new engine type will replace the previous engine types. Or, depending on the purpose, you may need to introduce a new engine type that has not been used before. In this case, you can refer to similar types of engine type or engine archetype trend for prediction. In this section, we confirmed that HS performed well when only the information on engine archetype could be referenced. The prediction situation in this section is notable because it is the most difficult but necessary real situation. The accuracy was compared with Prophet and Arima. The result is shown in Table 4. In all new types, the HS model had a lower RMSE than the comparative model. In the case of type 6, it is located in the section of 10 to 20 years with the smallest number of train data. The HS model predicts relatively accurately by pooling information between layers even when the number of data is small.

Table 4. Test set data RMSE (New ship engine which engine type was not trained)

Model type	HS	Prophet	ARIMA
Type 6	1.1726	1.2141	1.2698
Type 7	0.9138	1.2417	1.6889
Type 8	1.17	1.2017	1.7933
Type 9	1.0244	1.1463	1.7604
Type 10	1.0328	1.0726	1.1365
mean	1.0627	1.17528	1.52978

In test set data (Fig. 5), data of type 1, 2, 3, 6, and 10 are commonly ship engines with an age of 5 years or more. These types showed the lowest RMSE of test set prediction results (Table 3, Table 4). In Section 3.1, the data of the initial part is said to reflect fewer data than the actual (because the data in this study only include the military direct maintenance workshop). In other words, it can be said that the data of the initial part is less reliable than other sections of data. In general, the warranty repair period of the ROK naval ship engine does not exceed 5 years. Type 1, 2, 3, 6, and 10 do not include data for the initial 5-year period with relatively low reliability. That is, the HS model proved better performance than the comparative model by showing a low RMSE in the test set data of all engine types with relatively reliable data. 5.3. Reflecting the qualitative knowledge Prediction can be improved in the presence of the qualitative knowledge, construction era of the new engine type, for example. This act of translating qualitative into quantitative knowledge could be justified by analyzing their relationship with the existing failure functions of five engine types. Fig. 9 and Table 5 give two interpretable results. First, the failure function is shifted down due to technical development. For example, type 5 engine which replaced type 2 engine, takes a similar form with its predecessor, the main difference being its intercept. Second, the period of engine design has a big effect on its failure results. Euclidean distance between each function (Table 5), indicates that engine pairs (4 vs 5) and (2 vs 3) are relatively close. This is in line with our knowledge that type 4 and type 5 are new engines while type 2 and type 3 are old (Type 4 and 5 engines were constructed after 2010 while type 2 and 3 engines were constructed in the 1990s.) Based on the qualitative knowledge on the closeness of new engine type with the existing engine types, posterior ofα_e andw_e of previous engine types could be used as a hyperprior for new ship'sα_s andw_s. Compared to using the original priorNormal(α_0, σ_α)andNormal(w_0, σ_w) from equation 1, this would give more accurate results as more prior knowledge could be reflected for the prediction.

Engine type	Euclidean distance
2 vs 5	0.2334
4 vs 5	0.2554
2 vs 3	0.2658
2 vs 4	0.2840
1 vs 3	0.2944
1 vs 2	0.3104
3 vs 4	0.3106
1 vs 5	0.3119
3 vs 5	0.3442
1 vs 4	0.4608

Table 5. Euclidean distances

Conclusions

We have proposed using HS to develop a hierarchical model for forecasting failure rates. This approach shines especially when the data have unbalanced category and structured characteristic. We demonstrated the applicability of the model using a real-world dataset of failure rate data from Naval ships and compared it with previous methods. Through these comparisons, we confirmed that the prediction performance of our novel model in the given dataset was greatly improved. Moreover, we have shown how qualitative knowledge, such as the belonging to the same series or construction era, could be incorporated into the model; this approach was justified by further analyzing the relationship between each parameter. These techniques could greatly improve Naval ship management efficiency. Some improvement could be noted for further studies. First, prevention repair which may affect the failure pattern could be considered. A more advanced model that incorporates the probability of failure after the prevention repair is needed to design a model. Second is convergence and evaluation measures. There were few instances with low E-BMI and effective sample size, n_eff. Improving the model in terms of higher E-BMI and n_eff measures would result in a better fit of the model. Thirdly, due to substantial operational differences between combat and non-combat ships, only combat ships are used in this paper. However, if the differences could be incorporated in the further models, by using categorical variables, a more accurate model could be possible based on a larger amount of data. The proposed model can contribute greatly to the following areas. First, failure rate prediction could be used as a quantitative reference when establishing a maintenance policy. Proper maintenance not only improves the availability and mission completion rates but also reduces the budget by reducing unnecessary maintenance. Second, from a broader perspective, the predicted failure trend can be a qualitative reference for designing the optimal life cycle of a ship. For instance, based on our results, the failure rate increases dramatically as the ship becomes senile. Therefore optimal retirement period could be decided by balancing the maintenance and construction cost. The proposed model can contribute greatly to the following areas. First, failure rate prediction could be used as a quantitative reference when establishing a maintenance policy. Proper maintenance not only improves the availability and mission completion rates but also reduces the budget by reducing unnecessary maintenance. Second, from a broader perspective, the predicted failure trend can be a qualitative reference for designing the optimal life cycle of a ship. For instance, based on our results, the failure rate increases dramatically as the ship becomes senile. Therefore optimal retirement period could be decided by balancing the maintenance and construction cost.

References

Afenyo, M., Khan, F., Veitch, B., & Yang, M. (2017). Arctic shipping accident scenario analysis using Bayesian Network approach. *Ocean Engineering,* 133, 224-230.

Barragan, J. F., Fontes, C. H., & Embiruçu, M. (2016). A *wavelet* -based clustering of multivariate time series using a multiscale SPCA approach. *Computers & Industrial Engineering, 95* , 144-155.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. & Riddell, A. (2017), Stan: A probabilistic programming language, *Journal of Statistical Software 76(1)* .

Chang, Y., Zhang, C., Wu, X., Shi, J., Chen, G., Ye, J., ... & Xue, A. (2019). A Bayesian Network model for risk analysis of deepwater drilling riser fracture failure. *Ocean Engineering* , 181, 1-12.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: a seasonal-trend decomposition. *Journal of official statistics, 6(1)* , 3-73.

Dagum, E. B., & Bianconcini, S. (2016). *Seasonal adjustment methods and real time trend-cycle estimation* . Springer International Publishing.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting, 22(3)* , 443-473.

Dikis, K., & Lazakis, I. (2019). Dynamic predictive reliability assessment of ship systems. *International Journal of Naval Architecture and Ocean Engineering* .

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* . Chapman and Hall/CRC.

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics, 48(3)* , 432-435.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* . OTexts.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis, 55(9),* 2579–2589.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22* , 679–688.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting, 18(3)* , 439-454.

Khanlarzade, N., Yegane, B., Kamalabadi, I., & Farughi, H. (2014). Inventory control with deteriorating items: A state-of-the-art literature review. *International journal of industrial engineering computations* , *5* (2), 179-198.

Kuo, R. J., & Li, P. S. (2016). Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform. *Computers & Industrial Engineering, 99* , 153-161.

Mackey, T. B., Barney, J. B., & Dotson, J. P. (2017). Corporate diversification and the value of individual firms: A Bayesian approach. *Strategic Management Journal, 38(2),* 322-341.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* . CRC press.

Scheu, M. N., Tremps, L., Smolka, U., Kolios, A., & Brennan, F. (2019). A systematic Failure Mode Effects and Criticality Analysis for offshore wind turbine systems towards integrated condition based maintenance strategies. *Ocean Engineering* , 176, 118-133.

Sherbrooke, C. C. (2006). *Optimal inventory modeling of systems: multi-echelon techniques* (Vol. 72). Springer Science & Business Media.

Shor, B., Bafumi, J., Keele, L., & Park, D. (2007). A Bayesian multilevel modeling approach to time-series cross-sectional data. *Political Analysis, 15(2)* , 165-181.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples.* Springer.

Tabandeh, A., & Gardoni, P. (2015). Empirical Bayes approach for developing hierarchical probabilistic predictive models and its application to the seismic reliability analysis of FRP-retrofitted RC bridges. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 1(2)* , 04015002.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72(1)* , 37-45.

Taieb, S. B., Yu, J., Barreto, M. N., & Rajagopal, R. (2017). Regularization in hierarchical time series forecasting with application to electricity smart meter data. In *Thirty-First AAAI Conference on Artificial Intelligence.*

Van der Auweraer, S., & Boute, R. (2019). Forecasting spare part demand using service maintenance information. *International Journal of Production Economics* , *213* , 138-149.

Vehtari, A., & Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural computation, 14(10)* , 2439-2468.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.*Statistics and Computing, 27(5),* 1413-1432.

Wang, J., & Yin, H. (2019). Failure Rate Prediction Model of Substation Equipment Based on Weibull Distribution and Time Series Analysis.*IEEE Access* , 7, 85298-85309.

Yoo, J. M., Yoon, S. W., & Lee, S. H. (2019). SNA-based Trend Analysis of Naval Ship Maintenance. *Journal of the Korea Society of Computer and Information* , 24(6), 165-174.

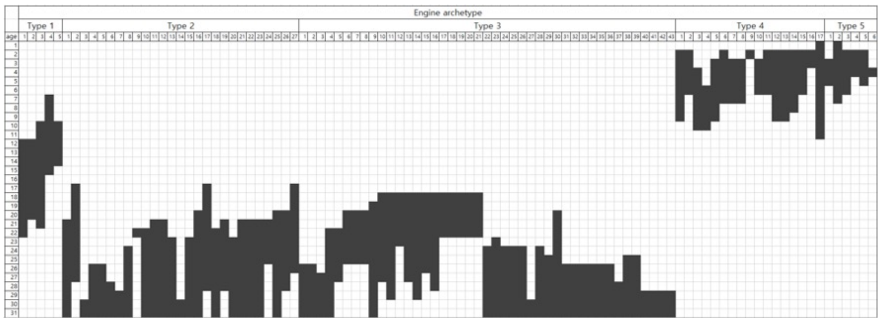
Zammori, F., Bertolini, M., & Mezzogori, D. (2020). A constructive algorithm to maximize the useful life of a mechanical system subjected to ageing, with non-resuppliable spares parts. *International Journal of Industrial Engineering Computations* , 11(1), 17-34.

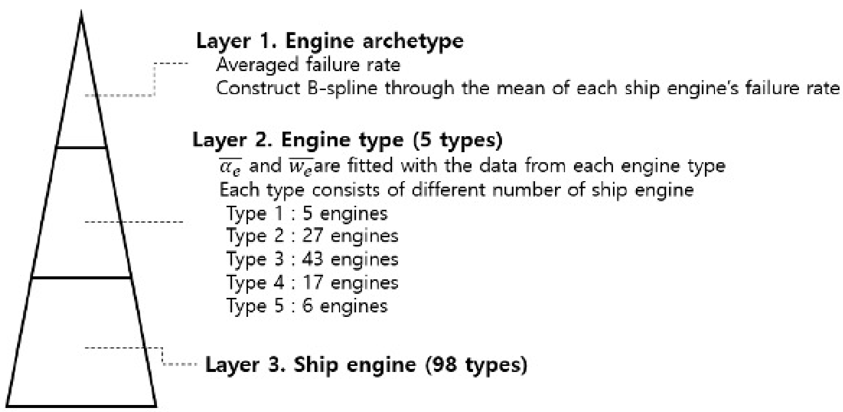
Appendix: Stan code HS model

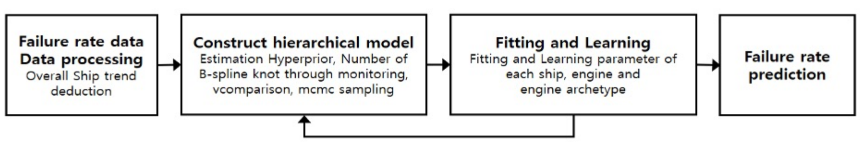
```
data {
int <lower = 1> K;
int <lower = 1> N;
int <lower = 1> T;
int <lower = 1> S;
int <lower = 1> E;
int <lower = 1> Age[N];
int <lower = 1> Ship[N];
int <lower = 1> S2E[S];
matrix[T,K] B;
real mu_a_bar;
real mu_w_bar[K];
vector [N] Y;
}
parameters {
```

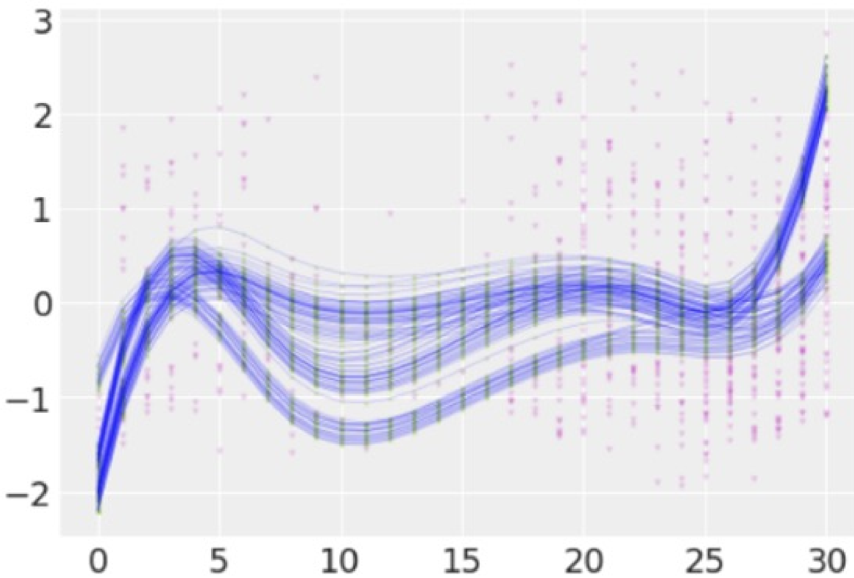
```
vector[S] a;
real a_bar[E];
vector[K] w[S];
vector[K] w_bar[E];
real<lower=0> s_a;
real<lower=0> s_w;
real<lower=0> s_a_bar;
real<lower=0> s_w_bar;
real<lower=0> s_Y;
}
transformed parameters {
vector [N] mu;
for (n in 1: N){
mu[n] = a[Ship[n]] + B[Age[n]] * w[Ship[n]];
}
}
model {
s_a ~ gamma(10,10);
s_w ~ gamma(10,10);
s_a_bar ~ exponential(1);
s_w_bar ~ exponential(1);
s_Y ~ exponential(1);
for (s in 1:S){
a[s] ~ normal(a_bar[S2E[s]], s_a);
w[s] ~ normal(w_bar[S2E[s]], s_w);
}
for (e in 1:E){
a_bar[e] ~ normal(mu_a_bar, s_a_bar);
w_bar[e] ~ normal(mu_w_bar,s_w_bar);
}
Y ~ normal(mu, s_Y);
}
generated quantities{
vector[N] log_likelihood;
```

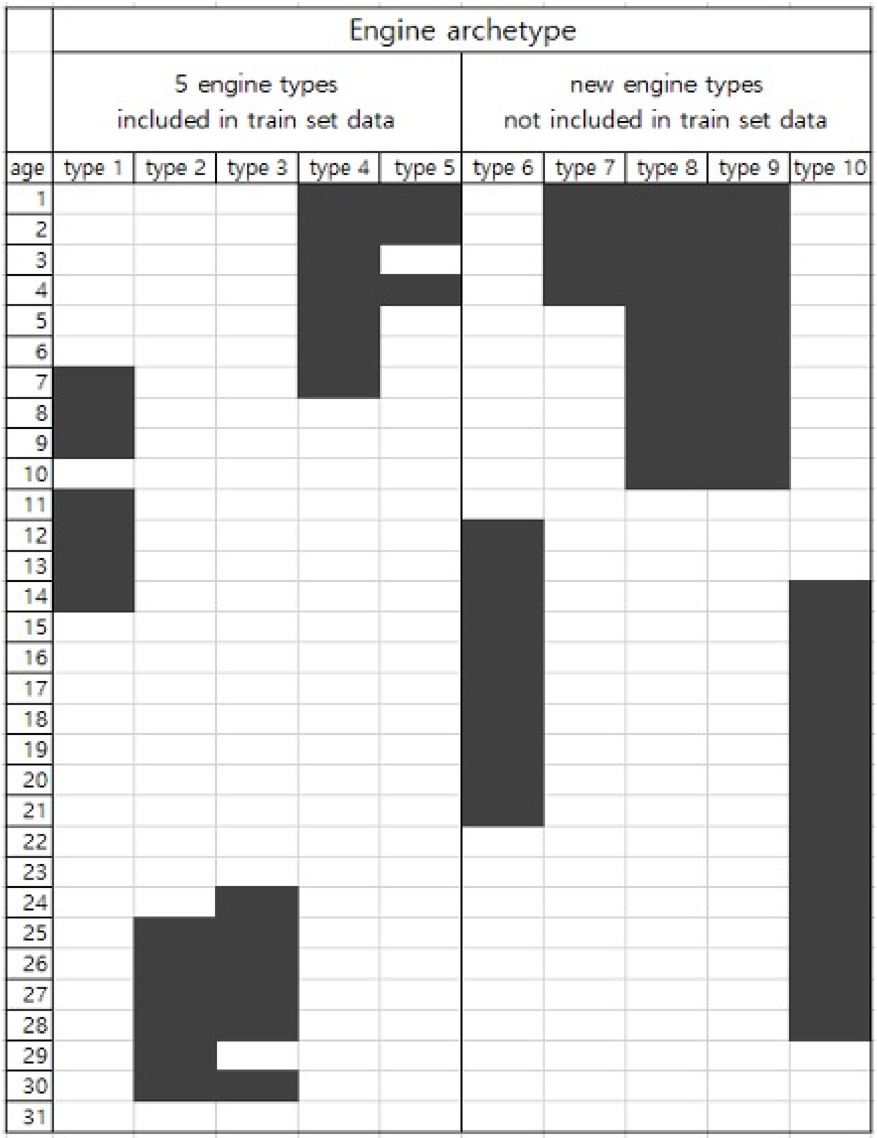
```
for (i in 1:N) {
log_likelihood[i] = normal_lpdf(Y[i]|mu[i], s_Y);
}
}
```

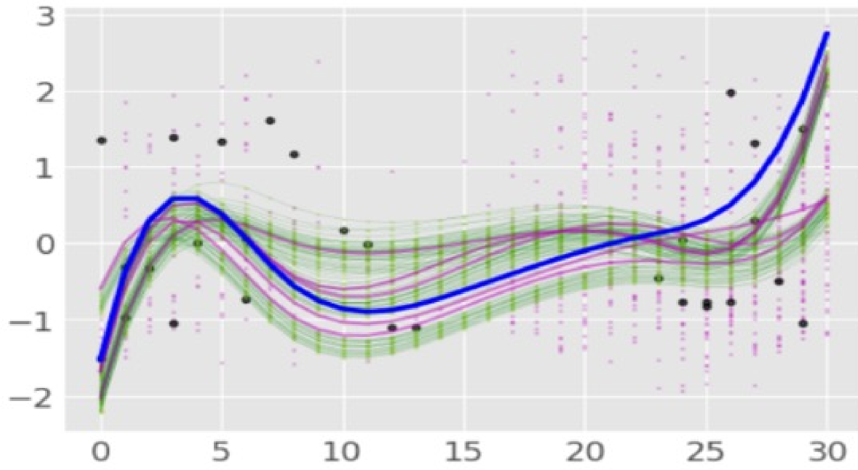


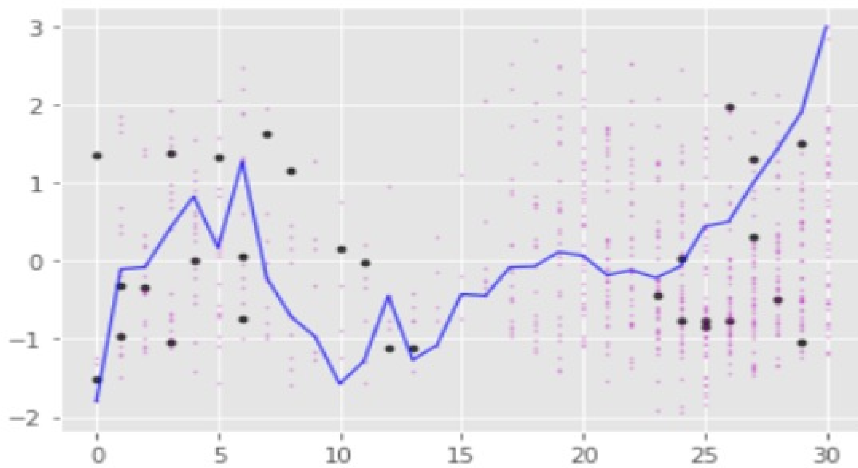


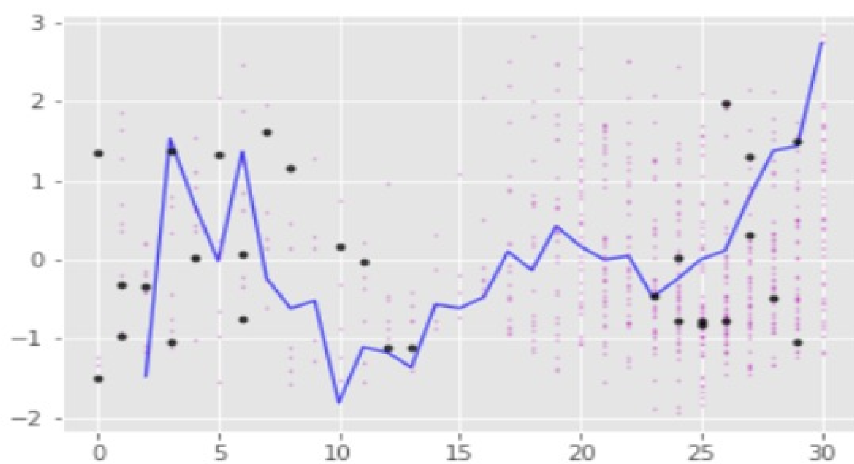


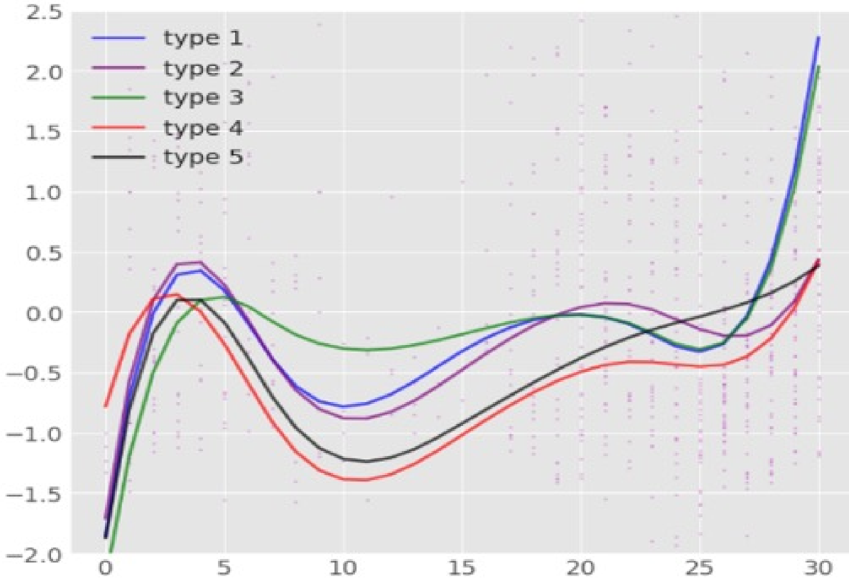

3
2
1
0
−1
−2
0
5
10
15
20
25
30


Engine archetype
5 engine types
included in train set data
new engine types
not included in train set data
age
type 1
type 2
type 3
type 4
type 5
type 6
type 7
type 8
type 9
type 10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31


3
2
1
0
−1
−2
0
5
10
15
20
25
30


3
2
1
0
−1
−2
0
5
10
15
20
25
30


3
2
1
0
−1
−2
0
5
10
15
20
25
30


type 1
type 2
type 3
type 4
type 5
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
5
10
15
20
25
30