Nuclear Engineering and Technology 55 (2023) 2844–2853

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Determining the adjusting bias in reactor pressure vessel embrittlement trend curve using Bayesian multilevel modelling

Gyeong-Geun Lee*, Bong-Sang Lee, Min-Chul Kim, Jong-Min Kim

Materials Safety Technology Research Division, Korea Atomic Energy Research Institute (KAERI), 111, Daedeok-daero, 989 Beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea

ARTICLE INFO

Article history:
Received 25 December 2022
Received in revised form
31 March 2023
Accepted 29 April 2023
Available online 9 May 2023

Keywords:
Reactor pressure vessel
Irradiation embrittlement
Embrittlement trend curve
Bayesian multilevel model
Markov chain Monte Carlo

ABSTRACT

A sophisticated Bayesian multilevel model for estimating group bias was developed to improve the utility of the ASTM E900-15 embrittlement trend curve (ETC) to assess the conditions of nuclear power plants (NPPs). For multilevel model development, the Baseline 22 surveillance dataset was basically classified into groups based on the NPP name, product form, and notch orientation. By including the notch direction in the grouping criteria, the developed model could account for TTS differences among NPP groups with different notch orientations, which have not been considered in previous ETCs. The parameters of the multilevel model and biases of the NPP groups were calculated using the Markov Chain Monte Carlo method. As the number of data points within a group increased, the group bias approached the mean residual, resulting in reduced credible intervals of the mean, and vice versa. Even when the number of surveillance test data points was less than three, the multilevel model could estimate appropriate biases without overfitting. The model also allowed for a quantitative estimate of the changes in the bias and prediction interval that occurred as a result of adding more surveillance test data. The biases estimated through the multilevel model significantly improved the performance of E900-15.

1. Introduction

The integrity of the reactor pressure vessel (RPV) in a nuclear power plant (NPP) is affected by irradiation embrittlement owing to neutrons generated during NPP's operation. Irradiation embrittlement of the RPV material is quantitatively evaluated using a surveillance test program. In the surveillance program, specimens made from the RPV material are stored in 4–5 capsules and subjected to high neutron flux during irradiation at specific locations between the nuclear fuel and the RPV. Over time, the capsules are removed individually to evaluate the mechanical properties of the irradiated specimens. The transition temperature shift (TTS), calculated as the difference in the Charpy impact test results between the unirradiated and irradiated specimens, is a key indicator of the RPV material's susceptibility to irradiation embrittlement.

To accurately predict the TTS of RPV materials in a NPP over the course of its life, it is crucial to develop an embrittlement trend curve (ETC) that considers the initial material properties of the RPV and operating conditions. In the early days of NPP construction, the surveillance test results of RPV materials were insufficient to develop an elaborate model [1–4]. As the operation time of NPPs gradually increased, surveillance test data also increased, and various ETCs were developed through statistical modeling or modeling based on the irradiation mechanism of RPV steels [5–8]. The ASTM E900-15 ETC, announced in 2015 [9], is a statistical nonlinear regression model that incorporates the irradiation damage on the matrix phase and the accumulation of Cu precipitates in RPV steel. It showed superior predictive performance compared to other ETCs. Recently, many elaborate models using machine learning with accumulated data have been introduced [10,11]. However, it is difficult to precisely determine the internal structure of a machine learning model, and extrapolation over time can introduce significant errors. Therefore, nonlinear models are preferred for use as regulatory ETCs.

Despite the excellent predictive performance of the entire range of surveillance test data in current ETCs, many factors can affect radiation embrittlement, and not all of them are included even in a modern ETC, such as E900-15. To increase the applicability of NPP regulation, E900–15 may require two additional considerations. First, more attention is required for the heat-based prediction of TTS. From a regulatory perspective, the integrity of an NPP is fundamentally dependent on the surveillance test results of an individual NPP. Therefore, a guideline for applying ETC to individual NPPs is required, in addition to finding trends for the entire

* Corresponding author.
E-mail address: gglee@kaeri.re.kr (G.-G. Lee).

https://doi.org/10.1016/j.net.2023.04.042

surveillance test data. The second concern is that the unirradiated Charpy transition temperature may continuously affect the subsequent TTS results in NPP materials. The unirradiated Charpy transition temperature at 41J was used as a reference point for the TTS group, which is expressed as follows:

$$TTS^i = T^i_{41J} - T^{unirradiated}_{41J} \left(i = 1, 2, 3, \cdots\right) \quad (1)$$

where i denotes the i-th surveillance test in the groups of an NPP. Even specimens manufactured from the same heat may have variations in the Charpy transition temperature owing to compositional non-uniformity and errors in the measurement process, and these variations arising from unirradiated specimens will affect the successive results of subsequent TTS measurements.

To address these drawbacks, ETC researchers have proposed introducing a bias or slope to the E900-15 global trend according to the amount of surveillance test data in the individual heat of an NPP [12–14]. This approach was simple and effective in bridging the gap between global-trending ETC and plant-specific trends. However, the bias adjustment was insufficient when there were fewer than three surveillance test results in an NPP, and the distribution of biases across the overall RPV remains to be determined.

This study introduced a grouping variable that included the NPP name, heat information with product form and notch orientation in the accumulated surveillance test data to calculate the biases of the groups. This grouping variable allowed the TTS data of individual plants to be grouped together with the same unirradiated Charpy transition temperature. Group biases of the global trend E900-15 in NPPs were estimated with a Bayesian multilevel model [15] using the Markov Chain Monte Carlo (MCMC) method [16], and the distribution of estimated group biases was analyzed quantitatively. The model performance was measured and compared with that of E900-15 ETC. The change in prediction interval according to the addition of the surveillance test data was measured. This paper briefly discusses future research.

2. Methodology

2.1. Dataset

The dataset used in this study was Baseline22, which was embedded in Plotter-22, an ETC evaluation software provided by the ASTM Committee E10.02, in 2022. Plotter-22 is a successor to Plotter-15 [9], which was developed to analyze the ASTM E900-15 database containing a collection of surveillance test data from pressurized water reactors (PWRs) and boiling water reactors worldwide. Baseline22 includes RPV surveillance test results after 2015 and corrects some data from the Plotter-15 dataset.

As mentioned in the introduction, this study aims to group data with the same unirradiated Charpy transition temperature for each NPP and to calculate an appropriate bias that can reduce the error between the global and plant-specific trends. Baseline22 includes much information to group data, but the unirradiated Charpy transition temperatures still need to be reported. However, even if the exact unirradiated Charpy characteristics are unknown, data grouping is possible using NPP information, material product forms, and notch orientation of the Charpy specimens. The operating temperature and irradiation flux are generally similar in the same NPP, and the composition of materials in the NPP is mostly the same, depending on the product form. Therefore, when there is data with the same product form of the same NPP, it becomes a highly related group. Notch orientation was also considered because the Charpy transition temperatures can vary considerably depending on the notch orientation when performing the Charpy impact test.

Baseline22 has a column “material lookup,” a key column that distinguishes the NPP name and product form. We created a new column by combining the material lookup and notch orientation columns. In the case of the notch orientation column, there were records with missing values. In these cases, we imputed it as an unknown. For example, in South Korea's Kori Unit 1, the data can be divided into three groups: KO1−F1-LT, KO1−F1-TL, and KO1−W1-LT. The first key in this grouping represents the unit name, the second key is the product form with a composition number, and the third represents the specimen direction. Generally, product forms are composed of a base metal and a weld metal in a plant, but some plants may have multiple base or weld metals, so a component number is added to distinguish them. Group IDs were created from this column by numbers assigned in ascending order based on the country. Through this process, all the data were assigned to a single group, and all the TTS data in a particular group were assumed to be associated with the same unirradiated transition temperature.

There were also many data points from Standard Reference Material (SRM) specimens irradiated with the same material at various NPPs, containing high-dose irradiation effects. SRM grouping was performed similarly to the process for the other data points, except that the NPP name was replaced with SRM. The total number of data points was 2044, and the number of groups was 677.

For regulatory purposes, validation of the data grouping is required to obtain accurate parameters for each group. However, as this study aims to prove the utility of the multilevel model, the grouped dataset verification is left for future research. Table 1 summarizes the grouped datasets. The range of data provides estimation coverage for the developed multilevel model.

2.2. Modeling

The E900-15 ETC was used to represent the global trend for all the data in the grouped dataset. The residuals were calculated using the difference between the measured and predicted TTS from E900-15. To minimize the between- and within-group residuals, three model parameters—the group bias mean, group bias standard deviation (SD), and within-group SD—were estimated with a Bayesian multilevel model using the MCMC technique [16]. Notably, MCMC is an advanced technique for obtaining the parameters of statistical models and can be briefly summarized as follows: A new parameter sample was created using random numbers, evaluated with the measured data, and accepted if it was an appropriate model parameter. The representative values and distributions of the model parameters can be estimated quantitatively as the sample collection gets large.

In this study, a multilevel model was fitted using *R* [17] and the *brms* package [18]. The fit generated 20,000 samples in approximately 2 min using the latest type of personal computer with eight cores, and the model size was approximately 200 MB. The model parameters were calculated to two decimal places in this study. The results may vary slightly depending on the random seed and the

Table 1
Summary of the Baseline22 dataset.

Variable	Unit	Minimum	Median	Mean	Maximum
TTS	°C	−28.9	33.0	42.0	253.0
Cu	wt%	0.01	0.09	0.11	0.41
Ni	wt%	0.04	0.64	0.62	1.70
Mn	wt%	0.58	1.36	1.33	1.98
P	wt%	0.0015	0.0100	0.0108	0.0240
Temperature	°C	255	286	285	304
Fluence	10^{19} n/cm^2	0.001	1.250	1.988	21.428
Product form	Plate (SRM), forging, weld				

Total datapoints: 2044.
No. of groups: 677.
No. of nations: 15.

calculation platform of the MCMC. A detailed description of the statistical model formula is provided in the results section.

3. Results

3.1. Necessity of group bias

The E900-15 ETC is a model that uses complete pooling of the entire data, ignoring the group information present in the surveillance test data. Complete pooling captures the overall trend of the data and is generally used when the amount of data within an individual group is limited. In the early days of ETC development, complete pooling was necessary because the amount of surveillance test data collected in most plants was small. The model can be expressed by a simple statistical formula as follows:

$$\begin{aligned} y &= f(x,\ \theta) + \varepsilon \\ \varepsilon &\sim Normal(0,\ \sigma) \end{aligned} \tag{2}$$

where y is the response variable, f is a nonlinear function representing the global trend, x is a vector of features or predictor variables, θ is a parameter vector of f, and ε is assumed to be a normal distribution with a mean of zero and SD σ. In this study, y is the TTS, f is the E900-15 model, x is a vector of seven feature variables, and θ is composed of the 26 parameters of E900-15. The seven feature variables are the product form, Cu, Ni, Mn, P, temperature, and irradiation fluence.

When the data are divided into groups, the TTSs for a particular group often found to show a systematic deviation from the predictions of the ETC. This is illustrated in Fig. 1 for the data groups related to Korean PWRs. Fig. 1 shows the change in TTS according to the irradiation amount, and the upper panels of Fig. 1 show the cases with several surveillance test results. The data points in the KR0005 group agreed with the predicted values of the E900-15. However, there were systematic deviations over 1 SD from E900-15 in the KR0001 and KR0002 groups. KR0001 and KR0002 were cases in which only the notch direction of the specimen was different in LT and TL under the same forge material and irradiation conditions of the same NPP. In the case of forgings and plates, there may be significant variations in the Charpy impact properties depending on the specimen orientation. Therefore, the specimen orientation should be indicated in the surveillance test [19,20].

E900-15, indicated by the black line, predicted the same value for both groups because E900-15 did not consider the specimen notch orientation. This result is an example of a contrasting trend based on the notch direction of the specimen. The lower panels of Fig. 1 show the groups that underwent only one surveillance test. Data points significantly different from the predictions of E900-15, such as KR0022, were noted.

The distribution of the data in Fig. 1 suggests that deviations from the predictions by E900-15 can be treated as varying intercepts or biases. The dotted red line in Fig. 1 is the E900-15 prediction curve, which introduces a bias b_i that minimizes the residuals between the measured data points and the E900-15 predictions. The model can be expressed as follows:

$$\begin{aligned} y_{ij} &= f(x_{ij},\ \theta) + b_i + \varepsilon_{ij} \\ b_i &= \text{constant with group } i \\ \varepsilon_{ij} &\sim Normal(0,\ \sigma) \\ i &= 1,\ \cdots,\ M \\ j &= 1,\ \cdots,\ N \end{aligned} \tag{3}$$

where i is the group index, j is the index of the data in the group, M is the total number of groups, N is the number of data points in the group, and b_i is the common bias of group i, which minimizes the residuals (that is, errors between the measured data and the E900-15 predictions in the group). Further, ε_{ij} represents the error of each data. Each group has a group bias b_i, and this case is known as no pooling. It is rational to have a bias b_i for each group, considering the characteristics of TTS. The bias relieves the uncertainty of unirradiated Charpy impact properties and group-specific errors, which can increase the goodness of fit of E900-15.

Fig. 2 shows the bias b_i for each group in the Korean surveillance test data. The gray points are the residuals between the measured values and the E900-15 predictions, and the red point is the b_i per group that minimizes the residuals within the group, which is equal to the intercept value of the red line in Fig. 1. There were many groups whose predicted values of E900-15 were smaller than the measured data, implying that E900-15 might underestimate and produce non-conservative predictions for these groups.

Calculating the bias independently for each group presents both advantages and disadvantages. When sufficient data points, such as three or more, are available for each group, group bias can be easily estimated by minimizing the residuals between global trends and measurements within the group. However, when only one or two surveillance test results are available, it can be challenging to accurately determine whether group bias is based on data within the group or is a result of TTS measurement errors. The inability to calculate reliable bias until several surveillance tests are performed is a significant disadvantage. Another disadvantage of no pooling of data is that b_i calculated independently for each group does not affect the calculation of the bias of other groups. A model that adequately represents the distribution of bias across all groups is required.

3.2. Multilevel model with partial pooling of data

Partial pooling is an effective method for reconciling the global trends obtained through complete pooling with plant-specific trends derived from group data. This approach incorporated a weighting system based on the quantity of data available for each group. Specifically, when the amount of data in a group is sufficient, the final estimate becomes closer to the no pooling estimate. Conversely, when data are limited, complete pooling estimates constitute a significant proportion of the final estimate [15].

A multilevel model is based on the partial pooling of data and calculates the errors of the model by dividing them into fixed and random effects. Here, the fixed effect represents the tendency of all the data, and the random effect is the tendency of each group. Assuming that the group bias b_i is a normal distribution, the multilevel model can be expressed by the following statistical model formula:

$$\begin{aligned} y_{ij} &= f(x_{ij},\ \theta) + b_i + \varepsilon_{ij} \\ b_i &\sim Normal(b_0,\ \sigma_b) \\ \varepsilon_{ij} &\sim Normal(0,\ \sigma) \\ i &= 1, \cdots,\ M \\ j &= 1, \cdots,\ N \end{aligned} \tag{4}$$

Note that b_i is a normal distribution with mean b_0 and SD σ_b. Two approaches have been proposed to estimate the model parameters and the group bias b_i. The first is the traditional method, where parameters are obtained through an optimization technique such as maximum likelihood or restricted maximum likelihood. This method, generally known as the frequentist approach, can produce results in a relatively short amount of time. However, it requires a complicated procedure to calculate the statistical interval of group b_i, and the interpreting each statistical interval can be complicated. Additionally, fitting a complex multilevel model may result in difficulties in converging the parameter values.

The second method is the Bayesian regression method, which

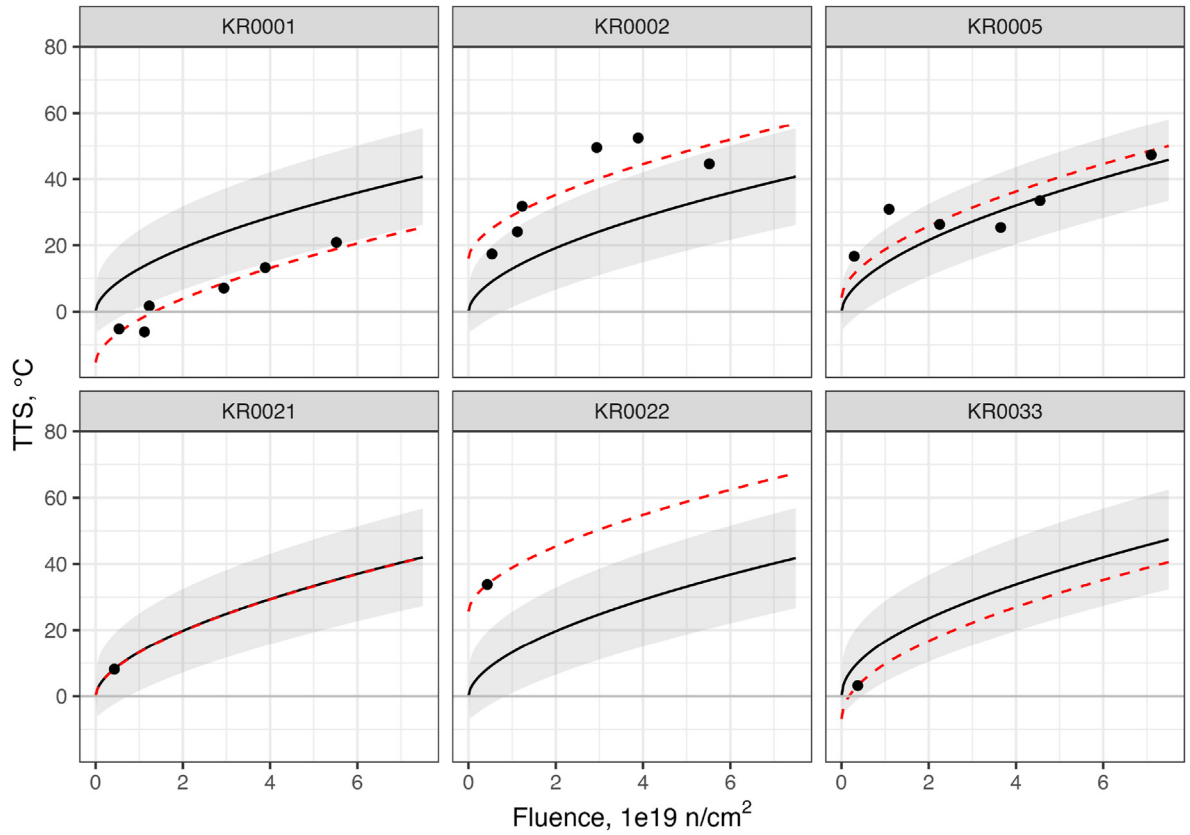

Fig. 1. TTS change with fluence in the selected data groups from Korean surveillance data. The black dots are the measured data. The black lines are the predicted value of E900-15, and the gray sections are the SD of E900-15. The red dashed lines are the trend line moved by the bias that minimizes the residuals.

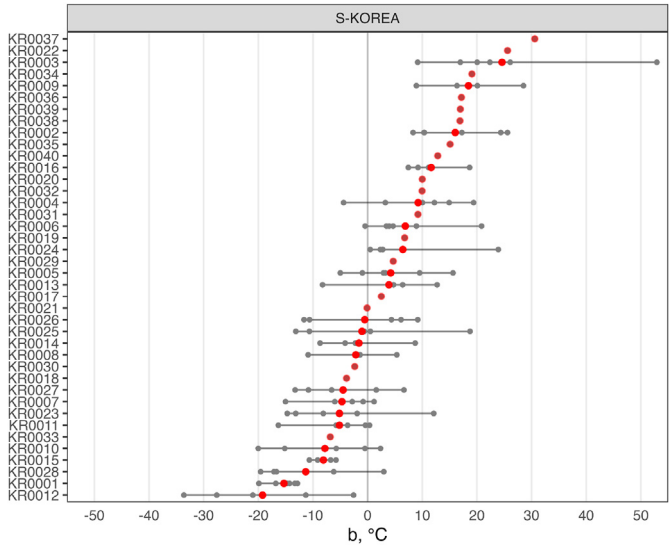

Fig. 2. Residuals and group bias in the Korean data groups. The black points represent the measured TTS, and the red points represent the no pooling bias that minimizes a group's residuals.

employs the MCMC method. The difference between the two approaches is based on a statistically complex explanation [15]. However, from the perspective of statistical applications, Bayesian regression can provide a quantitative distribution of model parameters and estimate the uncertainty of the prediction interval. In addition, it has the flexibility to set the prior probability distribution of the parameters flat to obtain results similar to the frequentist approach or adjust the prior distribution of parameters to increase the convergence and reliability of the model. For example, a reasonable range for a parameter to exist can be assumed to be a normal distribution with a mean and SD, in which case the parameter has a prior probability. If the SD is very large relative to the mean, this is close to a flat prior with little prior information. Despite these advantages, this approach requires significant amount of time and computing resources. In this study, to improve the applicability of various models for future research, we used the *brms* package [18] for MCMC. The *brms* package uses a flat prior as the default value, producing results similar to those of the frequentist approach.

Fig. 3 shows the three primary parameters of the fitted multilevel model. All three parameters have a shape similar to that of a normal distribution. Here, the round black dots are mean values, thick bars are intervals with a credible interval width of 66%, and thin bars are intervals with a credible interval width of 95%. Assuming a normal distribution, they are similar to the 1 SD and 2 SD intervals. The mean value of b_0 was 0.62, which is close to zero. This value almost satisfies the assumption that E900-15 is a global trend. The estimated σ_b for group b_i was 9.45, summarizing the distribution of bias among the 677 groups. The σ value, which represents the error between actual data points and the E900-15 shifted in parallel with the group biases, was 9.55. This σ was narrower than σ_b. Table 2 summarizes the representative parameters of the multilevel model.

Fig. 4 compares the prediction accuracy between E900-15 and E900-15 with group bias. The R-squared and root mean SD (RMSD) values were significantly improved by introducing group b_i. In

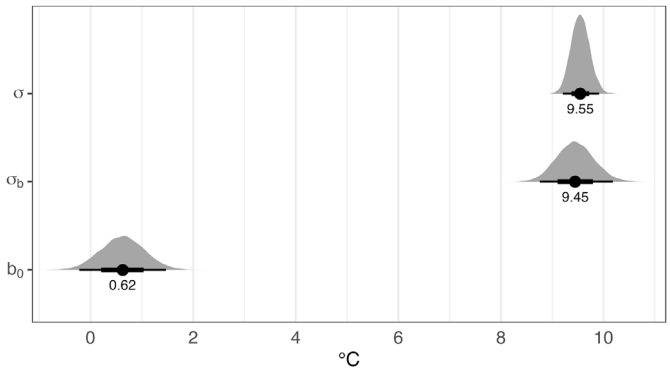

Fig. 3. Distribution and mean values of the multilevel model parameters.

particular, when the measured TTS was negative, the E900-15 model did not contain negative values, resulting in significant errors. However, in the case of E900-15 with group bias, group b_i could have a negative value; therefore, the restriction was significantly relaxed. Fig. 5 shows a graph depicting the distribution of the residuals according to the predicted values. The closer the residual slope is to zero, the better the model. The slope of E900-15 was calculated to be −0.018 because of the application of the Baseline22 data, which contains 195 additional data points compared to the Baseline15 data used to develop E900-15. This variation is thought to be influenced by Ni, but an evaluation using bootstrap methods [21] showed that modifications to the E900-15 model were not necessary [22]. The slope of E900-15 with group bias was −0.028, slightly lower than that of E900-15, but the addition of bias resulted in only the scatter of the residual decrease, and the slope did not change significantly. This confirmed the adequate performance of the multilevel model.

Fig. 6 represents the group b_i in the Korean surveillance test data. Here, the red points are the no pooling b_i calculated for each group in Fig. 2, and the blue points are the means of the group b_i calculated through the multilevel model. Above the blue dots, the histograms show the distribution of the group b_i. In general, b_i of the multilevel model moves closer to the overall average than b_i calculated by no pooling. This phenomenon is known as *shrinkage* [15]. The extent of shrinkage depended on the number of data points in each group. A group with more than four data points approaches no pooling b_i, and when there are fewer data points, it approaches the overall average. Moreover, the amount of data affects the width of the b_i distribution. When there is only one datum in a group, the position of b_i can exist over a wide range. The location of b_i became more reliable with increasing data points. In the frequentist approach, it is not easy to calculate the distribution of group b_i, which is an example of one of the advantages of MCMC.

Fig. 7 shows the change in the SD of the distribution of group b_i according to the number of data points. A slight jitter was added to the horizontal axis of the data points to express the dispersion of SD change. As the amount of data in the group increase, the SD of b_i show a decreasing trend. The maximum number of data points for a typical surveillance test is usually six, and more is the case for SRM data. The width of the bias distribution is significantly reduced because of the large number of SRM data points. The credible interval of b_i gradually decreases as the number of data points increases. This decrease was directly related to the prediction interval for future surveillance tests. The estimation of the prediction interval with increasing data points is described in the next section.

To easily visualize the trend of Figs. 6 and 8 shows an example of the addition of group b_i. The solid black line represents the predicted value of E900-15 and the dotted red line denotes the trend of the no pooling b_i. The dotted blue line represents the trend in Group b_i of the multilevel model. As shown in the upper panel of Fig. 8, no pooling and partial pooling result in similar biases when

Table 2
Fitted parameters of the multilevel model.

Parameter	Estimate	Est.Error	95% Credible Interval Lower	95% Credible Interval Upper
b_0	0.62	0.43	−0.22	1.47
σ_b	9.45	0.36	8.76	10.18
σ	9.55	0.18	9.21	9.91

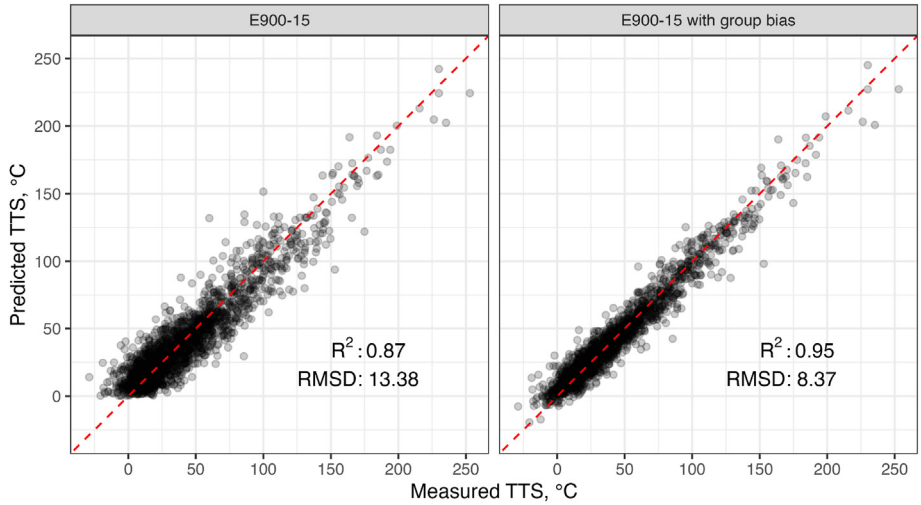

Fig. 4. Measured vs predicted TTS plots of E900-15 and E900-15 with group bias.

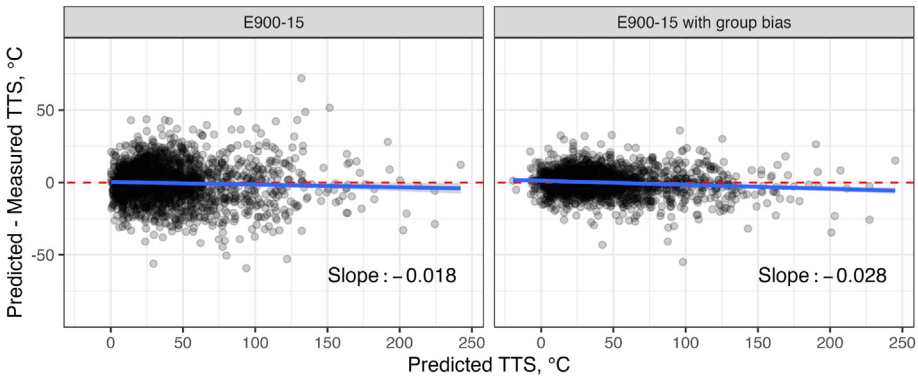

Fig. 5. Predicted TTS vs residuals plots of E900-15 and E900-15 with group bias.

the number of data points in a group is large. The lower panel of Fig. 8 shows only one data point, and the dotted blue line is closer to the solid black line. The multilevel model provides a group trend that is close to the overall trend when the number of data points is small. This implies that the multilevel model reduces overfitting of the no pooling model.

Since a method for considering bias in the global trend based on the surveillance test data of each plant was reported [12], it has been described in detail in recent technical reports [13,14]. These reports propose a procedure for calculating bias to reduce errors in the global trend in groups with more than three surveillance tests, and this adjustment can significantly reduce the SD of E900-15. The multilevel model of this study further considers the explicit inclusion of notch orientation in the group for bias calculation. In this case, errors due to differences in notch orientation can be included in the variation of bias, but the number of groups to be analyzed increased, and the data within the groups decreased. The decrease in data within groups makes it easier to obtain overfitted results rather than appropriate biases. However, by introducing a multilevel model, it was possible to calculate appropriate biases even when the number of surveillance test data points was less than three.

Ortner et al. reported that variations in the mechanical properties of RPV materials across different countries can result in disparities in irradiation embrittlement [23]. To confirm this finding, the distribution of bias and shrinkage tendency of the groups in various countries were analyzed (Fig. 9). The blue dots are group b_i of the multilevel model and the red dots are group b_i with no pooling. The group biases of no pooling were overestimated or underestimated depending on the country. Japan showed overestimated tendencies for E900-15, whereas Germany presented underestimated tendencies. Group b_i was appropriately moderated in all countries using a multilevel model. Calculating the mean bias of the country may be possible through additional modeling. However, the small number of data groups in many countries makes it difficult to obtain reliable parameters for a multilevel model for each country. This result emphasizes the requirement for joint collection and in-depth analysis of country-specific surveillance data.

3.3. Calculation of prediction intervals

A multilevel model can easily estimate the prediction interval of the model based on the addition of new data. This prediction interval simultaneously included between- and within-group errors. For example, we removed KR0005 group data from the Baseline22 dataset and created a multilevel model. We then added a new surveillance data point and updated the model serially. In a surveillance test, data points are not added simultaneously, but the number of data points slowly increases over time. The experimental model reproduced this situation. Fig. 10 shows the bias and prediction intervals based on the model update. The solid black line is the predicted trend of E900-15, and the gray shading indicates the range of one SD of E900-15. The dotted blue line and shading denote trends predicted by the multilevel model.

The group bias was estimated as 0.6 when there was no surveillance test, n = 0. This value was similar to the average value of b_0, 0.62. Because the MCMC depends on a random number, it is slightly different. The SD of the prediction interval was ~13.5, which is a value that considers both between-group σ_b and within-group σ. The prediction interval of the multilevel model was similar to that SD of E900-15, except for the initial irradiation period. The SD of E900-15 increased gradually according to the dose [9], but the

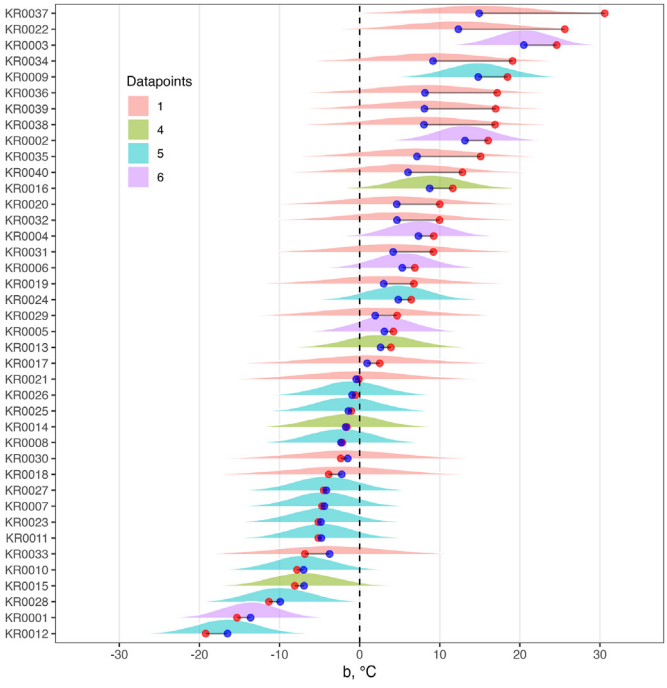

Fig. 6. Distribution and mean value of the bias from the multilevel model in grouped Korean surveillance data. Blue points represent the biases from the multilevel model. Red points represent the no pooling biases.

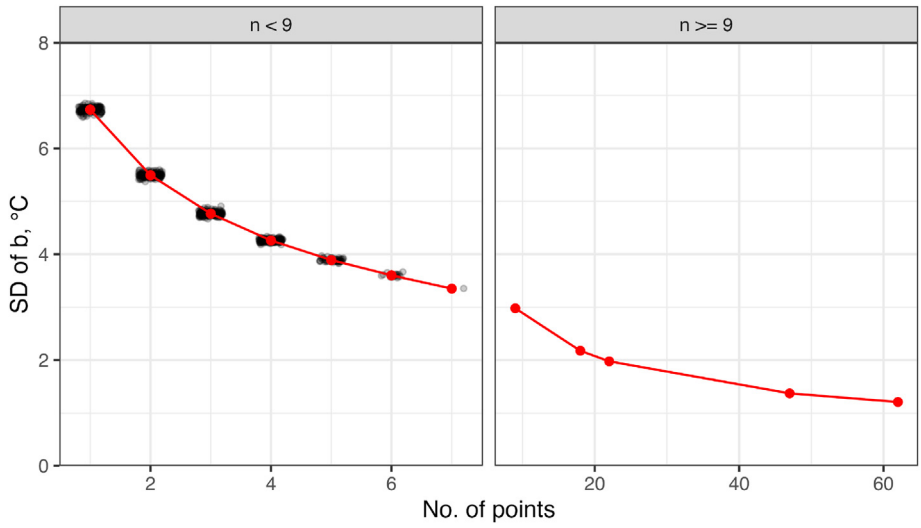

Fig. 7. Change in SD of the distribution of group b_i according to the number of data points.

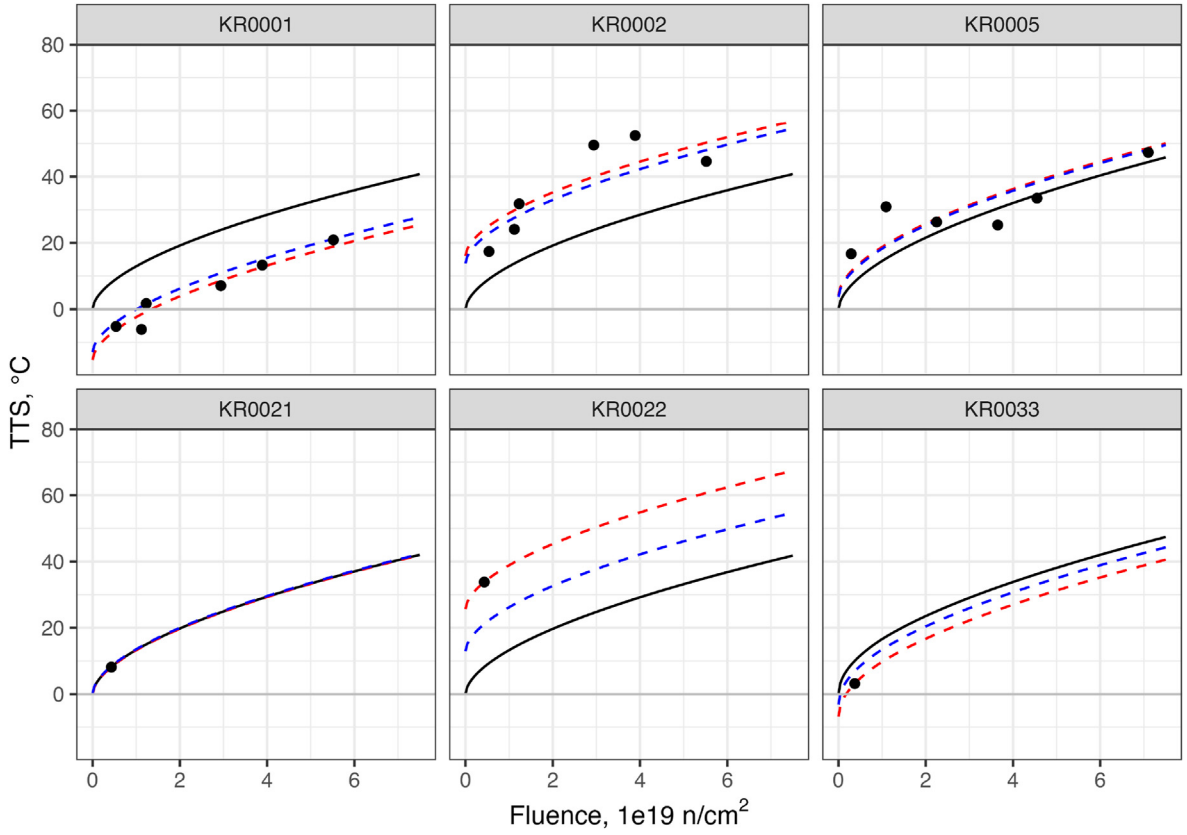

Fig. 8. TTS changes of the multilevel and no pooling models according to irradiation fluence in the selected data groups from Korean surveillance data. The blue line is from the multilevel model, and the red line is from the no pooling model.

multilevel model had the same prediction interval over the entire irradiation range. Thus, the prediction interval of the multilevel model was wider during the early stages of neutron irradiation.

After conducting the surveillance test and adding one data point, the updated model's first result in depicted in the n = 1 panel in Fig. 10. The group b_i increased to 5.2 based on the added data, and the prediction interval decreased to 11.7. Panel n = 2 shows that an additional datum was added. The large residuals of the new data points were reflected in the multilevel model, resulting in increased group bias. However, the width of the prediction interval continued to decrease with the addition of data. When n = 3, the group bias decreased because the newly measured value was close to E900-15. In the case of n = 4, the measured value was lower than the predicted value of E900-15, and the bias approached zero. Group bias

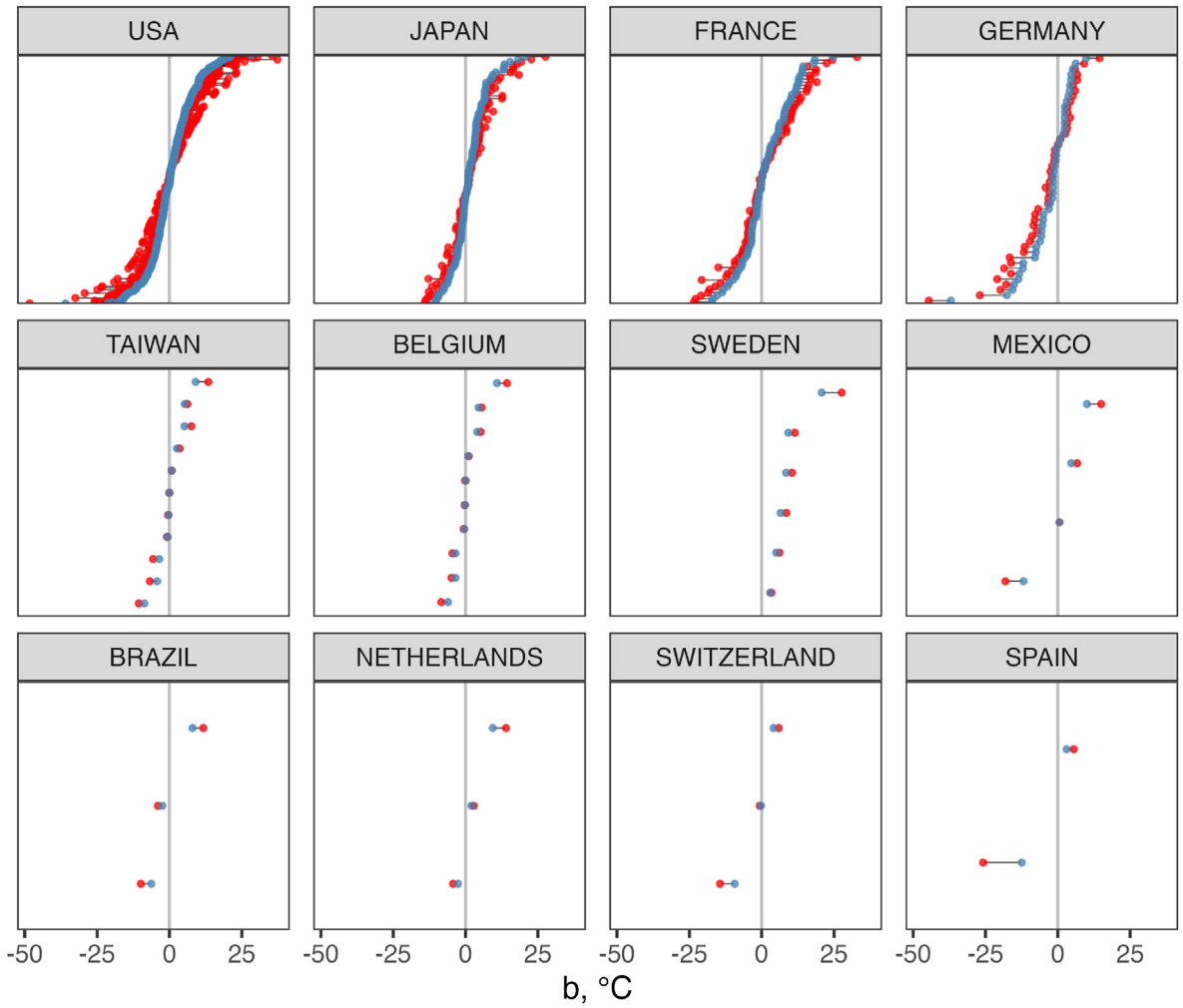

Fig. 9. Differences between the bias from the no pooling and multilevel models from the grouped data in various countries. Blue points represent the biases from the multilevel model. Red points represent the no pooling biases.

was adequately determined based on the addition of the surveillance test data.

Note that, even with the addition of just one data point, the multilevel model suggests a reduced SD compared with E900-15. This is related to the SD of the credible interval of the group bias with the addition of data (Fig. 7). With one data point, the group bias SD was approximately 6.8 (Fig. 7), and the within-group SD was 9.55; therefore, combining the two SDs yielded 11.7, which is the square root of the sum of the variances. The prediction interval continued to decrease as data were added, but it did not fall below 10. This is because the within-group SD σ was 9.55, and the prediction interval could not contract below this value, even if the number of data points increased. As demonstrated in this example, the multilevel model can quantify the group bias and prediction interval according to the addition of data, and it can be appropriately applied to evaluate plant units for regulatory purposes.

3.4. Improvement of the multilevel model

The current MCMC model considers only the bias change in E900-15. Some factors affecting embrittlement may not best be described as a simple bias. A more generalized model was formulated to incorporate the slope term, as described below:

$$y_{ij} = slope \times f(x_{ij}, \theta) + b_i + \varepsilon_{ij} \quad (5)$$

The preliminary modeling with the slope parameter estimated that the slope and b_0 attain a value of 1.04 and −0.73, respectively. Because E900-15 was a global trend, the slope and bias were close to one and zero, respectively. Adding the slope variable to the global trend E900-15 had little effect on the model performance, and the slope change according to the three product forms (forgings, plates, and welds) also did not significantly increase the performance.

To consider the random effects in the model, we can simultaneously estimate the bias and slope in a group. However, when

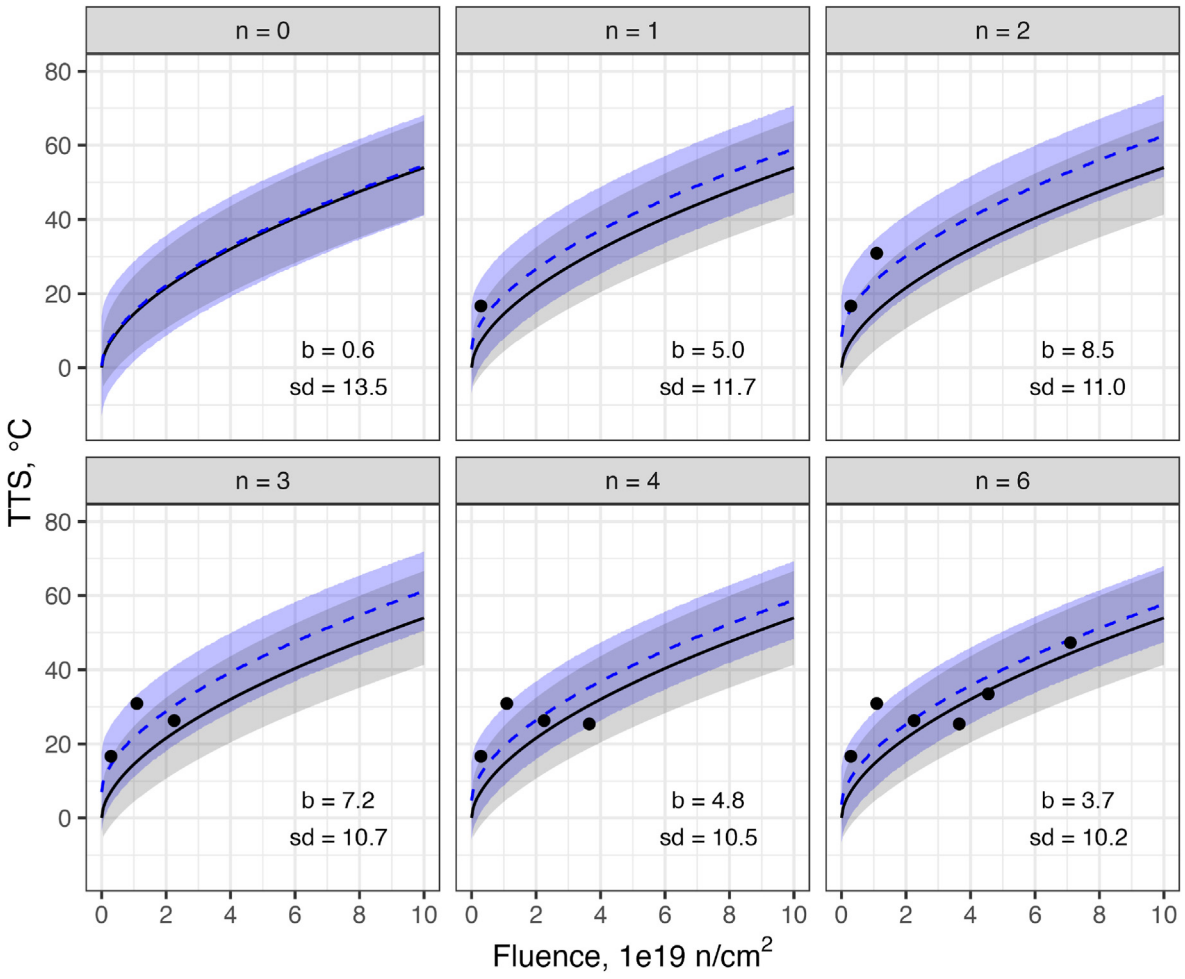

Fig. 10. Evolution of the bias and prediction interval of the multilevel model with increasing data points and refitting of the multilevel model.

there is only one data point in a group, it is difficult to calculate the two unknown parameters (bias and slope) with one data point. In this case, a flat prior distribution for the slope cannot be applied. An adequate prior distribution for the slope should be introduced, and further studies will address the details related to this topic.

4. Summary

In this study, a Bayesian multilevel model with group bias was developed to increase the applicability of the global trend E900-15 ETC to NPPs. The surveillance dataset Baseline22 was divided into 677 groups to group the data by combining plant name, product form, and notch orientation. By adding the notch direction to the grouping criterion, it was possible to calculate the bias of groups of the same material with different unirradiated Charpy transition temperatures in a plant. A multilevel model was formulated by dividing the error of E900-15 into between- and within-group errors, assuming that the group bias was normally distributed. The distribution of model parameters and the bias of each group were estimated using the MCMC method. With an increase in data points within a group, the group bias approached the mean residual of the data points in each group, and the width of the credible interval of the mean decreased. Conversely, a smaller number of data points within a group resulted in a group bias value closer to the overall trend and a wide credible interval of the mean. Changes in the bias and prediction interval due to the addition of surveillance test data could also be quantitatively estimated using a multilevel model. Notably, an appropriate bias without overfitting could be calculated even when the number of surveillance test data points was less than three. The biases estimated from the multilevel model could significantly improve the performance of the E900-15 ETC. Advanced modeling, such as considering group slope changes, is required to enhance the performance of the model further.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the ASTM E10.02 committee for providing the dataset for this study. This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2022-00144399).

References

[1] U.S. NRC, Regulatory Guide 1.99 Revision 1: Effects of Residual Elements on Predicted Radiation Damage to Reactor Vessels Materials, U.S. Nuclear Regulatory Commission, Washington DC, 1977.

[2] J.F. Perrin, R.A. Wullaert, G.R. Odette, P.M. Lombrozo, Physically based regression correlations of embrittlement data from reactor pressure vessel surveillance programs. https://www.osti.gov/biblio/5336511, 1984. Final report.
[3] G.L. Guthrie, Charpy trend-curve development based on PWR surveillance data, Nucl. Eng. Des. 86 (1985) 79–86, https://doi.org/10.1016/0029-5493(85)90211-0.
[4] U.S. NRC, Regulatory Guide 1.99 Revision 2: Radiation Embrittlement of Reactor Vessel Materials, U.S. Nuclear Regulatory Commission, Washington DC, 1988.
[5] E.D. Eason, J.E. Wright, G.R. Odette, NUREG/CR-6551: Improved Embrittlement Correlations for Reactor Pressure Vessel Steels, U.S. Nuclear Regulatory Commission, Washington DC, 1998.
[6] E.D. Eason, J.E. Wright, G.R. Odette, NP-3319: Physically Based Regression Correlations of Embrittlement Data from Reactor Pressure Vessel Surveillance Programs, Electric Power Research Institute, 1998.
[7] JEAC, JEAC4201-2007: Method of Surveillance Tests for Structural Materials of Nuclear Reactors, Japan Electric Association, Chiyoda-ku, Tokyo, Japan, 2007.
[8] M. Kirk, A wide-range embrittlement trend curve for western reactor pressure vessel steels, in: Effects of Radiation on Nuclear Materials: 25th Volume, ASTM International, 2013, pp. 20–51, https://doi.org/10.1520/STP103999.
[9] E10 Committee, ASTM E900-15e2: guide for Predicting Radiation-Induced Transition Temperature Shift in Reactor Vessel Materials. https://doi.org/10.1520/E0900-15E02, 2021. ASTM standards.
[10] G.-G. Lee, M.-C. Kim, B.-S. Lee, Machine learning modeling of irradiation embrittlement in low alloy steel of nuclear power plants, Nucl. Eng. Technol. 53 (2021) 4022–4032, https://doi.org/10.1016/j.net.2021.06.014.
[11] D. Ferreño, M. Serrano, M. Kirk, J.A. Sainz-Aja, Prediction of the transition-temperature shift using machine learning algorithms and the plotter database, Metals 12 (2022) 186, https://doi.org/10.3390/met12020186.
[12] N. Soneda, K. Nakashima, K. Nishida, K. Dohi, High Fluence Surveillance Data and Recalibration of RPV Embrittlement Correlation Method in Japan, American Society of Mechanical Engineers Digital Collection, 2014, https://doi.org/10.1115/PVP2013-98076.
[13] PWROG-20025-NP, Methodology for Adjusting Reactor Pressure Vessel Embrittlement Predictions Using Heat-specific Surveillance Data, PWROG, 2020.
[14] E. Long, Methods to Address the Effects of Irradiation Embrittlement in Section XI of the ASME Code, 2021. MRP-462).
[15] R. McElreath, Statistical Rethinking: A Bayesian Course with Examples in R and Stan, CRC Press/Taylor & Francis Group, Boca Raton, 2016.
[16] W.R. Gilks, S. Richardson, David Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman; Hall/CRC, 1995, https://doi.org/10.1201/b14835.
[17] R Core Team, R, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. https://www.R-project.org/.
[18] P.-C. Bürkner, Brms : an *R* package for bayesian multilevel models using *stan*, J. Stat. Software 80 (2017), https://doi.org/10.18637/jss.v080.i01.
[19] E28 Committee, ASTM E23-18: test methods for notched bar impact testing of metallic materials, ASTM International, 2018, https://doi.org/10.1520/E0023-18.
[20] E08 Committee, ASTM e399-20a: test method for linear-elastic plane-strain fracture toughness of metallic materials. https://doi.org/10.1520/E0399-20A, 2020. ASTM International.
[21] D. Ferreño, M. Kirk, M. Serrano, J.A. Sainz-Aja, Assessment of the generalization ability of the ASTM e900-15 embrittlement trend curve by means of Monte Carlo cross-validation, Metals 12 (2022) 481, https://doi.org/10.3390/met12030481.
[22] M. Kirk, D. Ferreño, jose A. Sainz-Aja, Evaluation of the ASTM e900-15 ΔT41J prediction equation in light of new data, April 19-22, in: ASTM Symposium – Radiation Embrittlement Trend Curves/Equations and Their Use for RPV Integrity Evaluations, Prague, Czech Republic, 2022.
[23] S. Ortner, M. Brumovsky, Age 60+ — applicability of ageing related data bases and methodologies for ensuring safe operation of LWR beyond 60 years, V06AT06A075, in: Volume 6A: Materials and Fabrication, American Society of Mechanical Engineers, Waikoloa, Hawaii, USA, 2017, https://doi.org/10.1115/PVP2017-65283.