Marine Structures 91 (2023) 103466

Contents lists available at ScienceDirect

Marine Structures

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

Analysis of S–N data for new and corroded mooring chains at varying mean load levels using a hierarchical linear model

Erling N. Lone [a,*], Philippe Mainçon [b], Øystein Gabrielsen [c], Thomas Sauder [b,a], Kjell Larsen [a,c], Bernt J. Leira [a]

[a] Department of Marine Technology, Norwegian University of Science and Technology, Trondheim N-7491, Norway
[b] SINTEF Ocean, P.O. Box 4762, Torgarden, 7465 Trondheim, Norway
[c] Equinor ASA, Arkitekt Ebbells veg 10, 7053 Ranheim, Norway

ARTICLE INFO

Keywords:
Studless chain
S–N analysis
Hierarchical model
Mean load
Corrosion

ABSTRACT

Results from full scale fatigue tests of offshore mooring chains are analyzed. The data set includes new and used chains, tested at a variety of mean load levels. The used chains have been retrieved after operation offshore and include samples with varying surface conditions, ranging from as-new to heavily corroded. Based on a parameterized S–N curve intercept parameter, the effects of mean load and chain condition are estimated empirically by regression analysis. A hierarchical linear model is used, to account for and quantify correlations within subsets of the data. The choice of grouping criterion for the hierarchical model is discussed, and assessed based on the current data. Results show that the mean load and corrosion effects are both significant. Differences in the fatigue performance of new versus used chains are quantified and discussed.

1. Introduction

The S–N approach to tension–tension fatigue of mooring chain is considered. Capacity curves for fatigue assessment of mooring chain components are then traditionally given on the form

$$\log N = \log A - m \cdot \log S \tag{1}$$

where N is the number of cycles to failure at stress range S [MPa], A is referred to as the intercept parameter, m is referred to as the slope parameter and $\log(\cdot)$ usually refers to the common logarithm. The parameters A and m are estimated by fitting Equation (1) to the results of constant amplitude tests performed at various stress range levels, typically by use of linear regression analysis [1].

As pointed out by Gabrielsen et al. [2], the parameters of current fatigue design curves for mooring chain [3,4] are based on full scale tests for new chain, performed at a mean load of 20% of the minimum breaking load (MBL). In fatigue calculations following the procedure outlined by the standards, degradation due to corrosion is accounted for in a simplified manner by a reduction of the cross section area of the chains – representing the general material loss caused by corrosion – giving an increase in the effective stress ranges entering the calculations. Actual mean tension loads in the mooring lines are disregarded. However, fatigue tests performed in recent years for used chains retrieved after operation offshore [2,5–9] have revealed that (i) increased surface roughness and pits caused by corrosion have a detrimental effect on the fatigue capacity, and this effect is larger than that obtained from the simplified correction prescribed by the standards; (ii) the fatigue capacity is strongly dependent on the mean load, so that mooring chains tested

* Corresponding author.
E-mail address: erling.lone@ntnu.no (E.N. Lone).

https://doi.org/10.1016/j.marstruc.2023.103466
Received 28 March 2023; Received in revised form 17 May 2023; Accepted 6 June 2023

Nomenclature

$\log(\cdot)$	Common logarithm
$\mathrm{se}(\cdot)$	Standard error
$\sim$	Distributed as
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
α_j	Group-specific intercept, for $j \in \{1, \ldots, J\}$
β_1, β_2	Mean load and corrosion grade effect coefficients, respectively
λ_m	Mean load [% MBL]
μ_α	Mean intercept of hierarchical linear model
σ	Total standard deviation
σ_m	Nominal mean stress
σ_α	Group-level standard error
σ_ϵ	Data-level standard error
$A(\sigma_m, c)$	Mean load and corrosion dependent intercept parameter of S–N curve
$A_D(\sigma_m, c)$	Mean load and corrosion dependent *design* intercept parameter of S–N curve
c	Corrosion grade, support $[1, 7]$
$g_1(\sigma_m)$, $g_2(c)$	Mean load and corrosion grade functions, respectively
J	Number of groups in the hierarchical model
m	Slope parameter of S–N curve (stress range effect)
N	Number of cycles to failure
S	Stress range
ICC	Intraclass correlation
LOO	Leave-one-out
MBL	Minimum breaking load
MCMC	Markov chain Monte Carlo
REML	Restricted maximum likelihood

at lower mean loads performed better than chains tested at higher mean loads with similar degree of corrosion. The beneficial effect of a lower mean load has been reported also from full scale tests of new chain [10,11] and from numerical computations [12].

As it stands, the S–N relation in (1) suggests that for a given set of the parameters (A, m) the fatigue capacity depends solely on the stress range. Properly addressing mean load and corrosion effects then requires the use of an effective stress range that implicitly takes these into account. With regards to corrosion, stress concentration factors could be used to scale the stress ranges to represent corrosion effects more realistically than that obtained by a mere reduction of the cross section. However, quantification of such factors would not be straightforward under the S–N assumption.

Empirical correction models for the mean load effect on fatigue exist in the literature. Examples of such are the Gerber, Goodman, Morrow or Smith–Watson–Topper models, see e.g. [13,14]. These models may be used to transform the stress range to an equivalent stress range at the reference mean stress for which the relevant S–N curve is given (i.e., typically 20% MBL for mooring chain). However; the validity of these correction models for fatigue of mooring chain is questionable. See Gabrielsen et al. [9] for a discussion on this matter.

To account for the effects of both mean load and corrosion, Lone et al. [15] expressed the intercept parameter as a function of mean load and chain condition:

$$\log A(\sigma_m, c) = B_0 + B_1 \cdot g_1(\sigma_m) + B_2 \cdot g_2(c) \tag{2}$$

where $(B_j)_{j\in\{0,1,2\}}$ are coefficients; σ_m is the nominal mean stress; c is a *corrosion grade* describing the condition of the chain on a custom scale from 1 (as-new) to 7 (severe corrosion); and $g_1(.)$ and $g_2(.)$ are monotonically increasing functions of σ_m and c, respectively, which are parameter-free and chosen *a priori* (this is addressed later in Section 4 of the present paper). The coefficients of the model were then estimated empirically from a database consisting of full scale fatigue tests for new and used chains, using least-squares regression. However, the database contained only one new chain sample tested at a mean load different from 20% MBL and rather few used chain samples tested at mean loads below 10% MBL.

Mendoza et al. [16] used the same corrosion grade scale as a measure for chain condition (c), and quantified its effect by considering a S–N model similar to that described by Eqs. (1) and (2). Mean load variations within the test database were accounted for by existing mean stress correction models, but not used to quantify their effect empirically. They pointed out the importance of addressing the adverse influence on estimated model parameters from correlations within subsets of the data, and accounted for these by applying a hierarchical model with used chain samples grouped by the platform from which they originated.

The main contributions of the present paper are as follows. It unites the works presented in [15,16] by (i) estimating empirically both mean load and corrosion effects, and (ii) using a hierarchical linear model for the regression analysis. Furthermore, (iii) the

Table 1
Overview of full scale fatigue tests included in the data analysis. Total number of tests is 154 (new chain: 67 tests, used chain: 87 tests). Diameter and MBL are nominal values.

Project,	Batch[a]	#Tests	Material	Diameter[b]	MBL[b]	Mean load[b]	Corrosion	Service	Ref.
Platform			grade	[mm]	[kN]	[% MBL]	grade (1–7)	life [years]	
ND	ND-R3-76	14	R3	76	4884	20.0	1	–	[17]
	ND-R4-76	12	R4	76	6001	20.0	1	–	[17]
TWI	TWI-R4-76	3	R4	76	6001	20.0	1	–	[10]
	TWI-R4-127	3	R4	127	14 955	20.0	1	–	[10]
	TWI-R5-76	3	R5	76	7008	20.0	1	–	[10]
	TWI-R5-127	13	R5	127	17 465	10.0, 20.0	1	–	[10]
VIC	VIC-R4	7	R4	[102, 171]	[10 216, 24 292]	[7.9, 14.7]	1	–	[11,18]
	VIC-R5	12	R5	[70, 165]	[6021, 26 833]	[6.7, 14.4]	1	–	[11,18]
P1	P1-R4-114a	6	R4	114	12 420	16.0	2, 5, 6	10	[2,5,6]
	P1-R4-114b	4	R4	114	12 420	[15.7, 16.9]	5	13	–
	P1-R4-114c	5	R4	114	12 420	[16.7, 20.0]	5, 6	12	[2,8]
	P1-R4-114d	4	R4	114	12 420	[16.8, 18.0]	4	20	–
	P1-R4-114e	10	R4	114	12 420	8.0, 18.0	7	12	[9]
P2	P2-R4-142	9	R4	142	18 033	14.0	1	12	[6]
P3	P3-R4-138	4	R4	138	17 199	14.0	1	7	–
P4	P4-R4-126	4	R4	126	14 775	13.5	2	12	[2,5]
	P4-R4-136	7	R4	136	16 785	11.9	5, 7	19	[2,7]
P5	P5-R4-130	8	R4	130	15 559	6.4, 20.0	5, 6, 7	20	[2,8]
P6	P6-R3-137	9	R3	137	13 829	13.4	4, 7	18	–
	P6-R4-114	3	R4	114	12420	[15.7, 16.8]	4	5	–
P7	P7-R4-145	4	R4	145	18 665	9.7	1	15	[2,6]
P8	P8-R4-130	3	R4	130	15 559	17.0	2	16	–
	P8-R4-139	7	R4	139	17 407	16.0	5, 7	19	–

[a]Batch identifier, general rule: ⟨project/platform⟩–⟨material grade⟩–⟨diameter⟩. See main text for additional comments.
[b]Comma-separated values without brackets list unique entries for the given batch, whereas bracketed values are [min, max].

test database that forms the basis for the present study has been extended through inclusion of additional fatigue tests performed at a range of mean load levels for both new and used chain; (iv) the choice of a grouping criterion for the hierarchical model is discussed, and we argue in favor of a more refined subset division than that used in [16].

The paper is organized as follows. In Section 2, we describe the fatigue test database and review the corrosion grade scale. In Section 3, hierarchical linear modeling is reviewed. In Section 4, we describe the data analysis and the choice of a grouping criterion for the present data set is presented. Results are presented in Section 5 and discussed in Section 6. Conclusions are given in Section 7.

2. Fatigue test data

2.1. Test database

The following attributes are common for all the tests included in the present study:

- All tests are for studless chain.
- All tests were performed in artificial seawater.
- Only first fractures are included (i.e., continued tests to obtain failure in additional links are not considered).

An overview of the tests is given in Table 1. The new chain tests have been provided from three previous projects: the Noble Denton (ND) joint industry study on 76 mm studless chain of grade R3 and R4 [17]; the TWI joint industry project on fatigue performance of high strength and large diameter studless chain [10,19]; and a Vicinay (VIC) study on the fatigue performance of R4 and R5 chains for a range of diameters [11,18].

The used chains have been retrieved by Equinor after service at eight different platforms (corresponding to the notation P1 to P8 in Table 1) in the North Sea and the Norwegian Sea, and then tested at constant amplitude cyclic loading until failure. Results from several of these tests have previously been published in [2,5–9]. These samples cover a variety of diameters and chain conditions — see listed references for additional details where available. Note the following for the used chain samples: (i) They have been visually inspected for presence of fatigue cracks prior to testing, and no such cracks have been observed. (ii) A corrosion grade has been assigned to each sample, as a quantitative measure of the chain's surface condition. (iii) The used chains have experienced limited general corrosion and corresponding material loss, with typically less than 1 mm for all groups but one. The exception is the R3 chains for platform P6 (referred to as P6-R3-137 in Table 1), for which the diameter is reduced by 4 mm on average.

Table 2
Description of corrosion grades.

Grade (c)	Surface condition
1	New chain, or mild corrosion.
2	Some scattered local corrosion (pitting), less than 1 mm deep.
3	Larger areas effected, local corrosion (pitting) around 1 mm deep.
4	Large areas affected by pitting, 1–3 mm deep, the deeper sharper in nature.
5	Severe and widespread pitting, up to 4 mm deep, but somewhat less than grade 6.
6	Severe and widespread pitting, 3–6 mm.
7	Severe and widespread pitting, 3–6 mm and deeper. Sharp and deep pitting.

Fig. 1. Examples of corrosion grades (c). None of the used chain samples in the database were graded with $c = 3$.

In the table, the tests from each project and platform have been divided into subsets referred to herein as "batch". A general rule used to define these subsets has been to distinguish between different chain sizes (diameter) and material grades (R3/R4/R5).[1] For most of the platforms (P2–P8), the general rule ensures that used chain samples retrieved from different positions along the mooring lines (e.g., top or bottom chain) and after the same service life duration are grouped into different subsets. For one of the platforms (P1), the top and bottom chains are of the same diameter and the samples have been retrieved at various points in time. The subsets defined for this platform have therefore been further refined based on service life and time of retrieval.

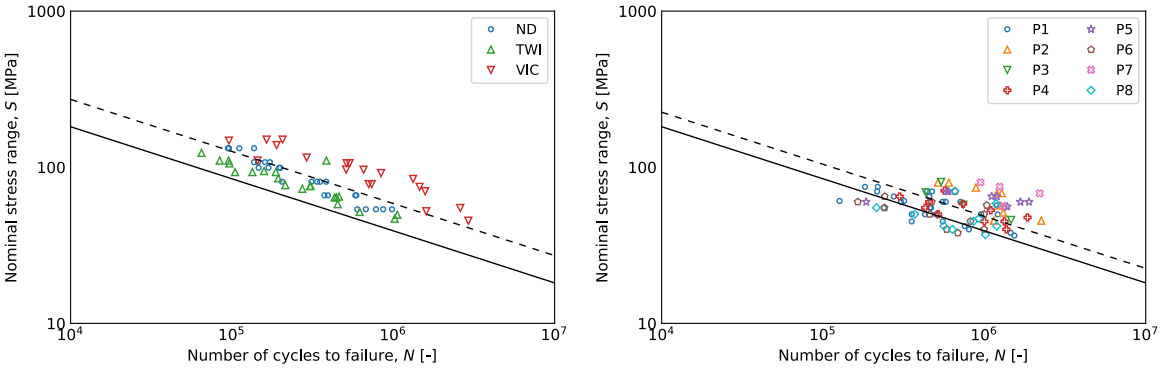
A difference across subsets which is not reflected in Table 1 is the number of links included in the test rig for each of the tests. The ND tests were conducted with 7 links; TWI tests were conducted with 7 links for 127 mm chain and 11 links for 76 mm chain; Vicinay's setup ranged from 2 to 9 links; and the used chain tests varied between 3 and 6 links, with 5 links for the majority of the tests. Also, the solution of the artificial seawater applied in the test varied slightly depending on test facility and project: a sodium chloride concentration of 3.5 wt% was used for the majority of the tests [5,10,18,19], with the exception of the tests reported by [8] with 3.8 wt%, and the pH was typically in the range from around 7 [5,19] to around 8 [8,17].

The corrosion grade scale applied for assessment of the surface condition of the used chains is described in Table 2. Examples of graded chain links are shown in Fig. 1. A shortcoming of this scale is the inevitable subjectivity involved, making its usage vulnerable to inconsistent categorization. Assessment of the samples used for the present study has therefore been performed by a small group of corrosion experts, to ensure that the corrosion grade scale is applied consistently across samples.[2]

Fig. 2 shows S–N plots of the test data included in the present study. The mean curve shown in each plot is obtained as the least-squares fit of Eq. (1) with the slope fixed at $m = 3$. Compared to the new chain data (Fig. 2(a)), which vary only in terms of

[1] This general rule is inadequate for the data from Vicinay, consisting of samples with a number of unique diameters. Subsets have therefore been defined based on material grade only for this data.

[2] Gabrielsen et al. [9] present preliminary results from ongoing work, aiming to develop computer algorithms that may be used to determine the corrosion grade based on 3-D scans of the chain links. More work is needed for this approach to prove its robustness. However; if successful, future assessment of surface condition may be based on objective algorithms, in lieu of the subjective assessment on which the current study relies.



(a) New chain. (b) Used chain.

Fig. 2. S–N data from fatigue tests for new chain (left) and used chain (right). Mean curve ($m = 3$) for data in each plot (dashed line) and DNV-OS-E301 design curve (solid line) are included for reference.

stress range and mean load, the used chain data (Fig. 2(b)) shows a larger horizontal scatter at each stress range level due to the additional variation of chain condition. For S–N plots distinguishing between mean load and corrosion grade levels for the used chain data, see [15].

It is worth noting that a few of the used chain samples have some degree of interlink wear and/or wear marks caused by fairlead contact or from the manufacturing process. A limited reduction of the cross section is observed for these samples. Furthermore, the fatigue test results have indicated that the observed wear has minor impact on the fatigue life obtained in the tests, particularly when compared to the effect of pitting corrosion [6]. A possible explanation for the limited importance of the observed interlink wear is that it may have two effects with opposite impact on fatigue life. It reduces the cross section, but also provides a better link-to-link fit which could be beneficial to the hotspot stresses [5].

2.2. Joint occurrences of mean load and corrosion grade

Since the fatigue test results will be used to estimate empirically the mean load and corrosion grade effects, it is of interest to examine the joint occurrences of these attributes within the test database. This is presented in Fig. 3. The combination of corrosion grade 1 and mean load 20% MBL is the most common (47 tests), since all ND tests and all but one test from TWI were performed at this mean load level. For used chain, a fairly large share of the tests are for corrosion grade 1–2 with a mean load of around 14% MBL (17 tests), and for corrosion grade 4–6 with a mean load around 16%–18% MBL (23 tests). It should further be noted that there are no tests with corrosion grade 2–4 combined with a mean load below 14% MBL, only two tests with a corrosion grade above 1 combined with a mean load of 20% MBL, and no tests with mean load above 20% MBL.

Compared to the data used for the study presented in [15], the database has been enriched with more tests at mean loads below 20% MBL. The Vicinay data [11,18] adds 19 tests for new chain (corrosion grade 1) with mean load in the range from 6.7% to 14.7% MBL, and the subset P1–R4–114e [9] adds 10 tests with corrosion grade 7 at mean loads 8% and 18% MBL (5 tests at each level).

2.3. Fracture locations

Data on fracture location for new versus used chain samples are presented by Gabrielsen et al. [6]. These are consistent with similar data presented by Mendoza et al. [16] and are representative also for the data set considered within the presented study. They show that for new chain, around 50% of the fractures occurred from cracks initiated at the inner bend of the chain links (referred to as “Kt point” in [16]), around 30% were initiated from the outside of the crown and the remaining fractures were on the straight part of the chain link. For used chain, more than 80% of the fractures occurred at the crown whereas less than 15% occurred at the bend. Furthermore, Gabrielsen et al. [6] also reported that the fracture locations depend on the surface condition of the used chain samples: for mildly corroded chain without corrosion pits (i.e., close to as-new surface), around 75% of fractures were at the crown and 25% were at the bend; and for chain links with corrosion pits, 95% of the fractures were at the crown and none occurred at the bend.

The more frequent fractures at the crown for used chain samples could be due to interlink wear, causing a reduced cross section, or due to fatigue crack initiation from corrosion pits at the crown [16]. However, as (i) the observed interlink wear is similarly represented among mildly and heavily corroded chain samples and appears to have a limited impact on fatigue life and (ii) the presence of corrosion pits has a considerable impact on fracture location, the most plausible explanation is that the crack initiation is driven by the surface condition at the crown.

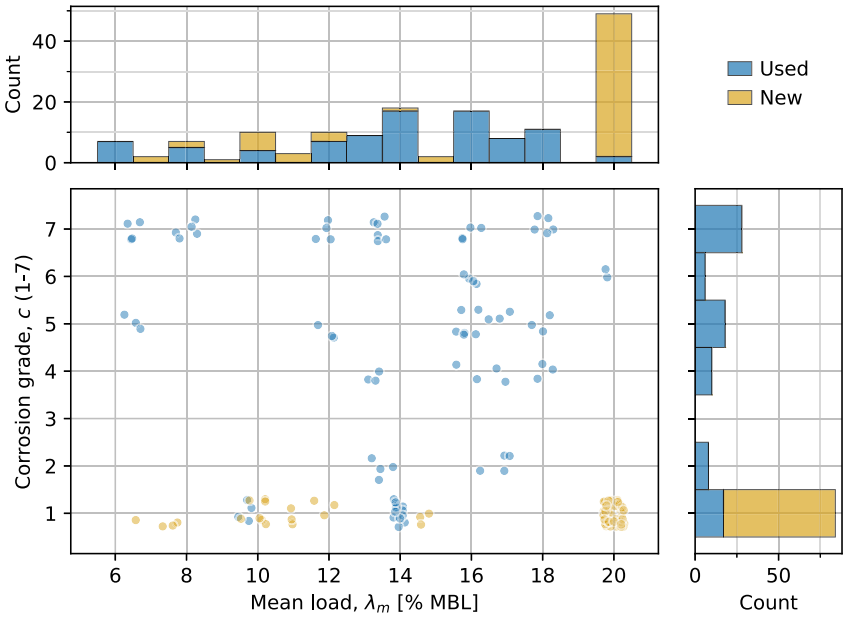


Fig. 3. Combinations of mean load (λ_m) and corrosion grade (c) in database, and marginal histograms. Scatter plot data has been jittered (up to ±0.3 in both directions).

3. Regression analysis with hierarchical model

3.1. Motivation for using a hierarchical model

Classical regression under the assumption of normal distributed errors may be expressed as

$$\begin{aligned} y_i &= \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i, \qquad \text{for } i = 1, \dots, n \\ \epsilon_i &\sim \mathcal{N}(0, \sigma_\epsilon^2) \end{aligned} \tag{3}$$

where y_i is the test outcome for the ith sample, β_0 is the intercept, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a column vector with p slope coefficients, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})$ is a row vector containing p predictors, and ϵ_i is the model error. This model implies that the errors are assumed to follow independent normal distributions, $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I})$, where σ_ϵ is referred to as the standard error of the model and $\boldsymbol{I}$ is the $n \times n$ identity matrix. However; if it is natural to sort the samples into groups (e.g., based on different size, material grade or other attributes that vary across subsets), the data may be clustered, with correlated errors. This is referred to as a hierarchical data structure, and it may then be incorrect to assume that the data points (and thus, the errors) are independent. When faced with hierarchical data, two extreme alternatives for the regression analysis are [20, Chapter 12]:

- *Complete pooling:* fit a single model according to (3), ignoring the hierarchical structure of the data. This may lead to underfitting, with biased estimates for the model coefficients and possibly poor predictive performance for new (unseen) observations.
- *No pooling:* fit separate models, applying the regression model in (3) to each group of data separately. This may lead to overfitting and high variance estimates for the coefficients, unless each group contains a sufficient amount of data (which is not the case in many practical problems). Furthermore, this approach (i) does not allow to enforce that groups may be sharing some of the parameters of the model (e.g., one or more of the slope coefficients), and (ii) is inadequate for predicting new observations in new (unseen) groups.

Regression analysis by use of a hierarchical linear model is referred to as *partial pooling*, and is a compromise between these extreme approaches. It combines the data from all groups, thereby maximizing the information extracted from the data, while seeking to avoid the issues of over- or underfitting the model. Also, the resulting model may be used for prediction of new observations from both existing and new groups.

It is worth noting that a less extreme version of the no-pooling approach is obtained by adding categorical predictors for each group to the complete-pooling model. In this case, the single regression model defined in (3) is adjusted as follows: the common intercept (β_0) is omitted, and one predictor is added per group with an indicator value that is either 1 (if the sample is a member of the corresponding group) or 0 (otherwise). The coefficients associated with these categorical predictors then represent group-specific

intercept parameters (instead of the common one, β_0). This allows enforcing shared slope coefficients across groups while estimating separate intercept parameters, but will still yield high-variance estimates and be inadequate for prediction of observations from new groups.

3.2. Hierarchical linear model

The focus for the present study is on the model

$$\begin{aligned} y_i &= \alpha_{j[i]} + \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, \sigma_\epsilon^2) \\ \alpha_j &\sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2) \end{aligned} \tag{4}$$

where $j \in (1, \dots, J)$ is a group index, $j[i]$ denotes group membership for the ith sample, and α_j is the intercept associated with group j. This type of model is called a varying-intercept model [20], and the model parameters are $(\alpha_1, \dots, \alpha_J, \beta_1, \dots, \beta_p, \sigma_\epsilon, \mu_\alpha, \sigma_\alpha)$. The error ϵ_i is not a model parameter. A particularity of this model is that both the group intercepts, α_j (data-level parameters), and the parameters μ_α and σ_α of the distribution of these intercepts (higher-level parameters), are parameters of the model: they are all estimated from the data in a maximum likelihood procedure, or have their distributions updated in Bayesian inference. Note that under this model, we regard each group intercept α_j as a realization from a common distribution, $\mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$. This has the effect of constraining them, so that they are pulled towards a common mean value (μ_α) compared to if they were estimated solely from the data within each group.

An equivalent way of writing the varying-intercept model in (4) is to substitute $\alpha_j = \mu_\alpha + \eta_j$, where $\eta_j \sim \mathcal{N}(0, \sigma_\alpha^2)$ represents the group-level error. In other words; compared to the classical regression model, the hierarchical linear model splits the model error into two terms such that σ_ϵ describes the standard deviation of data-level errors (i.e., within-group variations) and σ_α is the group-level standard error (i.e., between-group variations).[3] This implicitly allows for modeling of correlations within groups.

Appendix A briefly describes some aspects of the hierarchical linear model of relevance for the data analysis and interpretation of results presented subsequently: the standard error of estimated group intercepts, the intraclass correlation (ICC) and the predictive distribution for new (unseen) observations. For further details on hierarchical modeling, see e.g., [20], [22, Chapters 5 and 15] or [23, Chapter 9].

4. Data analysis

4.1. Assumptions and limitations

The following assumptions and limitations apply to the present study. The consequences of these assumptions are discussed in Section 6. (i) The study is limited to considering stress range effect fixed at $m = 3$. (ii) Prior fatigue loads and damage accumulated during service life are disregarded. (iii) Differences in number of links in the test rigs are neglected. All test results are assumed to represent failure of a single link. (iv) Nominal stress ranges are considered, rather than effective stress ranges accounting for possible reductions of the cross section. An implicit assumption is thus that the material loss due to general corrosion is negligible, and that the condition of the chain is fully represented by the corrosion grade.

4.2. Model

A varying-intercept model as defined in (4) is applied, with predictors for mean load and chain condition. The stress range would generally also be included as a predictor. However, since we have assumed that the stress range effect is fixed at $m = 3$, it is technically included in the outcome variable instead. The outcome and predictor variables are then:

$$\begin{aligned} y_i &= \log N_i + m \cdot \log S_i \\ x_{i1} &= g_1(\sigma_{m,i}) \\ x_{i2} &= g_2(c_i) \end{aligned} \tag{5}$$

where N is the number of cycles to failure at stress range S, and $g_1(\sigma_m)$ and $g_2(c)$ are functions of the mean stress (σ_m) and corrosion grade (c), respectively. The choice of these functions is based on findings from the previous study presented in [15], and is described in detail subsequently.

Model parameters are estimated by using the Python module `statsmodels` [24] and its `MixedLM` class, which is an implementation of the restricted maximum likelihood (REML) procedure described in [21].

[3] The model in (4) is sometimes called a linear mixed-effects model, with α_j (or η_j) referred to as random effects and $\boldsymbol{\beta}$ referred to as fixed effects. See e.g., [21].

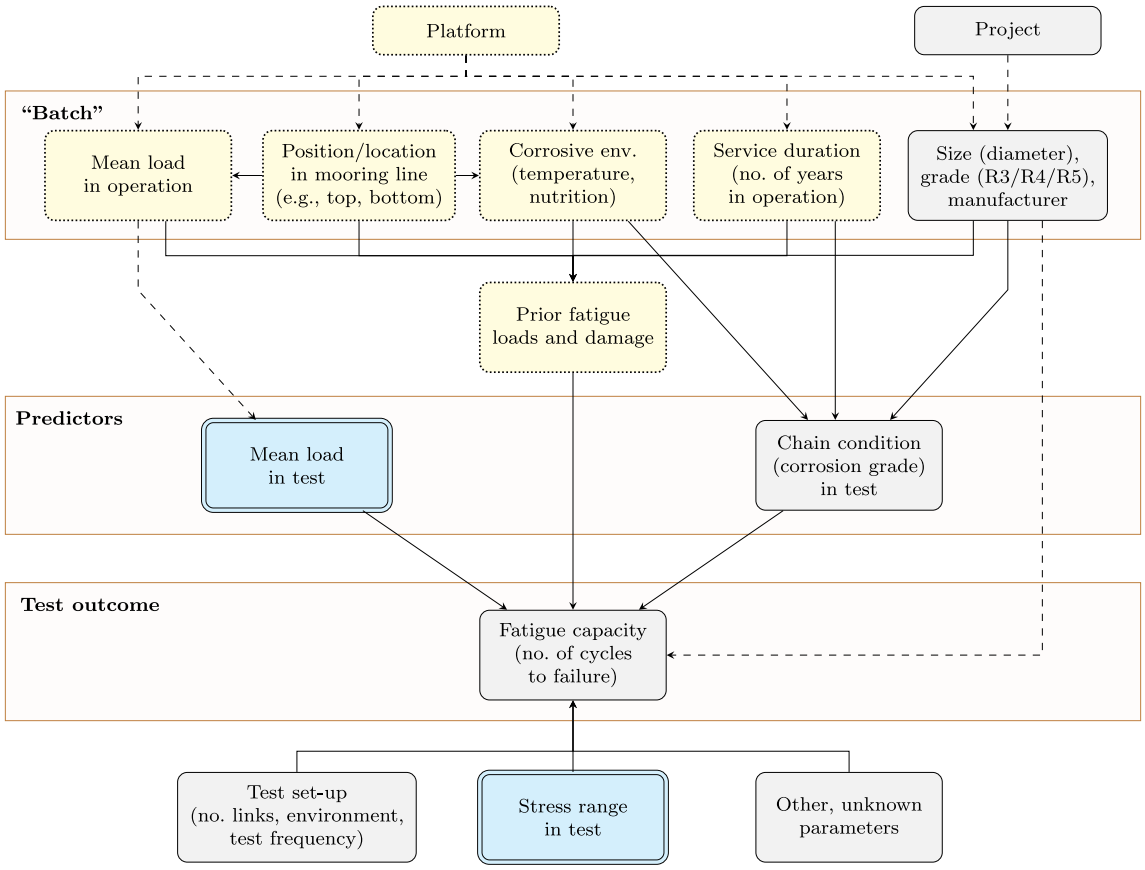


Fig. 4. Hierarchical structure of data. Grey box: relevant parameter for all tests. Yellow box (dashed border): relevance for used chains only. Blue box (double border): controlled variable in test. Dashed arrow: indirect and/or presumably weak relationship.

4.3. Grouping criterion for the hierarchical model

An essential step in setting up the hierarchical model is the choice of a grouping criterion. That is, the attribute or property, or set of such, by which we divide the data into groups $j \in (1, \ldots, J)$. This choice may have a large influence on the estimated model parameters and therefore on the predictive performance of the model. The grouping must therefore be done as wisely as possible. To describe how the various attributes and properties associated with each sample may influence clustering of the data, we will distinguish between *controlled*, *non-controlled* and *hidden* variables: (i) A *controlled* variable may be assigned a value of choice in connection with the experimental design of the tests, to make it independent of other variables affecting the outcome (e.g., the stress range applied in a fatigue test); (ii) A *non-controlled* variable may be measured, but not controlled (e.g., the corrosion grade of a sample); (iii) A *hidden* variable is any attribute or property that affects the outcome of the test, directly or indirectly, but is not explicitly included in the regression model (e.g., the material grade or the prior fatigue loads of a sample).

Firstly; hidden variables that influence simultaneously both non-controlled predictors and the test outcome may cause correlations within subsets of the data, resulting in biased estimates for the effects of these predictors if not accounted for. Secondly; in practice, the values of the controlled variables are not necessarily as independent of the hidden variables as one would like them to be, in particular for experimental data as costly as full scale fatigue tests. Hence, ideally, we want any hidden variable to have the same value within each group. A dilemma is, however, that a too strict grouping criterion may result in many groups with few data points each, causing high-variance estimates for the model parameters. Nevertheless, a fairly large number of groups containing few observations is fully acceptable for the hierarchical model [20, p. 276]. In general, increasing the number of samples per group should therefore not be a goal in itself, at the cost of defining a less suitable grouping criterion. This choice should therefore be based on a careful evaluation of available data and the process behind them.

The hierarchical structure of the data included in the present study is illustrated in Fig. 4. In this diagram, boxes represent attributes and properties of the sample or the test, whereas arrows indicate direct or indirect relationships. As indicated in the figure, stress range applied in the test is a controlled variable. In principle, so is the mean load applied in the tests. This is largely the case for the new chain tests, and for the latest tests for platform P1 (batch P1-R4-114e, see Table 1). The 10 samples within this batch, providing test results for mean loads 8% and 18% MBL, are all retrieved from the same mooring chain segment (that is, the

same mooring line and roughly the same location along the line). This ensures that variation in mean load across these samples is made largely independent of hidden variables, thereby limiting their adverse influence to the test outcome within that batch. For most of the other used chain batches, however, the tests have been conducted for a single mean load or over a small interval of mean loads (with the exception of P5-R4-130, tested at 6.4% and 20% MBL). Moreover, the mean loads applied for these batches have been selected partly based on mean loads experienced during operation. Despite being a controlled variable in principle, the mean loads applied in these tests are therefore influenced to some degree by a hidden variable that also affects the test outcome indirectly through its impact on the prior fatigue loads and damage. Grouping the tests across these batches may then cause clustering within the groups, possibly resulting in biased estimates for the mean load effect.

Corrosion grade is inevitably a non-controlled variable. It cannot be assigned a value of choice to obtain an optimum experimental design, and is largely influenced by hidden variables such as service duration, position in mooring line (which affects the corrosive environment) and possibly also the material grade and the base material chemistry used by the manufacturer of that particular sample [25]. Moreover; (i) all the hidden variables shown in the second level of Fig. 4 influence the fatigue loads and damage accumulated during operation; and (ii) although previous studies for new chain have concluded that there is no significant effect of size and material grade [10,11,17], we cannot presume *a priori* that this is the case also for corroded chain.

On this basis, the "batch" criterion defined in relation to Table 1 appears as a reasonable grouping criterion for used chain. This is not to suggest that the prior fatigue loads and damage are the same for used chain within each batch. Rather, division by batch ensures that relevant hidden variables have roughly the same value within each group. Note also that grouping purely by platform may be a poor choice: although the hidden variables for each sample are certainly influenced by the platform at which they were operated, their values may vary largely for samples originating from the same platform. As an example, two chain samples retrieved from the same platform may be from different manufacturers or of different material grades, and may have been exposed to very different mean loads levels or service durations depending on the chain diameters and positions within the mooring line.

Batch is applied as the grouping criterion also for new chain, enabling investigation of group effects related to size and material grade for these samples as well — particularly in relation to different mean load predictors. Note that in doing so, we retain the slightly inconsistent batch definition for the Vicinay data (i.e., based on grade only, ignoring size differences). This is believed to be an acceptable simplification, as new chain samples are presumably less affected by hidden variables compared to used chain.

4.4. S–N intercept and design curve

Combining the S–N relation in (1), the predictor variables in (5) and the predictive distribution in Appendix A, the S–N intercept for a sample from a new group has mean value

$$\log A(\sigma_m, c) = \mu_\alpha + \beta_1 \cdot g_1(\sigma_m) + \beta_2 \cdot g_2(c) \tag{6}$$

and standard deviation $\sigma = \sqrt{\sigma_\alpha^2 + \sigma_\epsilon^2}$. This formulation implies that β_1 and β_2 quantify the effects on fatigue capacity from mean load and corrosion grade, respectively, and is consistent with that used in [15] as given in (2).

According to DNV-OS-E301 [3], the *design curve* intercept should be associated with 97.72% probability of exceedance. Under the assumption of normal distributed model errors, this is obtained by subtracting two standard deviations if the model coefficients and σ are known. When $\mu_\alpha, \beta_1, \beta_2$ and σ are instead estimated from data, the inferential uncertainty (also referred to as epistemic uncertainty) of the estimated coefficients requires a larger offset of the design curve. This may be expressed as

$$\log A_D(\sigma_m, c) = \hat{\mu}_\alpha + \hat{\beta}_1 \cdot g_1(\sigma_m) + \hat{\beta}_2 \cdot g_2(c) - k_n \cdot \hat{\sigma}, \qquad k_n \geq 2 \tag{7}$$

where subscript $\cdot_D$ denotes design curve value, and superscript $\hat{\cdot}$ implies that an estimate of the parameter is used. The offset k_n depends on the fatigue test sample size (n), and in principle it may also depend on the mean load and corrosion grade level for which the intercept is calculated. The latter dependency is here neglected for simplicity, and the design curve offset (k_n) is calculated from the empirical function given in [3, Ch. 2, Sec. 2, clauses 6.6.3 and 6.6.4]. A convenient representation of the design curve intercept is obtained by defining a *design* value for μ_α:

$$\hat{\mu}_{\alpha,D} = \hat{\mu}_\alpha - k_n \cdot \hat{\sigma} \tag{8}$$

The design curve intercept is then obtained by substituting $(\hat{\mu}_{\alpha,D}, \hat{\beta}_1, \hat{\beta}_2)$ into (6).

4.5. Model variations

A summary of the base case considered for the data analysis is presented in Table 3, along with model variations that are applied one at a time. The basis for addressing these particular variations is briefly described here.

A partial pooling model is the base case. Using batch as the grouping criterion gives $J = 23$ groups with group sizes ranging from 3 to 14 samples per group at an average of $154/23 = 6.7$. For data as scattered as fatigue tests, in particular when exposed to varying mean loads and chain conditions, this amount of data per group may intuitively appear insufficient and associated with a risk of overfitting. To get a sense of how the model performs in this regard besides the measures described previously, and how partial pooling affects the estimated slope coefficients, we compare to results from complete-pooling and no-pooling models. The complete-pooling estimates are obtained from classical least-squares regression with all data in a single group. The no-pooling estimates are computed by adding categorical variables for groups to the complete-pooling model. This no-pooling implementation

Table 3
Overview of base case and model variations.

Model and data	Base case	Variations
Pooling:	Partial pooling	Complete-pooling, no-pooling[a]
Predictors:	$g_1(\sigma_m) = \lambda_m$ [% MBL] $g_2(c) = c$	$g_1(\sigma_m) = \sigma_m$ [MPa] $g_2(c) = \log c$
Data set:	All data	New chain, used chain
Grouping criterion:	Batch	–

[a]Unpooled estimates for group intercepts (α_j), and completely pooled estimates for slope coefficients ($\boldsymbol{\beta}$).

Table 4
Results obtained from REML. Units are associated with stress ranges in MPa.

Variation:	Pooling			Predictors		New vs. used chain	
Description:	Partial[a]	Complete	None[b]	Mean load	Corrosion	New	Used
Model no.:	1	2	3	4	5	6	7
Pooling	Partial	Complete	None[b]	Partial	Partial	Partial	Partial
Model parameters							
n	154	154	154	154	154	67	87
J	23	1	23	23	23	8	15
$g_1(\sigma_m)$	λ_m	λ_m	λ_m	σ_m	λ_m	λ_m	λ_m
$g_2(c)$	c	c	c	c	$\log c$	–	c
Coefficient estimates and associated standard errors, se(·)							
$\hat{\mu}_\alpha$	12.236	12.270	12.177[c]	12.129[d]	12.154[d]	12.150	12.231
$\hat{\beta}_1$	−0.0514	−0.0518	−0.0510	−0.0072[d]	−0.0508	−0.0503	−0.0547
$\hat{\beta}_2$	−0.0977	−0.1052	−0.0783	−0.1003	−0.6616[d]	–	−0.0893
se($\hat{\mu}_\alpha$)	0.0811	0.0601	–	0.0851	0.0815	0.1245	0.1034
se($\hat{\beta}_1$)	0.0044	0.0033	0.0059	0.0007[d]	0.0046	0.0068	0.0060
se($\hat{\beta}_2$)	0.0093	0.0058	0.0200	0.0111	0.0686	–	0.0119
Model errors and intraclass correlation (ρ)							
$\hat{\sigma}_\epsilon$	0.144	0.168	0.144	0.145	0.145	0.142	0.146
$\hat{\sigma}_\alpha$	0.092	–	0.116	0.123	0.104	0.079	0.099
$\hat{\sigma}$	0.171	0.168	0.185	0.190	0.178	0.162	0.176
$\hat{\rho}$	0.29	–	–	0.42	0.34	0.24	0.32
S–N intercept and fatigue life ratio at reference level ($\lambda_m = 20,\ c = 1$)							
$\log A_{\text{ref}}$	11.111	11.129	11.079	–	11.138	11.144	11.048
N_{ratio}	1.00	1.04	0.93	–	1.06	1.08	0.87
Design curve values							
k_n	2.02	2.02	2.02	2.02	2.02	2.05	2.04
$\hat{\mu}_{\alpha,D}$	11.891	11.931	–	11.744	11.794	11.817	11.872
$\log A_{D,\text{ref}}$	10.766	10.790	–	–	10.778	10.811	10.689
$N_{D,\text{ratio}}$	1.00	1.06	–	–	1.03	1.11	0.84

[a]Base case.
[b]No-pooling model with unpooled estimates for group intercepts and pooled estimates for slope coefficients.
[c]For no-pooling model, μ_α is here calculated as weighted mean of group intercepts (weighted by no. of samples per group).
[d]Not directly comparable to base case estimate due to differing predictor unit(s).

corresponds to the aforementioned, less extreme version of a no-pooling model as described in Section 3.1: it does not result in 23 separate models, but a single model with unpooled estimates for group intercepts and completely pooled estimates for the slope coefficients — consistent with the varying-intercept model applied for partial pooling.

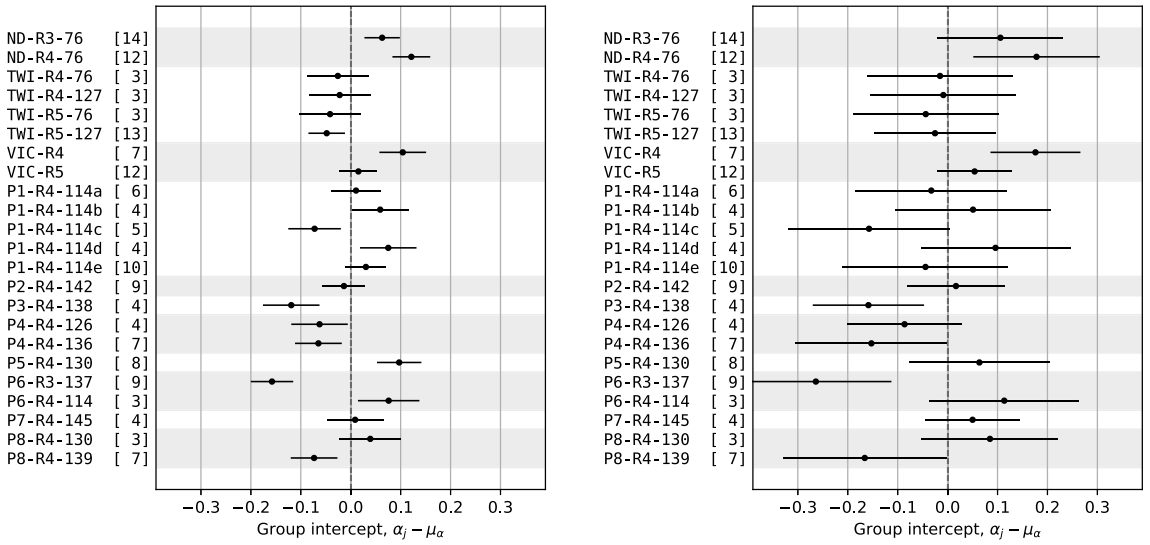
Base case predictors, $g_1(\sigma_m) = \lambda_m$ [% MBL] and $g_2(c) = c$, are defined based on findings from the previous study presented in [15]. Variation of predictors is here limited to consideration of one alternative function for mean load and corrosion grade effects, respectively. In [15], these alternative predictors did not perform as well as those defined for the base case, however; it is of interest to reassess them in light of the hierarchical model and the extended data set.

The model applied assumes implicitly that the effect of mean load is independent of the chain condition. To assess the validity of this assumption, we include partial pooling models considering new and used chain samples separately. These results will also serve to quantify differences in fatigue performance and between-group variability for new versus used chain.

5. Results

5.1. Main results

Results for all model variations considered are listed in Table 4, and relate to stress ranges with unit MPa. From Eq. (6) it is seen that the estimated mean intercept, $\hat{\mu}_\alpha$, corresponds to the S–N intercept, $\log A(\cdot)$, for the combination of $g_1(\sigma_m) = 0$ and $g_2(c) = 0$. With the base case predictors this occurs for zero mean load, $\lambda_m = 0$, which is irrelevant for operation of mooring chain, and the



(a) Partial pooling, $\hat{\mu}_\alpha = 12.236$.

(b) No pooling, $\hat{\mu}_\alpha = 12.177$.

Fig. 5. Group intercepts for partial pooling and no-pooling models with base case predictors; $g_1(\sigma_m) = \lambda_m$, $g_2(c) = c$. Error bars correspond to plus/minus one standard error of group intercepts, se($\hat{\alpha}_j$). Number of tests (n_j) given in brackets. Alternating background shading is used to distinguish between platforms. Note that difference in average intercept (μ_α) between models is not revealed by the figure.

fictitious corrosion grade $c = 0$. Direct comparison of $\hat{\mu}_\alpha$ across models is thus of limited value, even when identical predictors are used. A more convenient reference level is the combination of $\lambda_m = 20$ and $c = 1$. That is, new (or as-new) chain operated at a mean load of 20% MBL, which is consistent with the current design curves given in standards. The S–N intercept for this reference level is denoted A_{ref} in Table 4.

Another useful measure for comparison across models is the fatigue life ratio, defined as $N_{\text{ratio}} = 10^{\log A_{\text{ref}} - \log A_{\text{ref}}^{(\text{basecase})}}$, where $A_{\text{ref}}^{(\text{basecase})}$ is the reference value calculated from the base case model. Hence, N_{ratio} expresses the estimated fatigue capacity for a given model, at the reference level, relative to the base case.

Design curve values are also given in the table. These serve two main purposes: (i) they enable comparison across models accounting for differences in model errors, and (ii) the design curve intercept for the reference level, $\log A_{D,\text{ref}}$, is directly comparable to the S–N design intercept for studless chain given in DNV-OS-E301 [3] ($\log A_D = 10.778$). As an example: at the reference level ($\lambda_m = 20$, $c = 1$), the base case model ($\log A_{D,\text{ref}} = 10.766$) gives a fatigue design capacity which is 3% on the low side of the DNV-OS-E301 design curve (fatigue life ratio = 0.97) whereas the complete-pooling model ($\log A_{D,\text{ref}} = 10.790$) is 3% on the high side.

In Appendix B, the parameters fitted by REML for models 1, 4 and 5 are verified through comparison to Bayesian posterior modes obtained from Markov chain Monte Carlo (MCMC) sampling. The appendix also includes a comparison of the predictive accuracy of these models by means of leave-one-out (LOO) cross-validation, to support the discussion on alternative predictors in the capacity model.

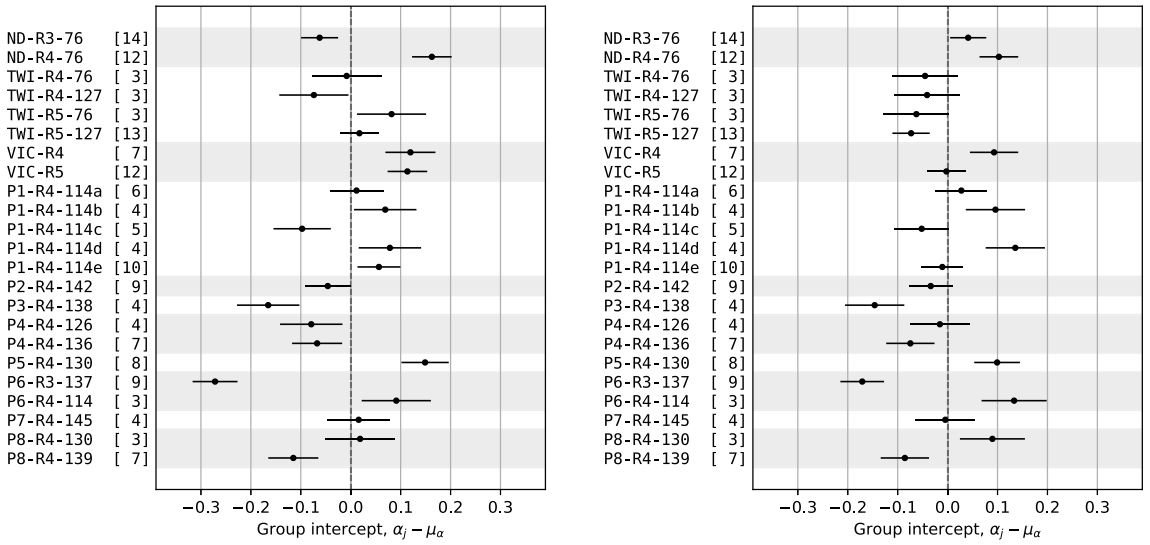
5.2. Group intercepts

Group intercepts and associated standard errors for the partial pooling and no-pooling models are presented in Fig. 5. The partial pooling model (Fig. 5(a)) shows limited variations for the intercepts of new chain groups, particularly between groups originating from the same projects (ND, TWI and VIC, respectively). The used chain groups exhibit a slightly larger between-group variation, even for the P1 groups containing samples of the same size and material grade. One group stands out in particular: P6-R3-137, which is the only group with used R3 chain. The group intercept for this group is the one that is shifted the most to the left in Fig. 5(a), meaning that it is the group that underperforms the most compared to what the model predicts on average.

Similar variations are seen for the unpooled estimates (Fig. 5(b)), but with more pronounced deviations from the average intercept and with considerably larger inferential uncertainty. Corresponding visualizations of group intercepts are shown in Fig. 6 for the models with alternative predictors.

5.3. New vs. used chain

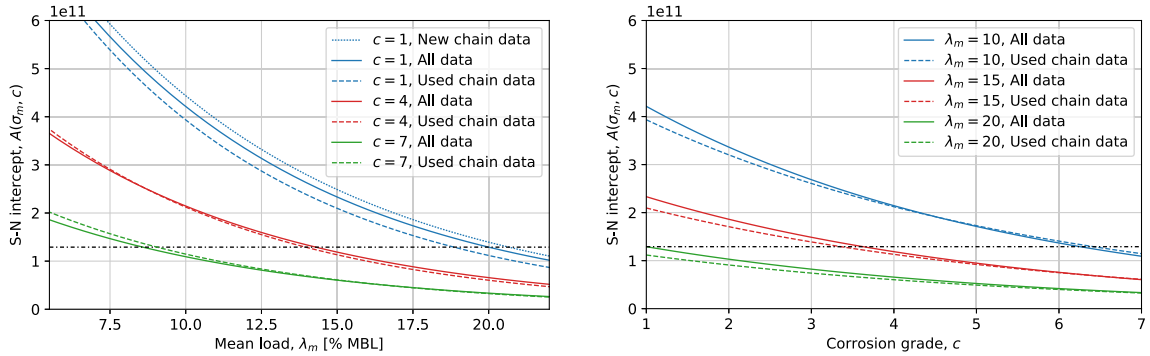
From comparison to the base case analysis ($\hat{\beta}_1 = -0.0514$, $\hat{\beta}_2 = -0.0977$, $\hat{\sigma}_\epsilon = 0.144$, $\hat{\sigma}_\alpha = 0.092$), the following initial observations are made for the results of applying the partial pooling model for analysis of new chain and used chain data separately (models



(a) Mean load predictor $g_1(\sigma_m) = \sigma_m$ [MPa] (model 4).

(b) Corrosion grade predictor $g_2(c) = \log c$ (model 5).

Fig. 6. Group intercepts for partial pooling model with alternative predictors. See further explanations in caption of Fig. 5.



(a) Mean load dependency, for $c \in \{1, 4, 7\}$.

(b) Corrosion grade dependency, for $\lambda_m \in \{10, 15, 20\}$.

Fig. 7. S–N intercepts based on different data sets: base case (all data), new chain only and used chain only. The horizontal dash-dotted line is located at the reference level $\lambda_m = 20$ [% MBL] and $c = 1$ for the base case model.

6 and 7 in Table 4, respectively). *For new chain data:* there is good agreement on the mean load effect ($\hat{\beta}_1 = -0.0503$), the model error is slightly reduced ($\hat{\sigma}_\epsilon = 0.142$), the between-group variation is reduced ($\hat{\sigma}_\alpha = 0.079$), and the predicted fatigue capacity for the reference mean load is a little less than 10% on the high side ($N_{\text{ratio}} = 1.08$). *For used chain data:* the model estimates a higher mean load effect ($\hat{\beta}_1 = -0.0547$) and a lower corrosion grade effect ($\hat{\beta}_2 = -0.0893$), the model error is slightly increased ($\hat{\sigma}_\epsilon = 0.146$), the between-group variation is increased ($\hat{\sigma}_\alpha = 0.099$), and the predicted reference fatigue capacity is roughly 15% on the low side ($N_{\text{ratio}} = 0.87$).

The difference in estimated fatigue capacity for new or as-new chain ($c = 1$) at mean load $\lambda_m = 20$ [% MBL] is thus around 25%, when predicted based on new chain data compared to used chain data, and the analysis that is based on all data yields a compromise between these two results. This is visualized in Fig. 7, showing S–N intercepts as functions of mean load and corrosion grade. Note that the vertical axis of this figure shows $A(\sigma_m, c) = 10^{\log A(\sigma_m, c)}$, as the predicted fatigue life is proportional to the S–N intercept on this scale. At $c = 1$, the deviations between the curves are fairly consistent across mean loads, although the difference in capacity estimated from new versus used chain data is reduced to around 13% at mean load $\lambda_m = 10$, due to the slightly higher mean load effect for the latter. Since the new chain data contains no information for $c > 1$, the S–N intercept based on all data approaches that based on used chain data for increasing corrosion grades. For $c \geq 3$, these curves agree very well (Fig. 7(b)), indicating that the differences in estimated mean load and corrosion effects cancel out.

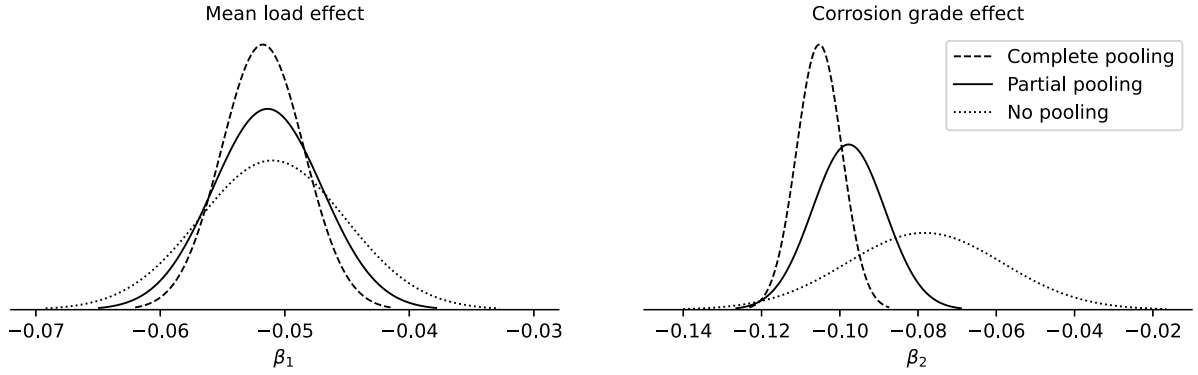


Fig. 8. Comparison of slope coefficients estimated from complete-pooling, partial pooling and no-pooling. Probability density function of $\beta \sim \mathcal{N}(\hat{\beta}, \mathrm{se}(\hat{\beta})^2)$ where $\hat{\beta}$ is estimated coefficient and $\mathrm{se}(\cdot)$ is the standard error. Base case predictors: $g_1(\sigma_m) = \lambda_m$ and $g_2(c) = c$.

6. Discussion

6.1. Grouping criterion

We consider the base case model with partial pooling and base case predictors (model 1 in Table 4). The ICC is $\hat{\rho} = 0.29$ and the ratio of within-group to between-group variances is $\sigma_\epsilon^2/\sigma_\alpha^2 \approx 2.4$. This implies that the group-level model (the common distribution for α_j) is moderately important for the current data set, carrying the same amount of information as 2–3 samples within groups. Hence, the intercepts for the smallest groups ($n_j = 3$) are approximately halfway between $\hat{\mu}_\alpha$ and the unpooled estimate. For groups with more data, group-specific information is emphasized more than the overall average, increasingly so with increasing group size.

Consider now the group intercepts in Fig. 5. The partial pooling model (Fig. 5(a)) clearly constrains the intercepts and limits the estimation variance compared to the unpooled estimates (Fig. 5(b)). The limited variation between new chain groups originating from the same projects (ND, TWI and VIC, respectively) suggests that the new chain samples could have been collected into three groups, representing each project. However, this does not disqualify the current grouping criterion. The larger between-group variation for used chain groups (P1–P8) supports the choice of batch as the grouping criterion, and weighs against a coarser subset division with larger group sizes.

Fig. 8 visualizes the estimated slope coefficients and associated uncertainty across the partial, complete- and no-pooling models. All three models agree well on the mean load effect (β_1), except for increasing uncertainty as the degree of pooling is reduced. A plausible explanation for the good agreement is that much information about the mean load effect is contained within some few groups with samples tested at both high and low mean loads (cf. Table 3). This is not the case for chain condition, which varies mostly *between* groups and less *within* groups, leading to larger deviations in the estimated corrosion grade effect (β_2). The partial pooling estimate appears as a compromise between the complete- and no-pooling estimates, and slightly closer to the complete-pooling model, indicating that it is able to extract information from chain conditions across groups without overfitting the model.

In conclusion, the partial pooling model performs well for the current data set and grouping criterion, with limited risk of overfitting, despite the large number of groups and the small average group size.

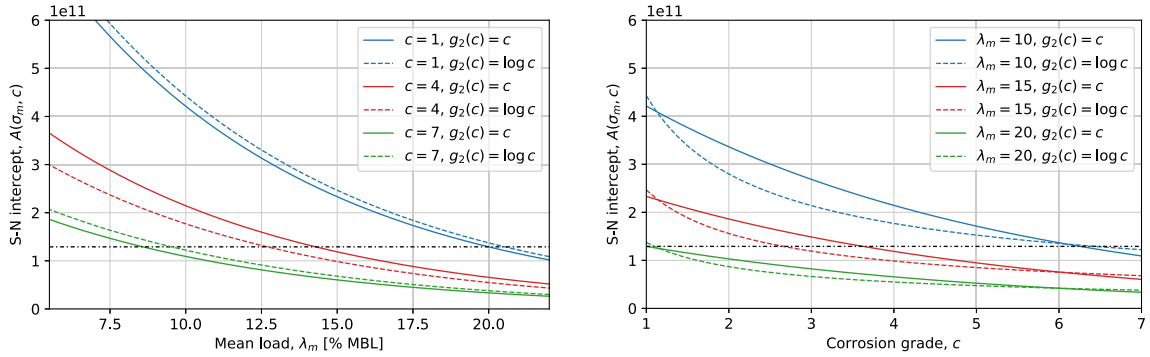
6.2. Alternative predictors

Focus is now shifted to the alternative mean load and corrosion grade predictors (models 4 and 5 in Table 4). Compared to the base case, both models perform equally well with regards to standard deviation of residuals ($\hat{\sigma}_\epsilon$) but estimate a higher group-level error ($\hat{\sigma}_\alpha$). This implies that they are able to describe within-group variations of the data well, but perform slightly worse in describing between-group variations. This results in a larger spread for the group intercepts, shown in Fig. 6.

6.2.1. Mean load predictor

The alternative mean load predictor, $g_1(\sigma_m) = \sigma_m$ [MPa], differs from using λ_m [% MBL] as follows. Consider two samples of the same diameter but of different material grades, that are tested at the same mean load when expressed in percentage of MBL (i.e., same value of λ_m). The sample with the higher material grade is then tested at a higher nominal mean stress (σ_m) due to a higher MBL. Specifically, the nominal mean stress will be around 23% higher for R4 compared to R3 chain, and around 17% higher for R5 compared to R4 chain [26]. In other words; if σ_m were representative as a predictor for the mean load effect, we would expect the lower material grade to perform better than the higher grades when tested at the same mean load level in percentage of MBL.

In practice, by comparison to group intercepts for the base case model with $g_1(\sigma_m) = \lambda_m$ (Fig. 5(a)), the effect of $g_1(\sigma_m) = \sigma_m$ (Fig. 6(a)) is that the intercepts for R3 groups are shifted left compared to R4 groups, and intercepts for R5 groups are shifted right. As a result, the ND and TWI group intercepts now exhibit a distinct spread. Considering that these groups contain fairly



(a) Mean load dependency, for $c \in \{1, 4, 7\}$.

(b) Corrosion grade dependency, for $\lambda_m \in \{10, 15, 20\}$.

Fig. 9. S–N intercepts for $g_2(c) = c$ (base case, solid lines) and $g_2(c) = \log c$ (model 5, dashed lines). The horizontal dash-dotted line is located at the reference level $\lambda_m = 20$ [% MBL] and $c = 1$ for the base case model.

homogeneous samples, all but one tested at a mean load of 20% MBL (the one exception being a TWI test at 10% MBL), this larger spread weighs against using nominal mean stress as the mean load predictor. Note also that (i) the one used chain group with R3 chain (P6-R3-137) is shifted left and now deviates even more from the average intercept, and (ii) the intercepts for VIC groups are actually shifted to agree very well for this model. Strictly, the latter observation favors the use of nominal mean stress, but it does not outweigh the larger spread for ND and TWI tests and the overall increase in between-group variation ($\hat{\sigma}_\alpha$). On this basis, the base case model with mean load predictor $g_1(\sigma_m) = \lambda_m$ is thus considered more adequate. This conclusion is supported by the model comparison presented in Appendix B.

A plausible explanation for the base case predictor's ability to better describe the mean load dependency are differences in proof load across material grades. For studless chain, the required proof load is 70% MBL [26], meaning that higher proof loads are applied to higher material grades. The proof load induces compressive residual stresses that are beneficial to the fatigue performance of the chain, increasingly so for increasing proof load [27]. A further discussion on residual stresses in mooring chain links and their impact of fatigue life is presented in [28].

6.2.2. Corrosion grade predictor

Two main observations are made by comparison of the model with alternative corrosion grade predictor, $g_2(c) = \log c$, to the base case model. Firstly, the alternative corrosion predictor has minor influence on the estimated mean load effect, both in terms of $\hat{\beta}_1$ and its estimated standard error (see model 1 vs. 5 in Table 4). This is consistent with the results discussed previously, indicating that much information about the mean load effect is contained within groups with no or limited variation of corrosion grade. Secondly, the between-group variation is only slightly increased, from $\hat{\sigma}_\alpha = 0.092$ to $\hat{\sigma}_\alpha = 0.104$, making it more difficult to observe from visual inspection of group intercepts (Fig. 6(b), compare to base case in Fig. 5(a)). It is further noted that the estimated S–N intercept at the reference level ($\log A_{\text{ref}}$) is higher for this model, with a fatigue life ratio of 1.06 compared to the base case. However; the $\log c$ function describes a more rapid degradation of the fatigue capacity when the corrosion grade increases from $c = 1$, as illustrated in Fig. 9. The difference is largest around $c = 3$ (Fig. 9(b)), with the base case model roughly 25% on the high side, then diminishes for increasing values of c. Since the model is purely empirical, however, it provides no basis for choosing one function over the other besides the difference in between-group variation.

In conclusion, these results support the base case model with $g_2(c) = c$ over the alternative predictor $g_2(c) = \log c$. One could argue that the latter option makes a conservative choice for most of the corrosion grade scale, but overall it has less support in the data set. Again, this conclusion is supported by the model comparison in Appendix B.

6.3. Mean load and corrosion grade effects

The S–N intercept in Eq. (6) with base case predictors implies $A(\sigma_m, c) \propto 10^{\beta_1 \cdot \lambda_m} \cdot 10^{\beta_2 \cdot c}$. The base case model thus predicts an increase in fatigue capacity by more than 30% for a mean load reduction of 2.5 [% MBL] ($10^{-0.0514 \cdot (-2.5)} = 1.34$), and by around 80% for a mean load reduction of 5.0 [% MBL] ($10^{-0.0514 \cdot (-5.0)} = 1.81$). This a substantial effect, considering that shifting the intercept by one standard deviation corresponds to an increase in capacity by around 50% ($10^{0.171} = 1.48$).

Similarly, the model predicts that increasing the corrosion grade by one unit reduces the fatigue capacity by 20% ($10^{-0.0977 \cdot 1} = 0.80$). This can be interpreted in terms of the classical approach of scaling stress ranges by the reduction in cross sectional area as follows. For a given tension, the effective stress range is inversely proportional to the effective area, implying $S \propto d_{\text{eff}}^{-2}$ where d_{eff} is the effective diameter after accounting for material loss due to general corrosion. From the S–N relation in Eq. (1), we then get $N \propto d_{\text{eff}}^{2m}$ and ultimately $d_{\text{eff}} \propto 10^{\beta_2 \cdot c/2m} = 10^{-0.0977 \cdot c/6}$. Hence, each increase by one unit on the corrosion grade scale would require a reduction of the effective diameter by around 4% if the classical approach were used instead. As an example: for a 114 mm chain

this corresponds to a diameter reduction of 12 mm for $c = 4$ or 23 mm for $c = 7$, which are both significantly more than the material loss observed for any of the used chain samples in the test database. It is thus concluded that the effect of pitting corrosion and surface condition, as described by the corrosion grade, is much larger than that which may be attributed to the general corrosion and material loss represented within the present data set.

A possible criticism of the model is the lack of an interaction term between mean load and corrosion grade effects. The mean load effect is thus assumed to be independent of chain condition and prior fatigue loads. However, the results reveal minor differences in the mean load effect estimated from new versus used chain data (Fig. 7). By further noting that (i) the used chain samples cover a large range of mean loads for $c = 7$ (cf. Fig. 3) and (ii) the used chain groups with the largest within-group variation of mean load contain heavily corroded chains (P1-R4-114e and P5-R4-130, see Table 1), it is concluded that the current data set does not support the need for a more complex model that includes such interaction.

6.4. New chain: effects of chain size, material grade and test setup

The distinct clustering of intercepts for, respectively, ND and TWI groups in Fig. 5(a) supports the previous reports on insignificant effects of size (diameter) and material grade for new chain. Both projects include a group of 76 mm R4 chain, so a reasonable explanation for deviations between ND and TWI tests is differences in test setup. One such difference is the number of links included in each test. Recall that ND used 7 links, whereas TWI used 11 links for 76 mm chain and 7 links for 127 mm chain. As the weakest out of 11 links is likely to fail before the weakest out of 7 links, there could be a minor size effect that is disguised by a difference in number of links. On the other hand, the deviations could be due to other differences related to the execution of the tests. For example: the TWI tests were conducted with a load cycle frequency between 0.2 and 0.5 Hz whereas the ND tests were mostly conducted with frequency 0.7 Hz, and previous studies have indicated that a higher load frequency leads to longer fatigue lives in the tests [10,17]. In any case, the deviations are practically insignificant in this context, and no conclusion can be drawn based on limited group sizes.

Slightly larger variation is observed between the VIC groups, containing tests performed at a range of mean loads. This does not necessarily suggest a difference in the mean load effect for R4 compared to R5 chain, because (i) deviations are still small, and (ii) the VIC tests were performed for a range of diameters, with a varying number of links and with different load frequencies [18].

6.5. Used chain

6.5.1. General corrosion and material grade

Batch P6-R3-137 stands out as the group with the largest material loss due to general corrosion and as the only used chain group of R3 grade. It is also the group that performs worst in comparison to the average model predictions (Fig. 5(a)): compared to the overall average ($\hat{\mu}_\alpha$), the group intercept for P6-R3-137 corresponds to a fatigue life that is around 30% lower based on the partial pooling model and almost 50% lower for the no-pooling model.

With a nominal diameter of 137 mm, the average material loss for this group (4 mm) corresponds to an increase in the effective stress range of 6% and a reduction of the fatigue life of roughly 20%. The use of nominal stress ranges for the data analysis, neglecting the effect of material loss, may thus partly explain the results for this group. Possible explanations for the additional deviation include: (i) Larger prior fatigue damage for this group compared to other groups (in this case it is not due to the long service life of 18 years alone, as there are several used chain groups with service life 18–20 years that performed better in the tests); (ii) The corrosion grade scale is not applied consistently for this particular group; (iii) The corrosion effect is different for R3 chain compared to R4 chain; (iv) General corrosion and material loss may affect the outer layer of compressive residual stresses and therefore influence the mean load effect; (v) It is by chance. However, until more used chain samples of R3 grade and/or with general corrosion are included there is no basis for distinguishing between these possible contributions to the observed deviation.

6.5.2. Prior fatigue loads and deviations from new chain results

A possible explanation for the deviations between new and used chain data, as seen from Fig. 7, is that the latter are influenced by prior fatigue loads and damage. The ratio of the allowable fatigue capacity in design to the estimated median capacity is $A_D/A = 10^{-k_n \cdot \hat{\sigma}}$, which for the base case model is around 0.45. In addition, safety factors in the range from 5 to 8 are required on estimated fatigue damage over the service life of the chain [3], reducing the allowable utilization to 6%–9% of the median capacity. Furthermore, the majority of the used chain samples included here were designed to a safety factor of 10 due to stricter company requirements, and retrieved before the end of design life was reached, further reducing the presumed utilization to below 5%. There are, however, other factors that influence the actual utilization of the chain prior to the fatigue tests:

- Actual fatigue loads that the chains were exposed to during the service life. These may be higher or lower than those estimated in design, depending on the accuracy of the numerical models and computational tools and the amount of conservatism (or, lack of such) included in the design analyses.
- Mean load during operation. A higher or lower mean load during service than in the subsequent test will influence the test result. For samples that were tested at mean loads similar to those experienced during operation, however, this effect should cancel out.

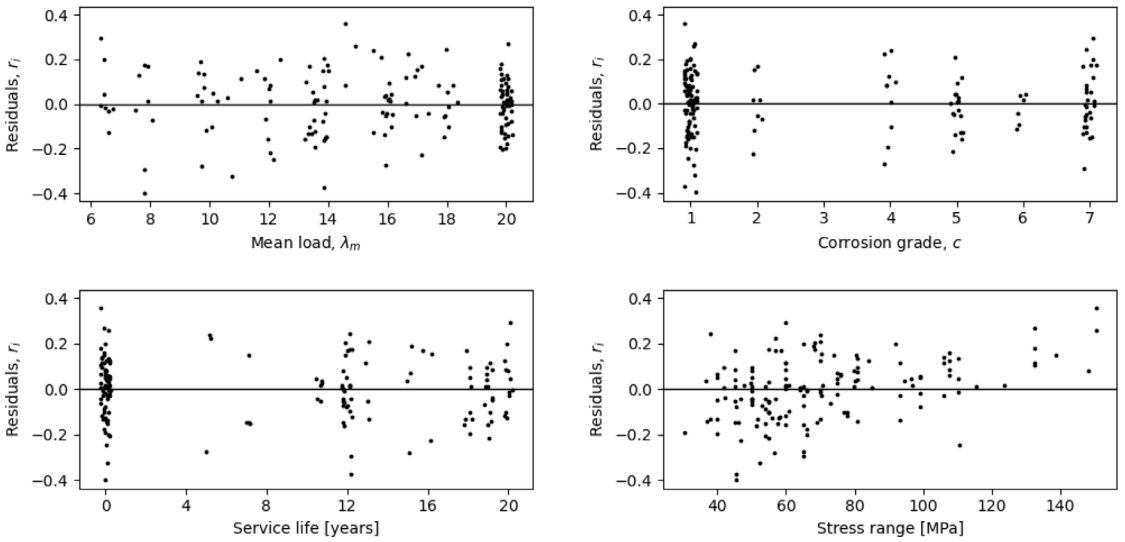


Fig. 10. Residual plots for base case. Plotting position has been jittered along horizontal axis for mean load, corrosion grade and service life.

- Chain condition during operation. This will, presumably, have the largest impact on the samples with the lowest corrosion grades. During operation, these will have accumulated loads at a chain condition similar to that in the test. For the samples with the highest corrosion grades, a large share of the fatigue loads will have accumulated for a better chain condition than during the test, thereby utilizing less of the fatigue capacity at this chain condition during the prior service life.

Altogether, it seems unlikely that prior fatigue loads alone would be the cause for as much as 25% difference, on average, in the fatigue performance of used chain compared to new chain.

Alternative explanations or contributions to the observed differences are the following, possibly in combination. (i) The condition of chains rated to $c = 1$ may not be strictly comparable to that of new chains. For instance, the surface roughness of the used chains may be slightly higher than that which is achieved from pre-soaking new chain samples in artificial seawater. (ii) The mean load and corrosion grade-dependent S–N intercept in (6) with $g_2(c) = c$ implies $A(\sigma_m, c) \propto 10^{\beta_2 \cdot c}$, and this empirical formulation does not likely represent the *true* effect of chain condition. (iii) None of the used chain samples with $c \in \{1, 2\}$ were tested at the highest mean loads (see Fig. 3), meaning that prediction of fatigue capacity at the reference level ($\lambda_m = 20, c = 1$) involves extrapolation beyond the used chain data set. With $\lambda_m = 14$, for which several such samples were tested, the difference in estimated capacity is reduced to around 17%.

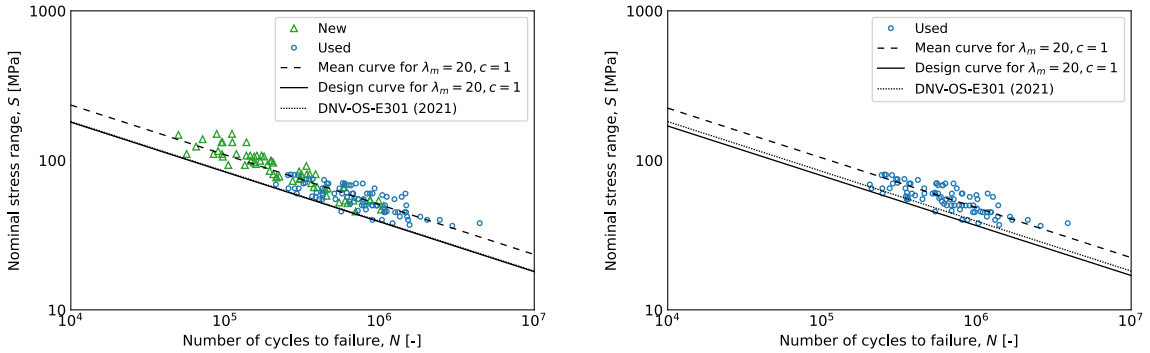
In any case, the differences between the results obtained for the two data sets are small compared to the estimated mean load and corrosion effects over the range of tested values: they correspond roughly to the effect of increasing the mean load by 2.5 [% MBL] ($10^{\hat{\beta}_1 \cdot 2.5} = 0.74$) or an increase in corrosion grade by around one unit ($10^{\hat{\beta}_2 \cdot 1} = 0.80$). Finally, keep in mind that the above discussion is based on point estimates for the model coefficients. Due to inclusion of fewer fatigue tests, the uncertainties associated with the coefficients are larger when estimated from either new chain or used chain data than for the base case model with all data. The difference in $\hat{\mu}_\alpha$ (or $\log A_{\text{ref}}$) between the three models is less than or equal to one standard error for the intercept.

6.6. Stress range effect

The residuals, describing differences between the data observed in the tests and those predicted by the model, are given by $r_i = y_i - \hat{y}_i$, where y_i is the test outcome, defined in (5), and $\hat{y}_i$ is the value predicted by the model. Residuals of the base case model are inspected in Fig. 10. They show no clear patterns when plotted versus the predictors (λ_m, c) or the service life of the chain. However; when plotted versus the stress range applied in the test, they reveal a clear trend with positive residuals for high stress ranges. This suggests that the data might favor a lower stress range effect (i.e., a lower value of m and a steeper curve in S–N plots), had it not been fixed at $m = 3$. Fig. 2 shows that the tests with highest stress ranges are all for new chain samples. This is consistent with previous studies on the new chain data from ND and TWI: when the stress range effect is estimated from the data, (i) the TWI data results in $m \approx 2.7$ for 127 mm chain and $m \approx 3$ for 76 mm chain [10], and (ii) the ND data results in an even lower stress range effect than that obtained for the TWI data [17].[4]

On the other hand, some used chain studies have suggested that a higher stress range effect may be appropriate for highly corroded chain. For the subset P1-R4-114e (cf. Table 1), with $c = 7$, Gabrielsen et al. [9] found that the data indicates $m > 3$ for the

[4] The steeper curve for the ND data may also be seen from Figure 4 in [10] or Figure 21 in [9].



(a) All data, base case model.

(b) Used chain, model based on used chain only (model 7).

Fig. 11. S–N data from fatigue tests with number of cycles to failure (N) transformed to represent $\lambda_m = 20$ [% MBL] and $c = 1$. DNV-OS-E301 design curve is included for reference (overlaps partly with solid line in the left figure).

tests with $\lambda_m = 18$ [% MBL], and even more so for the tests with $\lambda_m = 8$ [% MBL] (although, with only five tests at each mean load level the uncertainty in estimation of m is large for these two groups). Such trends are not revealed by the residual plots in Fig. 10, meaning that they could in principle be disguised by the model as corrosion grade effects or as group differences.

The model issues implied by Fig. 10 and the diverging reports for new chain compared to used and highly corroded chain advise caution as to how the stress range effect is treated. A step towards better understanding how the model parameters (including m) vary across groups, could be to generalize the varying-intercept model in (4) to a varying-intercept, varying-slope model [20]. However; a complicating factor is that the new chain tests cover a different interval of stress ranges compared to the used chain tests, as shown in Fig. 2. This implies that the resulting experimental design matrix is suboptimal with respect to the spread of stress ranges, mean loads and corrosion grades. Results from a model with even more parameters than the model applied for the present study should therefore be carefully evaluated before any conclusions are drawn, for instance by considering the posterior distributions of the parameters to assess how well defined they are by the data.

6.7. S–N design curves for reference level

Consider a sample that failed after N_1 cycles in a test conducted with stress range S_1 and associated with the parameters $\{\sigma_{m,1}, c_1\}$, and that we want to predict what the fatigue life would have been had the test been conducted with the same stress range but for a different mean load and corrosion grade, $\{\sigma_{m,2}, c_2\}$. By requiring that $S_2 = S_1$, where S_2 is the stress range in the hypothetical test for $\{\sigma_{m,2}, c_2\}$, the resulting transformation obtained from the S–N relation in (1) is simply the ratio of the S–N intercepts for the two states, $A(\sigma_{m,2}, c_2)/A(\sigma_{m,1}, c_1)$. For the base case model it is:

$$\frac{N_2}{N_1} = \underbrace{10^{\beta_1(\lambda_{m,2}-\lambda_{m,1})}}_{\text{Mean load correction}} \cdot \underbrace{10^{\beta_2(c_2-c_1)}}_{\text{Corrosion grade correction}} \tag{9}$$

where N_2 is the predicted fatigue life of the given sample in a test associated with the alternative state, $\{\sigma_{m,2}, c_2\}$. As noted, the transformation has been split into two factors to distinguish between the correction due to differences in mean load and corrosion grade, respectively.

Fig. 11 shows a S–N plot with all data transformed by (9) to represent the reference level, $\lambda_m = 20$ [% MBL] and $c = 1$. The scatter of the data, in terms of deviation from the mean curve, is considerably reduced for both new and used chain tests compared to that seen for the raw data in Fig. 2. It is emphasized that the estimated group-specific intercepts, $\hat{\alpha}_j$, are not utilized for adjustment of the data for this plot. The between-group variability is instead accounted for by using the total standard deviation, $\hat{\sigma}$, to establish the design curve in accordance with (7). It is noted that two of the used chain samples fall below the design curve of the model that is based on all data (Fig. 11(a)). The fact that these two least favorable results are both for used chain may be due to the larger variability for these data, or it could be that the corrosion grades assigned to them do not fully represent the condition of these samples. In any case, two samples falling below the design curve is acceptable, in principle, considering the total number of tests ($n = 154$) and the required 97.7% exceedance probability. A possible remedy could be to utilize the model which is based on used chain data only (model 7, cf. Table 4). For this model, the lower estimated mean curve and the higher standard deviation ($\hat{\sigma}$) shifts the design curve just enough towards lower fatigue life for all the samples to fall above the design curve (Fig. 11(b)).

6.8. On assumptions and limitations

Assumptions and limitations that have not yet been addressed are now briefly discussed.

Variability of used chain data. The model that is only based on used chain data results in a larger within-group variability (σ_ϵ) compared to the model which uses only new chain data (cf. Table 4). This could suggest that corrosion increases the variability of the fatigue performance, although, the differences are small and a systematic study of the variability within groups with different corrosion grades is needed to conclude. A more pronounced increase is seen for the between-group variability (σ_α), suggesting a larger influence from hidden variables for used chain compared to new chain or that the corrosion grade scale has not been applied consistently across groups. Furthermore; a concern is that the used chain data, consisting of mostly R4 chain, could be too homogeneous such that σ_α might be underestimated. This could be addressed by inclusion of more used chain samples with material grades other than R4.

Chain condition. The corrosion grades assigned to the used chain samples are treated by the regression model as certain values. Consequently, the uncertainty that arises from using this subjective measure is not quantified by the model. Until a more objective method for quantification of chain condition is available, the uncertainty in the estimated corrosion grade effect ($\hat{\beta}_2$) therefore cannot be accurately quantified. It is also emphasized that the used chains were visually inspected before the tests, without detection of fatigue cracks in any of the samples. Hence, the model is strictly valid only for chain segments for which the crack initiation has not evolved into visible cracks.

Validity of mean load effect. Tests have been conducted for mean loads in the range from 6% to 20% MBL, with the majority (143 out of 154 tests) from 8% MBL and upwards. The model should therefore be used cautiously for prediction of fatigue capacity for mean loads outside this range, as the data set includes no tests that can be used to quantify the fatigue performance at lower or higher mean loads.

Design curve estimation. S–N design intercepts presented here are estimated by applying an offset to $\hat{\mu}_\alpha$. This simplification neglects the inferential uncertainty of the slope coefficients which could, in principle, affect the required design curve offset differently at various mean load and corrosion grade levels. A formally more correct approach would be to account for this inferential uncertainty and estimate the intercept with 97.72% exceedance probability for a range of mean loads and corrosion grades, for instance by use of Monte Carlo simulation. Some examples on how such simulations may be used to account for the effect of inferential uncertainty on any quantity of interest are presented in [20].

Number of links. We have assumed that the test results represent single link failures, while in practice, they represent failure of the weakest out of multiple links. Our assumption has the following impact on the results: (i) the estimated distribution is shifted towards lower fatigue capacity, (ii) the within-group variability (σ_ϵ) is underestimated, and (iii) the assumption of normal distributed errors is questionable. The former two effects have opposite implications with respect to conservatism, and they are both amplified by considering first fractures only (as consideration of additional failures would provide data points at longer fatigue lives and presumably increase the variability). In sum, however, the assumption is considered to be conservative. Note also that differences in number of links applied in the tests lead to inconsistencies across groups. This is not addressed, but may in principle be perceived by the model as a group effect.

7. Conclusion

Full scale fatigue test data of new and used studless mooring chain have been analyzed. Both new chain data and used chain data include tests conducted at a range of mean load levels between 6% and 20% MBL. In addition, the used chain samples represented a range of surface conditions, quantified by means of a custom corrosion grade scale ranging from 1 (new or as-new chain) to 7 (heavily corroded chain). Based on an extended S–N formulation with a parameterized S–N intercept parameter, the effects of mean loads and corrosion grade were quantified empirically from the data. A hierarchical varying-intercept model was used for the regression analysis, to quantify correlations within subsets of the data and properly account for their effect on the estimated model coefficients. Detailed results of the data analysis have been presented and thoroughly discussed, including the goodness of the selected grouping criterion, the magnitude of mean load and corrosion effects, differences in the fatigue performance of new versus used chain samples and consequences of modeling assumptions.

We have shown that the data supports a subset division based on platform of origin, chain size, material grade and service life duration, and that this refined grouping criterion does not lead to an overfitted model. Furthermore, mean load and chain condition are shown to have a significant impact on the fatigue performance of the chains, such that the fatigue life (i) increases when the mean load is reduced and (ii) decreases when the corrosion grade increases. Over the range of mean loads and corrosion grades considered, these effects are considerably larger than the differences in estimated fatigue life for new versus used chain when evaluated at the same mean load and corrosion level.

The used chain data are all of R4 material grade except one group of R3 chain, which is also the only group with general corrosion and material loss of some importance. The results for this R3 group suggest that (i) material loss may need to be accounted for in addition to the corrosion grade describing surface condition, and (ii) the used chain data set may be overly homogeneous and lead to an underestimation of the variability of the fatigue capacity. Inclusion of more tests for used chain samples with considerable material loss and for material grades other than R4 could serve to further quantify this.

Further work. The present work has been limited to assuming fixed stress range effect ($m = 3$), and validation of the model indicates that this assumption is questionable. Although we have shown that the resulting model is still able to describe the effects of mean load and corrosion and predict the fatigue performance of mooring chains reasonably well, this somewhat stringent assumption and its impact on fatigue life prediction should be further assessed as part of future work. Also, methods to measure objectively the chain condition would greatly improve the model, and mitigate uncertainties related to the current subjective assessment. Finally, fatigue tests of both new and used chains at mean loads above 20% MBL would contribute to an improved understanding and quantification of the mean load effect beyond this level. This would be of relevance for the mooring systems of for instance floating wind turbines, possibly operating at such higher mean loads over long periods.

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.

Data availability

The authors do not have permission to share data

Acknowledgments

This study was co-funded by the Research Council of Norway, through the project 280705 "Improved lifetime estimation of mooring chains" (LIFEMOOR). TWI is acknowledged for providing test results from the TWI JIP on fatigue performance of mooring chain. Vicinay is acknowledged for sharing results from the new chain tests performed at a range of mean load levels.

Appendix A. Some aspects of the hierarchical linear model

This appendix briefly describes relevant properties and aspects of the hierarchical linear model described in Section 3 and defined in Eq. (4).

Standard error of $\hat{\alpha}_j$ and effect of group sample size. Given the data and model parameters, the posterior distribution of the group intercepts α_j follow independent normal distributions [20, p. 394], $\alpha_j \sim \mathcal{N}(\hat{\alpha}_j, V_j)$, where $\hat{\alpha}_j$ is the posterior mean of α_j and V_j is the corresponding estimation variance. These are given by

$$\hat{\alpha}_j = V_j \left(\frac{n_j}{\sigma_\epsilon^2} (\bar{y}_j - \bar{\boldsymbol{x}}_j \boldsymbol{\beta}) + \frac{1}{\sigma_\alpha^2} \mu_\alpha \right), \quad V_j = \left(\frac{n_j}{\sigma_\epsilon^2} + \frac{1}{\sigma_\alpha^2} \right)^{-1} \tag{A.1}$$

where n_j is the number of samples in group j, $\bar{y}_j$ is the mean of the outcomes for group j, $\bar{\boldsymbol{x}}_j$ is a vector containing the means of the predictors for this group. This implies that (i) for groups with sample size $n_j < \sigma_\epsilon^2/\sigma_\alpha^2$, the estimated intercept is closer to μ_α than to the unpooled estimate that would be obtained from group-specific information only (i.e., $\bar{y}_j - \bar{\boldsymbol{x}}_j \boldsymbol{\beta}$); (ii) for groups with $n_j > \sigma_\epsilon^2/\sigma_\alpha^2$, the group-specific information is more important than μ_α; and (iii) the uncertainty in $\hat{\alpha}_j$ decreases with increasing group sample size, n_j.

Within-group correlation. The hierarchical linear model in (4) implies that the outcomes for samples within the same group are correlated with correlation coefficient

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2} \tag{A.2}$$

This quantity is also known as the *intraclass correlation* (ICC).

Prediction of new observations. Two scenarios are of relevance for predicting the outcome for a new sample. (i) For a new sample, $\tilde{y}$, from an existing group j, the outcome may be predicted from $\tilde{y} \sim \mathcal{N}(\alpha_j + \tilde{\boldsymbol{x}}\boldsymbol{\beta}, \sigma_\epsilon^2)$, where $\tilde{\boldsymbol{x}}$ are the predictor values for the new sample. Note that as written here, the predictive distribution ignores inferential uncertainty and assumes that the coefficients $(\alpha_j, \boldsymbol{\beta}, \sigma_\epsilon)$ are known. In practice, estimated values $(\hat{\alpha}_j, \hat{\boldsymbol{\beta}}, \hat{\sigma}_\epsilon)$ must be used and the inferential uncertainty may be of importance as well. (ii) For a new sample from a new group (i.e., a group that was not included in the data analysis), the intercept is unknown and must be predicted from $\tilde{\alpha} \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$. This may be expressed as

$$\tilde{y} \sim \mathcal{N}(\mu_\alpha + \tilde{\boldsymbol{x}}\boldsymbol{\beta}, \sigma^2) \tag{A.3}$$

where $\sigma = \sqrt{\sigma_\alpha^2 + \sigma_\epsilon^2}$ is the total standard deviation of the outcome for a new sample from a new group.

Table B.5
Comparison of model parameters estimated from REML vs. MCMC (sample size 2×10^4). For MCMC, listed point estimates and standard errors se(·) are the modes and standard deviations of the simulated posterior distributions, respectively.

Model no.	1		4		5	
Description	Base case		Alt. mean load predictor		Alt. corrosion predictor	
$g_1(\sigma_m)$	λ_m		σ_m		λ_m	
$g_2(c)$	c		c		$\log c$	
Method	REML	MCMC	REML	MCMC	REML	MCMC
$\hat{\mu}_\alpha$	12.236	12.232	12.129	12.128	12.154	12.155
$\hat{\beta}_1$	−0.0514	−0.0512	−0.0072	−0.0072	−0.0508	−0.0508
$\hat{\beta}_2$	−0.0977	−0.0975	−0.1003	−0.0999	−0.6616	−0.6614
se($\hat{\mu}_\alpha$)	0.0811	0.0841	0.0851	0.0867	0.0815	0.0848
se($\hat{\beta}_1$)	0.0044	0.0045	0.0007	0.0007	0.0046	0.0047
se($\hat{\beta}_2$)	0.0093	0.0098	0.0111	0.0118	0.0686	0.0740
$\hat{\sigma}_\epsilon$	0.144	0.144	0.145	0.145	0.145	0.144
$\hat{\sigma}_\alpha$	0.092	0.089	0.123	0.120	0.104	0.102
$\hat{\sigma}$	0.171	0.169	0.190	0.188	0.178	0.176
$\log A_{\text{ref}}$	11.111	11.110	–	–	11.138	11.139

Table B.6
Model comparison by LOO cross-validation, obtained from PSIS-LOO. See main text for description of columns.

Model no.	Description	$\widehat{\text{elpd}}_{\text{loo}}$	$\hat{p}_{\text{loo}}$	$\widehat{\text{elpd}}_{\text{diff}}$	w
1	Base case	68.2	19.1	0.0	0.74
5	Alternative corrosion predictor	66.8	20.2	1.4	0.20
4	Alternative mean load predictor	65.6	21.3	2.6	0.06

Appendix B. Model comparison by Bayesian inference and cross-validation

As a verification of the model parameters estimated by REML, and as the basis for model comparison through LOO cross-validation, Bayesian inference is applied to the base case model and the models with alternative predictors (models 1, 4 and 5, cf. Table 4). The posterior distributions of the model parameters are obtained from MCMC sampling using the Python module `bambi` [29] (version 0.10.0), with weak priors for $\boldsymbol{\beta}$, σ_ϵ, μ_α and σ_α. In Table B.5, the REML estimates are compared to the estimated Bayesian posterior modes. They are seen to agree very well across all listed parameters and standard errors, for all three models.

For comparison of the predictive fit of the models, we apply Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO), a method proposed by Vehtari et al. [30]. This procedure uses the simulated posterior distributions to provide an estimate of each model's pointwise out-of-sample predictive accuracy, similar to the widely applicable information criterion (WAIC) which is asymptotically equal to LOO [30]. The calculations are conducted using Python module `arviz` [31] (version 0.15.1) and its function `arviz.compare(ic='loo')`, see Martin et al. [32, Sec. 2.5].

The results from this comparison are presented in Table B.6, with models listed in descending order with respect to the predictive accuracy. The columns of this table are defined as follows: (i) $\widehat{\text{elpd}}_{\text{loo}}$ is the expected log pointwise predictive density, estimated by PSIS-LOO and given on a "log score" scale (meaning that a higher value describes a better predictive fit); (ii) $\hat{p}_{\text{loo}}$ is the estimated penalization term for model complexity and may be loosely interpreted as the effective number of parameters; (iii) $\widehat{\text{elpd}}_{\text{diff}}$ is the difference between the value of $\widehat{\text{elpd}}_{\text{loo}}$ for the model with the highest predictive accuracy (out of those considered) and that obtained for each model; (iv) w is the "Akaike weight", which may be interpreted as the relative likelihood of each model [33, Sec. 2.9].

For the data set considered within the present study, the results in Table B.6 support the choice of the base case model over those with alternative mean load and corrosion grade predictors: it is nearly 4 times more likely than model 5 with $g_2(c) = \log c$ and roughly 12 times more likely than model 4 with $g_1(\sigma_m) = \sigma_m$. It is also noted that the estimated penalization term, $\hat{p}_{\text{loo}}$, is slightly higher for models 4 and 5 compared to the base case model. This is consistent with the lower constraint on the group-specific intercepts for these alternative models, as quantified by the higher between-group variability ($\hat{\sigma}_\alpha$).

References

[1] Schneider CRA, Maddox SJ. Best practice guide on statistical analysis of fatigue data. Doc. IIW-XIII-WG1-114 - 03, International Institue of Welding (IIW); 2003, p. 1–30.

[2] Gabrielsen Ø, Larsen K, Dalane O, Lie HB, Reinholdtsen S-A. Mean load impact on mooring chain fatigue capacity: Lessons learned from full scale fatigue testing of used chains. In: Proceedings of the ASME 2019 38th International Conference on Ocean, Offshore and Arctic Engineering, No. OMAE2019-95083. 2019, http://dx.doi.org/10.1115/OMAE2019-95083.

[3] DNV. Offshore standard - position mooring (DNV-OS-E301). 2021, Edition July 2021.

[4] ISO. Petroleum and natural gas industries - specific requirements for offshore structures - Part 7: Stationkeeping systems for floating offshore structures and mobile offshore units (ISO 19901-7). 2013.

[5] Fredheim S, Reinholdtsen S-A, Håskoll L, Lie HB. Corrosion fatigue testing of used, studless, offshore mooring chain. In: Proceedings of the ASME 2013 32nd international conference on ocean, offshore and arctic engineering, No. OMAE2013-10609. 2013.

[6] Gabrielsen Ø, Larsen K, Reinholdtsen SA. Fatigue testing of used mooring chain. In: Proceedings of the ASME 2017 36th international conference on ocean, offshore and arctic engineering, No. OMAE2017-61382. 2017, http://dx.doi.org/10.1115/OMAE2017-61382.

[7] Gabrielsen Ø, Liengen T, Molid S. Microbiologically influenced corrosion on seabed chain in the north sea. In: Proceedings of the ASME 2018 37th international conference on ocean, offshore and arctic engineering. 2018, p. 1–9. http://dx.doi.org/10.1115/OMAE2018-77460.

[8] Ma K-t, Gabrielsen Ø, Li Z, Baker D, Yao A, Vargas P, Luo M, Izadparast A, Arredondo A, Zhu L, Sverdlova N, Hø gsæt IS. Fatigue tests on corroded mooring chains retrieved from various fields in offshore West Africa and the North Sea. In: Proceedings of the ASME 2019 38th international conference on ocean, offshore and arctic engineering, No. OMAE2019-95618. 2019, http://dx.doi.org/10.1115/OMAE2019-95618.

[9] Gabrielsen Ø, Reinholdtsen S-A, Skallerud B, Haagensen PJ, Andersen M, Kane P-A. Fatigue capacity of used mooring chain - results from full scale fatigue testing at different mean loads. In: Proceedings of the ASME 2022 41st international conference on ocean, offshore and arctic engineering, No. OMAE2022-79649. Hamburg, Germany; 2022, http://dx.doi.org/10.1115/OMAE2022-79649.

[10] Zhang Y, Smedley P. Fatigue performance of high strength and large diameter mooring chain in seawater. In: Proceedings of the ASME 2019 38th international conference on ocean, offshore and arctic engineering, No. OMAE2019-95984. 2019, http://dx.doi.org/10.1115/OMAE2019-95984.

[11] Fernández J, Arredondo A, Storesund W, González JJ. Influence of the mean load on the fatigue performance of mooring chains. In: Proceedings of the annual offshore technology conference, No. OTC-29621-MS. 2019, http://dx.doi.org/10.4043/29621-MS.

[12] Martinez Perez I, Bastid P, Constantinescu A, Venugopal V. Multiaxial fatigue analysis of mooring chain links under tension loading: Influence of mean load and simplified assessment. In: Proceedings of the ASME 2018 37th international conference on ocean, offshore and arctic engineering. 2018, p. 1–11. http://dx.doi.org/10.1115/OMAE2018-77552.

[13] Dowling N, Thangjitham S. An overview and discussion of basic methodology for fatigue. In: Fatigue and fracture mechanics: 31st volume. 2000, p. 3–36. http://dx.doi.org/10.1520/stp14791s.

[14] Dowling NE. Mean stress effects in stress-life and strain-life fatigue. In: SAE technical papers, Vol. 2. SAE International; 2004, http://dx.doi.org/10.4271/2004-01-2227.

[15] Lone EN, Sauder T, Larsen K, Leira BJ. Fatigue assessment of mooring chain considering the effects of mean load and corrosion. In: Proceedings of the ASME 2021 40th international conference on ocean, offshore and arctic engineering, No. OMAE2021-62775. Virtual; 2021, http://dx.doi.org/10.1115/OMAE2021-62775.

[16] Mendoza J, Haagensen PJ, Köhler J. Analysis of fatigue test data of retrieved mooring chain links subject to pitting corrosion. Mar Struct 2022;81:103119. http://dx.doi.org/10.1016/j.marstruc.2021.103119.

[17] Denton N. Corrosion fatigue testing of 76mm grade R3 & R4 studless mooring chain. Joint industry study report H5787/NDAI/MJV, Rev.0, 2002.

[18] Fernández J, Storesund W, Navas J. Fatigue performance of grade R4 and R5 mooring chains in seawater. In: Proceedings of the ASME 2014 33rd international conference on ocean, offshore and arctic engineering, No. OMAE2014-23491. 2014, http://dx.doi.org/10.1115/OMAE2014-23491.

[19] TWI. Fatigue performance of mooring chains in seawater, final report. Techn. rep. report no. 22116/62/18, Rev. 2, 2018.

[20] Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Analytical methods for social research, Cambridge: Cambridge University Press; 2007, http://dx.doi.org/10.1017/CBO9780511790942.

[21] Lindstrom MJ, Bates DM. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J Amer Statist Assoc 1988;83(404):1014–22. http://dx.doi.org/10.1080/01621459.1988.10478693.

[22] Gelman A, Carlin J, Stern H, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Texts in statistical science, 3rd ed. Boca Raton, Fl: CRC Press; 2014.

[23] Kruschke JK. Doing Bayesian data analysis. 2nd ed. Academic Press; 2015, http://dx.doi.org/10.1016/B978-0-12-405888-0.09999-2.

[24] Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. In: Proceedings of the 9th python in science conference. 2010, p. 92–6. http://dx.doi.org/10.25080/Majora-92bf1922-011.

[25] Gabrielsen Ø, Liengen T, Rørvik G, Molid S, Stavang T. Corrosion experience with low carbon steel R4 grade mooring chain. In: Offshore technology conference, No. OTC-31233-MS. OTC; 2021, http://dx.doi.org/10.4043/31233-MS, URL https://onepetro.org/OTCONF/proceedings-abstract/21OTC/3-21OTC/466810.

[26] DNV GL. Offshore standard - offshore mooring chain (DNVGL-OS-E302). 2015, Edition July 2015.

[27] Denton GN. RR1093 - An assessment of proof load effect on the fatigue life of mooring chain for floating offshore installations - Mooring Integrity Joint Industry Project Phase 2. Tech. rep., 2017.

[28] Martinez Perez I, Bastid P, Venugopal V. Prediction of residual stresses in mooring chains and its impact on fatigue life. In: Proceedings of the ASME 2017 36th international conference on ocean, offshore and arctic engineering. 2017, http://dx.doi.org/10.1115/OMAE2017-61720.

[29] Capretto T, Piho C, Kumar R, Westfall J, Yarkoni T, Martin OA. Bambi: A simple interface for fitting Bayesian linear models in Python. 2022, arXiv:2012.10754.

[30] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 2017;27(5):1413–32. http://dx.doi.org/10.1007/s11222-016-9696-4.

[31] Kumar R, Carroll C, Hartikainen A, Martin O. Arviz a unified library for exploratory analysis of Bayesian models in python. J Open Source Softw. 2019;4(33):1143. http://dx.doi.org/10.21105/joss.01143.

[32] Martin OA, Kumar R, Lao J. Bayesian modeling and computation in python. Boca Raton: Chapman and Hall/CRC; 2021, http://dx.doi.org/10.1201/9781003019169.

[33] Burnham KP, Anderson DR. Model selection and multimodel inference. 2nd ed.. New York, NY: Springer New York; 2002, http://dx.doi.org/10.1007/b97636.