Testing for statistically significant differences in predictions obtained from competing creep models

M. Evans

Institute of Structural Materials, Swansea University Bay Campus, SA1 8EN, Swansea, Wales, m.evans@swansea.ac.uk.

ABSTRACT

It is important to be able to predict the creep life and other creep properties of materials used in power plants and aeroengines. Whilst existing studies have compared different creep models using different measures of predictive performance, none have identified whether these differences are real from a statistical significance perspective. This paper proposes several tests based on a review of the literature. These tests were applied to two creep models using failure time data on 2.25Cr-1Mo steel to develop a recommended approach for practitioners to adopt when selecting a creep model. All such tests concluded that the Evans model produced better long-term life predictions using only short-term data, and that this difference was statistically significant at the 5% significance level.

Keywords: creep, parametric tests, non-parametric tests, mean percentage squared and absolute errors, statistical significance

Introduction

It is important to be able to predict the creep life and other creep properties for materials used in power plants and aeroengines. When this can be done with a high degree of confidence, the results can potentially be used to justify the continued use of aging power plants beyond their original design lives - as a short-term solution to potential energy gaps for example. 2.25Cr-1Mo steel is a main stay alloy used for structural components operating at high temperature within such aging power plants - where the usual service conditions for heater tubes is around 840 K and 35 MPa.

There are many creep models available in the literature for carrying out such extrapolations and a good review can be found in Evans [1] and Holdsworth [2]. Such models generally fall into two broad categories – those that aim at predicting points on a creep curve (such as the time to failure or the minimum creep rate) and those that aim to predict the position and shape of the whole creep curve as a function of test conditions. In turn, there has been a recent subdivision of the former models into those that are parametric in nature and those that are semi or non-parametric in nature [3]. Many papers also exist in the literature that perform comparisons of these models. For example, Holdsworth [5] compared several models using 2.25Cr1Mo, 9CrMoVNb and 18Cr13NiMo steels, whilst Abdallah et. al. [6], Bueno and Sobrinho [7] have concentrated on 2.25Cr-1Mo.

In all these studies, differences in the predictive performance of different creep models were observed. However, in none of these studies, (and as far as the author is aware in any other published paper), was there an attempt to assess whether these differences were statistically significant - rather than simply occurring due to chance. Clearly, this is important for determining the selection of the most appropriate creep model for long term life predictions on a particular material. The objectives of this paper are therefore to review the literature to present a range of statistical tests that can be used for this purpose by future researchers. The paper also aims to illustrate how these tests should be applied using 2.25Cr-1Mo steels and two competing creep models as a test bed. It is not to identify a which of these creep models works best for any other material, or different batches of this material. That would require an extensive analysis of data on other materials, To meet these objectives the paper is structured as follows. The method reviews two performance measures and several statistical tests that are present in the literature. This section also describes two parametric creep models and how their unknown parameters can be estimated. In the results section these tests are applied to the predictions made by these creep models using data on 2.25Cr-1Mo. Conclusions are then drawn.

Method

Statistical tests

Evans [9,10] has suggested the use of two statistics for comparing the accuracy of competing creep models. Both are based on the approximate percentage prediction error, with one squaring this error and the other using the absolute values for these errors. Both these approaches prevent under and over predictions offsetting each other in the averaging process

$$\text{MPSE} \approx \frac{1}{n}\sum_{i=1}^{n}[A_i - P_i]^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \quad (1)$$

$$\text{MPAE} \approx \frac{1}{n}\sum_{i=1}^{n}|A_i - P_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (2)$$

where MPSE stands for mean percentage squared error and MPAE stands for mean percentage absolute error. $A_i = \ln(t_F)_i$ is the natural log of the time at which the ith test specimen fails (of which there are n such failure times in a data set) and P_i is a creep model's prediction of A_i. Note that in what follows A_i could equally stand for any other creep property such as the minimum creep rate, time to a specified strain or even strain itself. The reason for working with natural logs is that e_i is then approximately equal to the percentage difference between the actual time to failure and a model's prediction of it. This approximation is very good for percentages of around 30% or less (and obviously the smaller the better). In this paper e_i will be referred to as the percentage prediction error. Evans [9,10] has also shown how both these statistics can be decomposed into random and systematic errors.

The use of the MPSE has several limitations, however. The first is that an absolute-value-based measure, such as the mean percentage absolute error (MPAE), is much more interpretable. Whilst taking the square root of the MPSE (= RMPSE) helps with interpretation by converting the MPSE into the same units as e, this still can give very misleading assessments of a creep models adequacy. This is because the percentage prediction errors are squared, and this makes the MPSE and the RMPSE very sensitive to the presence of any outliers present in the data, i.e. very poor predictions. Thus, there is the potential for the MPSE to underestimate the predictive performance of a given creep model. This is not true of absolute percentage errors which are more robust to the presence of such outliers. But when it comes to developing tests of statistical significance, the squaring of errors enables several well know distributions to be used for determining statistical significance (a difference can be statistically significant no matter how small the difference).

Parametric based tests

When selecting an appropriate creep model, the obvious approach is to select the creep model that has the smaller MPSE or MPAE. But then there is a need to go one step further and determine whether this difference is significant for predictive purposes or simply due to the specific sample of data collected. Let e_{1i} be the percentage prediction errors obtained using creep model 1 and e_{2i} the percentage prediction errors obtained using creep model 2. The Diebold-Mariano [11] test (the DM test) then involves creating a new series called the differential loss d_i which can be defined in different ways. But for this paper, d_i will be defined in one of the following two ways

$$d_i = (e_{1i})^2 - (e_{2i})^2 \quad \text{or} \quad d_i = |e_{1i}| - |e_{2i}| \tag{3}$$

The Diebold-Mariano (DM) test statistic is then defined as

$$DM = \frac{\bar{d}}{s_{\bar{d}}} \tag{4a}$$

where

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad s_{\bar{d}}^2 = \frac{\sum_{i=1}^{n}[d_i - \bar{d}]^2}{n(n-1)} \tag{4b}$$

Under the null hypothesis that the population mean value for d is zero, (i.e. $\mu = E[d_i] = 0$, where $E[d_i]$ reads the expected value for d), DM has an asymptotic standard normal distribution: $DM \sim N(0, 1)$. Thus, there is a statistically significant difference between the predictions obtained from two creep models if $|DM| > Z_{crit}$, where Z_{crit} is the two-tailed critical

value from the standard normal distribution. Z_{crit} depends of course on the researchers chosen level of statistical significance (typically either 1%, 5% or 10%).

Diebold and Mariano provided some simulation evidence to suggest that their test performance was versatile and satisfactory relative to any other test in moderately large samples over a wide range of situations, including heavy-tailed as well as normal prediction error distributions. However, they did find the test tends to reject the null hypothesis too often in small to moderately sized samples (i.e., it is oversized or has lower power). Harvey, Leybourne, and Newbold [12] found that the following modified DM test (or the HLN test)

$$\text{HLN} = \sqrt{\frac{n}{n-1}}DM \tag{5}$$

performed better in all scenarios that they looked at. They also found that this modified test was not as over-sized, and that the performance of the test would be acceptable to practitioners. Under the null hypothesis that the population mean value for d is zero ($\mu = E[d_i] = 0$), HLM follows a Student t distribution with n - 1 degrees of freedom: HLN ~ T(n-1). Thus, there is a significant difference between the model's predictions if $|HLN| > T_{crit}$ where T_{crit} is the two-tailed critical value from the Student t distribution. The standard assumptions required for the above two tests to be valid are that the d_i are normally distributed and that the variance for d is constant.

Tests that were present in the literature prior to the publication of the DM test, include those proposed by Ashley, Granger and Schmalensee [13] (the AGS test) and that by Morgan-Granger-Newbold [14,15] (the MGN test). However, these are tests for equality of the MPSE's between creep models only. The foundation of the AGS test lies in the decomposition of the mean percentage squared error (MPSE) based on the work of Granger and Newbold [16]. They showed that

$$\text{MPSE} = (\bar{A} - \bar{P})^2 + s_e^2 \tag{6a}$$

where s_e^2 is the variance in the percentage prediction errors

$$s_e^2 = \frac{\sum_{i=1}^{n}[e_i - \bar{e}]^2}{n} \tag{6b}$$

If $MPSE_1$ is the mean percentage squared error associated with the n predicted creep properties obtained from creep model 1 and if $MPSE_2$ is the mean percentage squared error associated with the n predicted creep properties obtained from creep model 2, then from Equations (6a)

$$MPSE_1 - MPSE_2 = (\bar{e}_1^2 - \bar{e}_2^2) + (s_{e1}^2 - s_{e2}^2) \tag{6c}$$

where $\bar{e}_1$ and $\bar{e}_2$ are the average percentage prediction errors from creep models 1 and 2 respectively, and s_{e1}^2, s_{e2}^2 their variances. These variances are obtained in the same way as in Equation (6b) – with e_1 and e_2 replacing e. Equation (6c) can be rewritten using $s_i = e_{1i} + e_{2i}$ and $g_i = e_{1i} - e_{2i}$

$$MSPE_1 - MSPE_2 = (\bar{e}_1^2 - \bar{e}_2^2) + s_{sg} \tag{7a}$$

where s_{sg} is the covariance between s and g

$$s_{sg} = \frac{\sum_{i=1}^{n}[s_i - \bar{s}][g_i - \bar{g}]}{n} \quad (7b)$$

Equation (7a) comes about because $(s_{e1}^2 - s_{e2}^2) = s_{sg}$. Now, if model predictions P_2, are more accurate than model predictions P_1, then $MPSE_2$ will be smaller than $MPSE_1$ so that the terms on the right-hand side of Equation (6a) must in combination be positive. Therefore, a test of whether the $MPSE_2$ is smaller than $MPSE_1$, can be based on whether the combination of the covariance and the difference between the squares of the mean errors is positive. This test can be carried out by first estimating the parameters of the regression equation using least squares

$$g_i = \lambda_1 + \lambda_2(s_i - \bar{s}) + u_i \quad (8a)$$

where u is an error term with mean zero, and $\bar{s}$ is the average of all the s_i values. The least squares estimators of the parameters λ_1 and λ_2 are

$$\hat{\lambda}_1 = \bar{g} \quad \text{and} \quad \hat{\lambda}_2 = \frac{s_{sg}}{(s_{s-\bar{s}})^2} \quad (8b)$$

where $s_{s-\bar{s}}$ is the standard deviation in the mean adjusted values for s_i. From Equations (8a,b), the hypothesis that there is no difference in the MPSE's implies that λ_1 and λ_2 will be zero, which can be tested via the null hypothesis, H_o: $\lambda_1 = \lambda_2 = 0$. The alternative hypothesis, that the predictions P_2 have a smaller MPSE, implies that either or both λ_1 and λ_2 will be positive. Should the estimated coefficients in Equation (8a) be significantly negative, the forecasts P_2 are not more accurate than P_1. The test that λ_1 and λ_2 is jointly zero is based on the F distribution

$$\text{AGS} = \frac{(\sum g_i^2 - \sum \hat{u}_i^2)/2}{\sum \hat{u}_i^2/(n-2)} \quad (8c)$$

where $\hat{u}_i$ are the observed residuals in Equation (8a) obtained by using the estimated parameter values of Equation (8b). If the null hypothesis is true, AGS has an F-distribution with 2 and n-2 degrees of freedom, AGS ~ F(2, n-2). Thus, there is a significant difference between creep model predictions if AGS > F_{crit}, where F_{crit} is the critical value for the F distribution.

If creep model predictions are further assumed to be unbiased (so $(\bar{e}_1^2 - \bar{e}_2^2) = 0$), the AGS test then suggests that the equality of mean percentage squared errors is equivalent to equality of prediction error variances. This leads to the MGN test. This test starts with the regression equation

$$s_i = \lambda(g_i) + \varepsilon_i \quad (9a)$$

The null hypothesis H_o: $\lambda = 0$ states that the sum of the prediction errors from the two creep models equals random values given by ε_i, i.e., the prediction errors from each model differ from each other by random unexplained amount ε and so they are essentially the same. The standard t test can be used to test this hypothesis

$$MGN = \frac{\hat{\lambda} - 0}{\sqrt{\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{(\sum_{i=1}^{n}\{g_i^2\})(n-1)}}} \quad \text{with} \quad \hat{\lambda} = \frac{\sum_i^n s_i g_i}{\sum_{i=1}^n g_i^2} \quad (9b)$$

where $\widehat{\varepsilon_i}$ are calculated from Equation (9a) using the value for λ given by Equation (9b). Under the null hypothesis that $\lambda = 0$ (i.e. the error in predictions from each creep model are the same), MGN follows a Student t distribution with n - 1 degrees of freedom.

The standard assumptions required for the AGS and MGN tests to be valid are that the residuals u_i and ε_i both have a mean of zero, both have a constant variance, and that these residuals are uncorrelated with $s_i - \overline{s}$ and g_i respectively. It must be additionally assumed that the u_i and ε_i are normally distributed. However, Diebold and Mariano provide simulation evidence showing that the test can be seriously over-sized, even in very large samples, when the normality assumption does not hold. This drawback could be serious, as heavy-tailed distributions for prediction errors would seem to be quite plausible in practice.

Moreover, this problem persists, and, indeed, becomes worse as the sample size increases. It is therefore desirable to have a test that is robust to this type of non-normality. The source of this problem is that the variance of λ in the denominator Equation (9b) is an inconsistent estimator of the true variance for λ. This in turn is a consequence of the fact that, although g_i and s_i are uncorrelated under the null hypothesis, they are not independent.

Harvey et. al. [12] recommend a modification of the above test to overcome this problem

$$MGN^* = \frac{\widehat{\lambda} - 0}{\sqrt{\frac{\sum_{i=1}^{n} g_i^2 \hat{\varepsilon}_i^2}{\left(\sum_{i=1}^{n} g_i^2\right)^2}}} \tag{9c}$$

Although the distribution for MGN* is no longer exactly a Student t distribution, a reasonable practical procedure is to compare this test statistic with critical values from the Student's t distribution with (n - 1) degrees of freedom. Simulations by Harvey et. al. revealed that the MGN* test does yield a test of the correct size for large samples. However, in the smallest samples, the MGN* test is seriously over-sized even when the error distribution is normal – more so than the MGN test. Thus, they recommend use of the MGN* test only when moderately large samples are available. In the absence of large samples, they recommended a consideration of non-parametric approaches.

A non-parametric test

Non-parametric procedures are recommended when any of the assumptions behind the above tests are invalid. Non-parametric tests are said to be distribution free in that they don't assume a normally distributed population, and so can be used in the presence of non-normality. To compare the accuracy of two sets of creep property predictions, the sign test (Mendenhall and Reinmuth [17]) is based on the number of times that predictions from creep model 1 are better than predictions from creep method 2. Any ties are ignored. As such it can be used to see if there are difference between two MPSE's or MPAE's. The number of times creep model 1 is better than model 2 is given by the binomial distribution, and under the null hypothesis that both creep models are equally accurate, the probability parameter p of this distribution is 0.5. That is, if both creep models are equally accurate, the probability of a specific creep model producing a better prediction than another model is 0.5 for each time a prediction is made. The binomial distribution with p = 0.5 can be written as

$$P[X = x) = C_x^n 0.5^x (1 - 0.5)^{n-x} \tag{10a}$$

where X is the number of times creep model 1 produces better predictions than model 2, when n predictions are made. P[X = x] reads the probability that the random variable X will equal a specific value x (e.g. the probability that creep model 1 produces x = 10 better predictions in the n predictions produced). C_x^n are the number of different ways of getting x superior predictions from model 1, in the n predictions made.

If a large or small value for x is actual observed, but this value for x leads to P[X = x] being very small (and below a selected significance level) when using p = 0.5 in Equation (10a), then this low probability for observing something that has actually happened can only be explained by p not being equal to 0.5. In which case the null hypothesis is incorrect, and creep model 1 performs differently to creep model 2 at the selected significance level (whether its better or worse depends on whether x is small or large). More formally, let x equal the number of times creep model 1 produces a percentage prediction error smaller than creep model 2

$$x = \#e_{1i} < e_{2i} \qquad \text{for } i = 1 \text{ to } n \tag{10b}$$

No matter how the prediction errors are distributed, this count will always have a binomial distribution given by Equation (10a) under the null hypothesis that the two creep models are equally accurate in prediction (assuming independence). In the context of creep model comparisons, the null hypothesis is always of the form H_0: p = 0.5 so that each creep model is equally likely to produce the best prediction at any one test condition. The alternative is that H_a: p > 0.5. Here H_a says that creep model 1 is more likely to produce the best prediction at any one test condition. The probability that the null hypothesis is true, the p-value, can be evaluated using the following version of the binomial distribution

$$\text{p-value} = PX \geq x] = \sum_x^n \frac{n!}{(n-x)!x!} 0.5^x \, (1-0.5)^{n-x} \tag{10c}$$

While this sign test has the attraction of being distribution free so that very few assumptions need to be made, it is important to realise that they ignore some of the information which is available within the data - the sign test ignores the numerical size of the errors in prediction for example. This loss of information can result in the power of the sign test being considerably smaller than those for the parametric tests described above – depending on the specific circumstances.

Testing the assumptions behind parametric tests

Normality of u_i and ε_i can be tested using the test statistic proposed by Doornik and Hansen [18]. This is a more powerful version of the asymptotic version proposed earlier by Jarque and Bera [19]. Both these approaches involve testing for the presence of skewness and kurtosis in the least squares residuals. Testing for normality of the u_i in Equation (8a) can be done using the Jarque and Bera test statistic

$$JB = \frac{n(w_1)^2}{6} + \frac{n(w_2-3)^2}{24} \tag{11a}$$

where

$$w_1 = \frac{\sum_{t=1}^{N}(\hat{u}_t - \bar{u})^3/N}{\{\sum_{t=1}^{N}(\hat{u}_t - \bar{u})^2/N\}^{2/3}} \quad \text{and} \quad w_2 = \frac{\sum_{t=1}^{N}(\hat{u}_t - \bar{u})^4/N}{\{\sum_{t=1}^{N}(\hat{u}_t - \bar{u})^2/N\}^2} \tag{11b}$$

Under the null hypothesis of no skewness and kurtosis in the residuals, this test statistic has an asymptotic chi square distribution with 2 degrees of freedom, JB ~ $\chi(2)$. Thus, there is a significant deviation from normality when JB > χ_{crit}, where χ_{crit} is the critical value for the chi square distribution. Testing for normality of the ε_i in Equations (9a) can be done by replacing terms contain u with ε in Equations (11). The Doornik and Hansen [18] version simply adjusts this test statistic for small samples.

Homoscedasticity of the residuals u_i or ε_i can be tested using the test statistics proposed by White [20] and by Nicholls and Pagan [21] - both of which look to see if the variance in these residuals, as estimated using the squared residuals, depends on all the explanatory variables (and their squares) that are in the regression equation. The involves estimating the $\hat{u}_t$ in Equation (8a) using the λ estimates in Equation (8b). Then $\hat{u}_t^2$ is regressed on (s_i-$\bar{s}$) and (s_i-$\bar{s}$)^2 and the coefficient of determination R^2 extracted. Under the null hypothesis of homoscedastic residuals, the statistic W = NR^2 has a chi square distribution with 1 degree of freedom, $W_1 \sim \chi(1)$. Thus, there is a significant deviation from homoscedasticity when W_1 > χ_{crit}. Homoscedasticity of the residuals ε_i in Equations (9a) can be tested by first estimating the, $\hat{\varepsilon}_t$ in Equation (9a) using the λ estimate in Equation (9b). Then $\hat{\varepsilon}_t^2$ is regressed on g_i and (g_i)^2 and the coefficient of determination R^2 extracted. Under the null hypothesis of homoscedastic residuals, the statistic W = NR^2 has a chi square distribution with 1 degree of freedom. $W_2 \sim \chi(1)$. Thus, there is a significant deviation from homoscedasticity when W > χ_{crit}.

Specification of two parametric creep models

The above tests will be applied to two very different creep models to test whether they produce significantly different predictions of long-term life. The aim is not to establish superiority of one model over the other – simply to identify whether any statistically significant differences exist between them. The first model used in this paper is that attributed to Orr-Sherby-Dorn (OSD) [22] and is given by

$$\ln(t_F) = \ln(B) + n\ln(\sigma) + \frac{Q_c}{R}\left(\frac{1}{T}\right) \tag{12a}$$

where B and n are further model parameters. The model implies that there is a linear relationship between log failure time and log stress, and the role of temperature is to shift this linear relationship in a parallel fashion. Another approach that is increasingly common in the literature is to normalise stress using the high temperature tensile strength of the material σ_{TS} with the additional constraint that when $\sigma = \sigma_{TS}$, t_F will equal zero and as s tends to zero t_F tends to infinity. This approach is rather appealing because provided the tensile strength is measure at a high enough constant strain rate, σ_{TS} should represent the stress inducing instantaneous failure upon the application of a constant load in a creep test. Such a constraint is introduced using an inverted S shaped curve for the relationship between σ/σ_{TS} and t_F at constant temperature. Thus, Wilshire and Battenbough [23] introduced the creep model

$$t_F = A\left[-\ln\left(\frac{\sigma}{\sigma_{TS}}\right)\right]^m \exp\left\{\frac{Q_c}{RT}\right\} \tag{12c}$$

which at constant temperature is the same equation that describes the inverted Weibull cumulative density function. Yang et. al. and Wang et. al. [24,25] introduced the model

$$t_F = A\left[\frac{\sigma}{\sigma_{TS}-\sigma}\right]^{-m} \exp\left\{\frac{Q_c}{RT}\right\} \tag{12d}$$

which at constant temperature is the same equation that describes the inverted Logistc cumulative density function. Finally, Evans [26] unified these two models by adding an additional parameter k, so allowing for other S shaped curves that fall within these two special cases

$$t_F = A\tau^m \exp\left\{\frac{Q_c}{RT}\right\} \qquad \text{where} \quad \tau = \left[k\left(\frac{\sigma}{\sigma_{TS}}\right)^{-1/k} - k\right] \tag{12e}$$

This unification comes about because as $k \longrightarrow \infty$, $\tau^m \longrightarrow \left[-\ln\left(\frac{\sigma}{\sigma_{TS}}\right)\right]^m$ and when k = 1 $\tau^m = \left[\frac{\sigma}{\sigma_{TS}-\sigma}\right]^{-m}$. As a cumulative probability must range between 0 and 1, all the models contained within Equations (12e) have the constraint $t_F \longrightarrow \infty$ as $\sigma/\sigma_{TS} \longrightarrow 0$ and $t_F \longrightarrow 0$ as $\sigma/\sigma_{TS} \longrightarrow 1$. So, the transformed stress given in Equation (12e) is not some arbitrarily correction made to stress – instead it is integral to some new approach in the literature to creep life prediction. Further, these S shaped creep models have been successfully applied to many different high temperature materials for example, applications to 2.25Cr-1Mo [27], 316 stainless steels [28], and Inconel 740/740H [29].

Estimating these parametric model

In all applications of these models, there is a need to estimate their unknown parameters, and this is complicated by the fact that this should be done in a piecewise fashion. This is required because the model parameters will be different over different stress and temperature ranges to reflect the different creep mechanism operating in such ranges. This creates a further complication, namely that the test conditions at which such creep mechanisms change must also be part of the estimation procedure. The following approach estimates such break points from the data and allows the model parameters to be different either side of these break points.

Wilshire and Whittaker [28], when studying all the batches of 2.25Cr-1Mo steel in the NIMS creep data base, and using Equation (12c) identified two distinct breaks and so three distinctly different creep mechanisms. To explain this, they suggested that at the highest stresses creep takes place through the generation and movement of new dislocations, formed at appropriate sources, since the yield stress of the material is exceeded. These new dislocations that are continuously generated due to the high stress, leads to large net movement of atoms, and so contributes to high creep rates making failure times very sensitive to changes in stress (and so large values for m in Equation (12c)). This creep occurs largely because of these dislocations moving within the grains, and under these circumstances Q_c is expected to be high and equivalent to that for self-diffusion in bainitic matrices. At intermediates stresses, where the stress is below the yield stress, these authors suggest creep occurs not by the generation of new dislocations but by the movement of the dislocations pre-existing in the as received bainitic microstructure. With fewer atoms moving, the creep rate slows (and so m falls in value). Provided the movement of preexisting dislocation is predominantly within grains the activation energy will remain unchanged.

However, Whittaker and Harrison [30] have suggested that at these intermediates stresses creep takes place predominantly within the grain boundary zones rather than in the grains, and this would reflect itself in a lower activation energy for creep representing diffusion along dislocations and grain boundaries. They did not formally test for this and so this paper

will use an estimation strategy that does. Then at the lowest stresses, Wilshire and Whittaker state that the bainitic regions transform to ferrite and coarse molybdenum carbide particles due to the long-term nature of the tests conducted at the very low stresses. This they argued result in creep once again taking place within the grains where the activation energy would be higher and equal to that for self-diffusion. In comparison to the Wilshire and Whittaker study, Brear [31] identified only two distinct ranges of test conditions when using Equation (12a), and he suggested that this change may be due to the high amounts of oxidized material seen on the failed creep specimens at the very lowest stresses – the result of prolonged testing. This reduction in the number of implied creep mechanisms will also be tested in the results section below.

The presence of three different stress regimes suggested by the Wilshire and Whittaker study can be accommodated for in the Evans model by using the following modification of Equation (12e)

$$\ln[t_F] = \ln[A] + m\ln[\tau] + m_1\max\left[0, \ln(\tau) - \ln(\tau_1)\right] + m_2\max\left[0, \ln(\tau) - \ln(\tau_2)\right] + Q_c\frac{1}{RT} + +m_3\frac{D}{RT} \quad (13a)$$

where $\tau_1 < \tau_2$ and these are the values for τ where there is a change in creep mechanism and D is a dummy variable that equals 1 when $\tau_1 < \tau < \tau_2$ and zero otherwise. m_1 to m_3 are three additional parameters that require estimation. Thus, in the high stress regime where $\tau < \tau_1$, the model simplified to

$$\ln[t_F] = \{\ln[A]\} + (m)\ln[\tau] + \frac{Q_c}{RT} \quad (13b)$$

and in the low stress regime where $\tau > \tau_2$, the model becomes

$$\ln[t_F] = \left\{\ln[A] - m_1\ln(\tau_1) - m_2\ln(\tau_2)\right\} + (m + m_1 + m_2)\ln[\tau] + \frac{Q_c}{RT} \quad (13c)$$

and in the intermediate stress regime where $\tau_1 < \tau < \tau_2$, the model becomes

$$\ln[t_F] = \left\{\ln[A] - m_1\ln(\tau_1)\right\} + (m + m_1)\ln[\tau] + \frac{Q_c + m_3}{RT} \quad (13d)$$

Thus in the intermediate stress regime the activation energy drops form Q_c (which represents that for self-diffusion) to Q_c+m_3, as the expectation is for m_3 to be negative based on the Wilshire and Harrison [30] paper. In the high and low stress regimes the activation energy will be higher and estimated by Q_c . But if $m_3 = 0$, that will support the view put forward by Wilshire and Whittaker [28] that dislocation movement remains predominantly within the grains in the intermediate stress range. When jumping from the high to intermediate stress regime parameter m changes to a value of $m+m_1$, with the expectation (based on the analysis by Wilshire and Whittaker and Whittaker and Harrison) that m_1 will be negative so that the value for n diminishes as the stress falls (due to fewer moving dislocations). Parameter m changes to $m+m_1+m_2$ in the low stress regime. If the value for m_2 is such that $m+m_1+m_2$ approximates m, that would support the view that microstructural degradation is taking place so increasing m. But, if the value for m_2 is such that $m+m_1+m_2$ drops off towards a value of 1,

that would be consistent with Nabarro-Herring creep. This then represents a piecewise fitting procedure where the parameters are different over different stress ranges.

To estimate values for the parameters in Equations (13) the following procedure can be used. First k is set equal to 1,with τ_1 and τ_2 set equal to random values within the limits of the largest and smallest values for τ within the sample of data. This enables a τ value to be associated with each failure time in the NIMS creep data base so that all the variables on the right-hand side of Equation (13a) are fully quantified allowing $\ln(t_F)$ to be regressed on $\ln(\tau)$, $\max[0, \ln(\tau) - \ln(\tau_1)]$, $\max[0, \ln(\tau) - \ln(\tau_2)]$, 1/RT and D/RT to obtain least squares estimates for A, m, $m_1 - m_3$ and Q_c. This regression equation has a residual sum of squares (RSS) associated with it - that of course is minimised for the chosen value for k, τ_1 and τ_2 by the linear least squares procedure. Then a generalized reduced gradient non-linear search technique that uses centralised numerical derivatives is used to search for values of k, τ_1, and τ_2 that minimise the RSS associated with the regression line given by Equation (13a). This non-linear search is carried out using Excel's Solver [32] subroutine.

The presence of three different stress regimes suggested by Wilshire and Whittaker study can be accommodated for in the OSD model by using the following modification of Equation (12a)

$$\ln[t_F] = \ln[B] + n\ln[\sigma] + n_1\max[0, \ln(\sigma) - \ln(\sigma_1)] + n_2\max[0, \ln(\sigma) - \ln(\sigma_2)] + Q_c\frac{1}{RT} + +n_3\frac{D}{RT} \quad (14a)$$

where $\tau_1 < \tau_2$ and these are the values for τ where there is a change in creep mechanism and D is a dummy variable that equals 1 when $\tau_1 < \tau < \tau_2$ and zero otherwise. n_1 to n_3 are three additional parameters that require estimation. Thus, in the high stress regime where $\tau < \tau_1$, the model simplified to

$$\ln[t_F] = \{\ln[B]\} + (n)\ln[\tau] + \frac{Q_c}{RT} \quad (14b)$$

and in the low stress regime where $\sigma > \sigma_2$, the model becomes

$$\ln[t_F] = \{\ln[B] - n_1\ln(\sigma_1) - n_2\ln(\sigma_2)\} + (n + n_1 + n_2)\ln[\sigma] + \frac{Q_c}{RT} \quad (14c)$$

and in the intermediate stress regime where $\sigma_1 < \sigma < \sigma_2$, the model becomes

$$\ln[t_F] = \{\ln[B] - n_1\ln(\sigma_1)\} + (n + n_1)\ln[\sigma] + \frac{Q_c + n_3}{RT} \quad (14d)$$

This then represents a piecewise fitting procedure where the parameters are different over different stress ranges. The parameters of Equation (14a) can be estimated in the same way as for the Evans model by searching, using Excel's Solver, for the optimal break points (value for where the creep mechanism changes – σ_1 and σ_2)

Results: An application of statistical tests for 2.25Cr-1Mo steel

Data sources

This paper makes use of the creep failure times contained within Creep Data Sheets 3B & 50A, published by the Japanese National Institute for Materials Science (NIMS) [33-34]. This has extensive data on twelve batches of 2.25Cr-1Mo steel where each batch has a different chemical composition that underwent one of four different heat treatments - details of which are given in [33]. This paper makes use of just one of these batches, the MAF batch, which was in tube form that had an outside diameter of 50.8mm, a wall thickness of 8mm and a length of 5000 mm with a chemical composition as shown in **Table 1**. Specimens for creep testing were taken longitudinally from this material. Each test specimen had a diameter of 6mm with a gauge length of 30mm. The creep tests were obtained over a wide range of conditions: 400 MPa - 22MPa and 723K – 923K. Figure 1 plots the failure times obtained for this MAF batch at the different stresses and temperatures used.

Figure 1. Relationship between stress, temperature, and time to failure for the MAF batch of 2.25Cr-1Mo steel contained within NIMS creep data sheets 3B &50A [33-34].

The same NIMS data sheet also contains results from high temperature tensile testing. The tensile strength measurements shown in the first row of Table 2 were obtained using a constant strain rate of $0.0013s^{-1}$. The models defined by Equations (12c-e) require σ_{TS} to represent that stress that will induce -in practical terms - instantaneous creep rupture when it is instantly applied (so inducing a very high strain rate) to a specimen in a creep test. It follows therefore that σ_{TS} must be obtained from a high temperature tensile test that is conducted at a sufficiently high strain rate – one high enough to make the tensile strength values independent of the chosen strain rate. However, a recent paper by Evans [35] has revealed that the NIMS stain rate of $0.0013s^{-1}$ is not sufficiently high enough to meet this criteria, because data on the same2.25Cr-1Mo steel obtained by Bueno and Sobrinho [7] shows the tensile strength measurements remain strain dependent above a strain rate of $0.0013s^{-1}$. The steel tensile tested by Bueno and Sobrinho was supplied in plate form with 25.4 mm thickness, according to ASTM A 387, grade 22, in the normalized and tempered condition, with a chemical composition shown in the second row of Table 1. Metallographic analysis indicated the presence of 30% bainite and 70% ferrite in the as-received condition. The specimens for tensile testing were extracted from the rolling direction and a gauge length Lo = 25 mm and an initial diameter do = 6.25mm were used for all specimens. The constant strain rate tests were carried out in a servo-hydraulic 8802 model INSTRON machine.

Table I. Chemical Composition for Different Batches of 2.25Cr- 1Mo Steel (%/wt)

Source	Cr	Mo	C	Si	Mn	P	S	Ni	Cu	Al
MAF [33-34]	2.46	0.94	0.1	0.23	0.43	0.011	0.009	0.008	0.07	0.005
Bueno & Sobrinho[7]	2.09	1.08	0.097	0.32	0.50	0.007	0.002	0.03	0.01	0.05
Balance Fe.										

Whilst the chemical composition of the specimens tested by Bueno and Sobrinho (Table 1) are slightly different from those associated with the NIMS MAF batch, these differences are not enough to explain the extent to which the Bueno and Sobrinho tensile strength measurements obtained at the much higher strain rate of $0.013s^{-1}$ differ from the NIMS data

obtained at $0.0013s^{-1}$. For this reason, the results shown below for the Evans model make use the NIMS stresses normalised using the Bueno and Sobrinho tensile strength measurements shown in Table 2.

Table 2. Tensile strength measurements made on two different batches for 2.25Cr- 1Mo Steel (MPa)

Data Source	723K	748K	773K	798K	823K	873K	923K
MAF [33-34]	448	423	394	380	365	283	216
Bueno & Sobrinho[7]	534*	513*	486	471*	456	413	360

*Interpolated values based on the temperature dependency of the tensile strength

Estimating the Evans model

The estimation technique described above is applied to the NIMS failure times with values of 10,000h or less, i.e. to the solid data points in Figure 2(a). Failure times above 10,000h are shown as open circles. It was found that the RSS is minimised when k = 1 suggesting the Yang et. al. model [24] represents the creep data better than the Wilshire model (hence the vertical axis shows $\ln[(\sigma_{max}/\sigma) -1]$ – which is what τ in Equation (12e) collapses to when k = 1). The model's predictions are shown by the solid segmented line and this fits the short-term data (failure times less than 10,000h) very well– with an R^2 value of 98.04%. The model yields an activation energy of 309 $kJmol^{-1}$ in the high and low stress regions - which is quite close to that for lattice self-diffusion in this material (350 $kJmol^{-1}$). This value is therefore consistent with dislocation creep controlled by lattice self-diffusion in these stress regions. Whilst the activation energy is slightly different in the intermediate stress region, this difference is not statistically significant. This suggests that in this region dislocation movement is still predominantly within grains, but because the stresses are below the yield stress, it is the movement of preexisting rather than new dislocations – hence the observed increase in the value for m and mainly unchanged value for Q_c in this region. Hence the Whittaker and Harrison [30] suggestion that creep is predominantly the result of dislocation movement between grain is not fully supported by the data.

The predictions of failure times beyond 10,000h are summarised in Figure 2(b). Notice how close the trend line fitted to the data points in this Figure is to the 45^o line where actual and predicted values are identical. The equation of this trend line picks up systematic errors made by the creep model in predicting time to failure, whilst the scatter of the data points around the trend line picks up the random prediction errors that the model makes. With the power term of predicted time being almost unity, the trend line reveals that this creep model has systematic prediction errors averaging 5.9% (1/0.944) above the actual failure times.

Figure 2. Showing (a) the Evans representation of failure times using σ_{TS} from Bueno and Sobrinho, where the model is estimated using $t_F < 10{,}000h$ and, (b) actual v predicted t_F values beyond 10,000h.

Estimating the OSD model

Like for the Evans model above, the parameters of Equation (14a) were estimated using only failure times of 10,000h or less. The model's predictions are shown by the solid segmented line and the fit to the short-term data (failure times less than 10,000h) is again good – with an

R^2 value of 97.30% - slightly below that for the Evans model. There appears to be two breaks in this prediction line at a log stress of 4.44 and 4. Both breaks are statistically significant and so the data does not support Brear's [31] claim of a single break. The model yields an activation energy of 383 $kJmol^{-1}$ in the high and low stress regions - which is again quite close to that for lattice self-diffusion in this material (350 $kJmol^{-1}$). Whilst the activation energy is slightly different in the intermediate stress region, this difference is not statistically significant. In the low stress regime n continues to fall in absolute terms to a value of around 2. This decrease is not consistent with Brear's oxidation argument nor with Wilshire and Whittaker's particle coarsening argument, It is more consistent with a move towards diffusion creep within grains at lower stresses. The predictions of failure times beyond 10,000h are summarised in Figure 3(b). Notice the fitted trend line in this Figure has a slope different from the 45° line, implying a larger degree of systematic error compared to the Evans model. That is, at low failure times the model consistently underpredicts with the opposite occurring at higher failure times.

Figure 3. Showing (a) the OSD representation of failure times, where the model is estimated using $t_F < 10{,}000$h and, (b) actual v predicted t_F values beyond 10,000 h

Figure 4 shows the predictions from the Evans and OSD models in the more familiar stress v time space. The tendency for the OSD model to produce the larger predicted times at a given stress and temperature is clearly seen in this figure.

Figure 4. Relationship between stress, temperature, and time to failure for the MAF batch of 2.25Cr-1Mo steel contained within NIMS creep data sheets 3B &50A [31-32], together with the predicted values from the OSD and Evans models.

Statistical tests

Figures 2(b) and 3(b) reveal differences in the characteristics of the predicted failure times made by each creep model. The Evans model slightly overpredicts all the failure times beyond 10,000h, whilst the OSD model under predicts at the lowest times and over predicts and the higher failure times. If the Evans model is labelled creep model 1 and the OSD model creep model 2 then the data in Figures 2(b) and 3(b) produce $RMPSE_1 = 30.67\%$, $RMPSE_2 = 46.23\%$ with $MPAE_1 = 25.31\%$ and $MPAE_2 = 42.55\%$. Clearly then, the Evans model is producing the most accurate predicted lifetimes, but the question remains as to whether this result is a peculiarity of this sample of data on 2.25Cr-1Mo steel or whether these differences are real. Not also how squaring the prediction errors (leading to the RMPSE) underestimates the predictive accuracy of each model as RMSPE > MPAE.

Using $d_i = (e_{1i})^2 - (e_{2i})^2$, gave DM = -26.14 and HLN = -26.68, whilst using $d_i = |e_{1i}| - |e_{2i}|$ gave DM = -19.71 and HLN = -20.12. Using a 5% significance level, the value for $Z_{crit} = 1.96$ and the value for $T_{crit} = 2.06$. So, $|DM| > Z_{crit}$ and $|HLN| > T_{crit}$. Consequently, the Evans model produces statistically significantly better mean squared percentage errors and mean percentage absolute errors, irrespective of whether the original or modified DM test is used. That is, the chances of the population average value for d being zero is less than 5%. Further, given the sample of 25, and based on the results of Diebold-Mariano [11], these tests reject the null hypothesis only about 0.3 percentage points more than they should.

Estimation of Equation (8a) yielded $g_i = 0.0801 - 0.3634(s_i - \bar{s})$ with an R^2 value of 16.03%. The AGS statistic came out at 4.39, and with $F_{crit} = 3.42$ at the 5% significance level, it can again be concluded that the Evans model produces a statistically significantly lower mean percentage square error (i.e. the null hypothesis of $\lambda_1 = \lambda_2 = 0$ is rejected at the 5% significance level). The Student t statistic on λ_1 is 0.80 and -2.10 on λ_2. Consequently, λ_1 is not significantly different from zero, whilst λ_2 is significantly less than 1 - which together imply the predictions from the OSD model are not more accurate than those from the Evans model.

Recall that this test assumes that the u_i in Equation (8a) are normally distributed with a constant variance. The JB statistic comes out at 4.12, and with $\chi_{crit} = 5.99$ at the 5% significance level it can be concluded that the null hypothesis of normality cannot be rejected at this significance level. Further, the White test for homoscedastic residuals produced a test statistic value of W = 4.15, and with $\chi_{crit} = 3.84$ at the 5% significance level it can be concluded that the null hypothesis that the u_i have constant variance can be rejected at this 5% significance level. But with $\chi_{crit} = 6.63$ at the 10% significance level it can be concluded that the null hypothesis that the u_i have constant variance cannot be rejected at this higher 10% significance level. Perhaps then, some degree of cation should be attributed to the results of the above AGS test statistic.

Estimation of Equation (9a) yielded $s_i = -0.43(g_i)$, with an R^2 value of 15.23%. The MGN statistic associated with this regression comes out at -2.08 and for the modified version MGN^* = -2.77. Both |MGN| and $|MGN^*|$ exceed the value of T_{crit} of 2.06 and so once again the null hypothesis that both models produce the same MPSE can be rejected at the 5% significance level. Indeed, the negative value for λ means that the Evans model produces a significantly small MPSE.

Again, this test assumes that the ε_i in Equation (9a) are normally distributed with a constant variance. The JB statistic comes out at 0.89, and with $\chi_{crit} = 5.99$ at the 5% significance level it can be concluded that the null hypothesis of normality cannot be rejected at this significance level. Further, the White test for homoscedastic residuals produced a test statistic value of W = 3.99, and with $\chi_{crit} = 3.84$ at the 5% significance level it can be concluded that the null hypothesis that the ε_i have constant variance can be rejected at this 5% significance level. But with $\chi_{crit} = 6.63$ at the 10% significance level it can be concluded that the null hypothesis that the ε_i have constant variance cannot be rejected at this higher 10% significance level. Perhaps then, some degree of cation should be attributed to the results of the above MGN test statistic.

As there is an element of uncertainty surrounding the constancy of the variance for u_i and ε_i, it is worth looking at the results from the sign test that do not require this assumption. For the sign test, the Evans model produced a smaller percentage prediction error (measured either as an absolute or squared error) 20 times out of the 25 predictions made. The probability of getting such a count or higher under the assumption that both models are equally accurate (p = 0.5) is, from the Binomial distribution, 0.046%. That is, such counts are highly unlikely, yet given that such a result has occurred, the only way of explaining this is to conclude that p >

585 0.5. Consequently, the Evans model has a higher probability of producing a better prediction
586 of time to failure at any stated test condition compared with the OSD model.

587 ## Conclusions

588 This paper has put forward several statistical tests to assess whether one creep model
589 produced better predictions of creep properties than another. These tests were obtained from a
590 review of the literature – mainly within the field of economic time series forecasting where
591 such approaches are well developed. The tests reviewed were based on either the MPSE or the
592 MPAE, or indeed both. These statistics were then applied to predicted failure times for 2.25Cr-
593 1Mo steel obtained from two competing creep models - the well-known OSD model and a
594 recently proposed model by Evans. The purpose being, not to identify a model that will always
595 be superior for all materials or for different batches of this material, but to suggest a statistical
596 testing procedure that will be useful for practitioners to analyse their own results (it is unlikely
597 that the superiority of the Evans model displayed in this paper generalises to other materials).

598 More specifically, it was found that the Evans model produced both a lower RMPSE
599 and a lower MPAE. Furthermore, all the proposed statistical tests concluded that this observed
600 difference in the MPSE and the MPAE was real and not something that had occurred by chance.
601 Using the 5% significance level all tests concluded the Evans model produced statistically
602 significantly lower MPSE's and MPAE's (although the assumption of homogenous variances
603 required for these tests was only accepted at the 10% significance level). However, the
604 Binomial test, that does not require this assumption, came to the same conclusions as the other
605 tests. These statistical tests can be used by future practitioners to identify the best creep models
606 to work with for a particular high temperature material. Consequently, one area for future
607 research is to apply these methods to different materials and to different creep models.

608

609 Disclosure statement: The authors report there are no competing interests to declare.

610 Data: All data used in this publication are in the public domain: References [26-27]

611 Conflict of Interest**:** This research was not funded by research council grants or private sponsors
612 and as such there are no financial relationships to declare.

References

[1] Evans M. Creep in Materials - Origins, Analysis and Prediction Methodologies. *Journal of Materials Education*, 2008.

[2] Holdsworth SR, Askins B, Baker MA, Gariboldi E, Holmstro¨me S, Klenkf A, Ringelf 597 M, Merckling Z, Sandstromh GR, Schwienheeri N, Spigarellij S. Factors influencing creep 598 model equation selection. *International Journal of pressure Vessels and Piping*, 2008; 5: 80–88.

[4] Evans M. Semi-parametric estimation of the Wilshire creep life prediction model: an application to 2.25Cr-1Mo steel. *Materials Science and Technology*, 2019; 35(16): 1977-1987

[5] Holdsworth SR, Merckling G. ECCC developments in the assessment of creep-rupture data. In: Proceedings of sixth international Charles Parsons Conference on engineering issues in turbine machinery, power plant and renewables, Trinity College, Dublin, 16–18 September, (200.

[6] Abdallah Z, Gray V, Whittaker M, Perkins K. A Critical analysis of the conventionally employed creep lifing methods. *Materials*, 2014; 7(5): 3371-3398.

[7] José Francisco dos Reis Sobrinho, Levi de Oliveira Bueno, Correlation between creep and hot tensile behaviour for 2.25Cr-1Mo steel from 500°C to 700°C. Part 2: An assessment according to different parameterization methodologies, *Revista Matéria,*10(3) (2005), 463 – 471.

[9] Evans M. Estimating threshold stresses using parametric equations for creep: application to low-alloy steels, *Materials Science and Technology*, 2023; 1-16.

[10] Evans M. Assessing the predictive performance of creep models using absolute rather than squared prediction errors: An application to 2.25Cr-1Mo steel. *Materials at High Temperatures*, Published Online: 16 Oct 2023.

[11] Diebold FX, Mariano, RS. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 1995; 13: 253-63.

[12] Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 1997; 13: 281-91.

[13] Ashley LR, Granger CWJ, Schmalensee, R. Advertising and aggregate consumption: an analysis of causality. *Econometrica,* 1980; 48: 1149-68.

[14] Morgan WA. A test for significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika,* 1939-1940; 31: 13-19.

[15] Granger CWJ, Newbold P., *Forecasting Economic Time Series*, 1977, Academic Press, Orlando, FL.

[16] Granger CWJ, Newbold P. Some comments on the evaluation of economic forecasts. *Applied Economics*, 1973; 5: 35-47.

[17] Mendenhall W, Reinmuth J. E. *Statistics for Management and Economics*, 1982, Boston, Mass.: Duxbury Press.

[18] Doornik JA, Hansen H. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 2008; 70: 927–939.

[19] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 1980; 6: 255–259.

[20] White H. A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 1980; 48: 817–838.

[21] Nicholls DF, Pagan AR. Heteroscedasticity in models with lagged dependent variables. *Econometrica*, 1983; 51: 1233–1242.

[22] Dorn JE, Shepherd LA. What We Need to Know About Creep. In Proceedings of the STP 165 Symposium on The Effect of Cyclical Heating and Stressing on Metals at Elevated Temperatures, Chicago, IL, USA, 17 June 1954.

[23] Wilshire B, Battenbough, AJ. Creep and creep fracture of polycrystalline copper *Mater. Sci. Eng. A*, 2007, vol. 443, pp. 156–66.

[24] Yang M, Wang Q, Song XL, Jia J. Xiang ZD. On the prediction of long term creep strength of creep resistant steels. *International Journal of Materials Research*, 2016, 107(2).

[25] Wang Q, Yang M, Song XL, Jia J, Xiang ZD. Rationalization of Creep Data of Creep-Resistant Steels on the Basis of the New Power Law Creep Equation. *Metall. Mater. Trans. A*, 2016, vol. 47A (7), pp. 3479–87.

[26] Evans M. Testing Model Structure Through a Unification of Some Modern Parametric Models of Creep: An Application to 316H Stainless Steel. *Metallurgical and Materials Transactions A, Metall. Mater. Trans. A*, 2020, vol. 51 (2), pp. 697–707.

[27] Wilshire B, Whittaker M. Long term creep life prediction for Grade 22 (2·25Cr-1Mo) steels. *Materials Science and Technology*, 2011; 27(3): 642-647.

[28] Whittaker MT, Evans M, Wilshire B, Long-term creep data prediction for type 316H stainless steel. *Materials Science and Engineering: A, 2016 ;*552: 145-150.

[29] Chen L, Dong Z, Song XL, Jia, J, Xiang ZD. Determination of Activation Energy and Prediction of Long-Term Strength of Creep Rupture for Alloy Inconel 740/740H: A Method Based on a New Tensile Creep Rupture Model. *Metallurgical and Materials Transactions A*, 2022, Vol. 53, pp. 1–5.

[30] Whittaker MT, Harrison W. The evolution of the Wilshire Equations for creep life prediction. Materials at High Temperatures, 2014, 31(3), pp. 233-238.

[31] Brear JM. A perspective on the Wilshire creep equations. *Strength, Fracture and complexity*, 2022; 15(1): 79-98.

[32] Microsoft Corporation. Microsoft Excel [Internet]. 2018. Available from: https://office.microsoft.com/excel.

[33] NIMS Creep Data Sheet No. 3B: Data Sheets on the Elevated-Temperature Properties of 2.25Cr-1Mo Steel for Boiler and Heat Exchanger Seamless Tubes (STBA 24), National Research Institute for Metals, Tokyo, Japan.

[34] NIMS Creep Data Sheet No.50A: Long-Term Creep Rupture Data obtained after Publishing the Final Edition of the Creep Data Sheets, National Research Institute for Metals, Tokyo, Japan.

[35] Evans M. The Important Role Played by High-Temperature Tensile Testing in the Representation of Minimum Creep Rates Using S-Shaped Curve Models. *Metallurgical and Materials Transactions A*, 2023, 54, pp. 4796–4805.

725