METHODOLOGY OF MODEL DEVELOPMENT IN THE APPLIED SCIENCES

BARNA SZABÓ

Washington University

One Brookings Drive, St. Louis, Missouri, 63130 USA

szabo@email.wustl.edu

IVO BABUŠKA

Oden Institute, The University of Texas at Austin,

201 E. 24th Street, Austin, Texas 78712 USA

babuska@ices.utexas.edu

[Received: April 19, 2021; Accepted: June 14, 2021]

Abstract. The formulation and validation of mathematical models in the applied sciences are largely consistent with the methodology of scientific research programmes (MSRP), however an essential modification is necessary: The domain of calibration has to be defined. The ranking and systematic improvement of mathematical models based on objective criteria are described and illustrated by an example. The methodology outlined in this paper provides a framework for the evolutionary development of a large class of mathematical models.

Mathematical Subject Classification: 03C30, 62M86, 65N75, 65Z05

Keywords: Formulation, validation, calibration, uncertainty quantification, simulation.

1. INTRODUCTION

The formulation and validation of mathematical models is fundamental to predictive computational science [1]. These activities fit into the framework of the Methodology of Scientific Research Programmes (MSRP) as outlined by Imre Lakatos [2], subject to an essential modification; the definition of the domain of calibration, which is addressed in this paper.

We view a mathematical model as a transformation of one set of data (the input $\mathbf{D}$) into another set (the quantities of interest $\mathbf{F}$). The transformation is based on some idea of physical reality $\mathbf{I}$ cast in mathematical form. This is indicated in short hand by the expression

$$(\mathbf{D}, \mathbf{I}) \xrightarrow{\mathbf{p}} \mathbf{F}, \quad (\mathbf{D}, \mathbf{p}) \in \mathbb{C} \quad (1.1)$$

where the right arrow represents all operations needed to transform $\mathbf{D}$ into $\mathbf{F}$. These operations involve a set of parameters $\mathbf{p}$ that represent physical properties and their statistical distributions. These parameters are determined by calibration which is discussed in Section 2.2. The restrictions imposed on $\mathbf{D}$ by the assumptions incorporated into the model, and the intervals on which the parameters were calibrated, define the domain of calibration $\mathbb{C}$. We shall assume that $\mathbf{F}$ is a smooth function of $\mathbf{D}$ and $\mathbf{p}$, that is, a small change in $(\mathbf{D}, \mathbf{p})$ will result in a small change in $\mathbf{F}$. Consequently a calibrated model is a validated within the domain of calibration.

The domain of calibration is a fundamentally important attribute of a mathematical model: As long as $\mathbf{F}$ is a smooth function of $(\mathbf{D}, \mathbf{p})$, a condition usually satisfied in practice, just about any model, even pseudoscientific models, can be validated on a sufficiently small domain of calibration [3]. In chaotic systems, such as turbulent flow, depending on its definition, $\mathbf{F}$ may not be a smooth function of $(\mathbf{D}, \mathbf{p})$.

To highlight the importance of the concept of domain of calibration, consider Newtonian mechanics, arguably the greatest scientific achievement in history. For about 200 years, it was generally believed that Newton's three laws of motion and the law of universal gravitation were laws of nature. This was supported by the evidence that predictions based on Newton's laws were consistently confirmed by observations with almost perfect accuracy. However those observations involved only bodies that move at velocities much smaller than the speed of light. Therefore the domain of calibration of Newtonian mechanics was limited to velocities much smaller than the speed of light. The theory of relativity did not refute Newtonian mechanics but it did define its domain of calibration and downgraded its status from "law of nature" to that of a phenomenological model. Mathematical models that incorporate Newton's laws continue to be useful, as any civil engineer will testify. Newtonian mechanics has an amazingly large domain of calibration.

Arguably, all physical theories are phenomenological models. This was articulated by Hawking in [4]: *"I take the positivist viewpoint that a physical theory is just a mathematical model and that it is meaningless to ask whether it corresponds to reality. All that one can ask is that its predictions should be in agreement with observations."*

To this statement we append the explanatory note that we can reasonably ask the predictions to be in agreement with observation within the domain of calibration of the model only.

Popper expressed the same view in these words [5]: *"The empirical basis of objective science has nothing absolute about it. Science does not rest on solid bedrock. The whole towering structure, the often fantastic and audacious construction of scientific theories, is built over a swamp. Its foundations are pillars driven from above into the swamp - not down to any natural 'given', ground, but driven just as deeply as is necessary to support the structure. The reason why we stop driving the pillars deeper into the ground is not that we have reached solid rock. No, our decision is based on the hope that the pillars will support the structure."*

These observations are in line with the philosophical notion of model-dependent realism [6]: Mathematical models represent precisely formulated ideas of certain aspects of reality. They must not be confused with reality. Also, since more than one model can be formulated to simulate any aspect of reality, it is necessary to have metrics by which the relative merit of competing models is measured.

In this paper we illustrate the fundamentals of model development in the applied sciences through an example, the details of which have been documented in the technical literature in recent years.

2. THE METHODOLOGY OF MODEL DEVELOPMENT

We illustrate the methodology of model development with reference to a specific project, the details of which have been documented in the technical literature in references [7], [8], [9], [10] and [11]. The goal of that project was to develop and validate a mathematical model for estimating the probability of failure in critical mechanical and structural components subjected to cyclic loading in the high cycle regime. This problem arises in the management of the maintenance of high-value assets, such as helicopters: There is a strong incentive to minimize the number of unscheduled maintenance operations because it is much more economical to perform maintenance at regularly scheduled times than at the times when the presence of a flaw is detected.

In a scientific research programme, as the term was defined by Lakatos [2], there are generally accepted “hard core” theorems and assumptions that are not open to investigation. In addition, there are auxiliary hypotheses that may be altered or abandoned to conform with the available empirical evidence.

In the example discussed in this paper, our hard core assumptions will be the set of assumptions on which the linear theory of elasticity is based, augmented by the assumption that localized nonlinear effects can be correlated with the elastic stress field. The limitations imposed by the hard core assumptions are taken into consideration in the definition of the domain of calibration.

Figure 1. Schematic view of the model development process.

The auxiliary hypotheses are concerned with the prediction of failure events in metallic objects subjected to cyclic loading and the statistical dispersion associated with mechanical fatigue. The auxiliary hypotheses have their own domains of calibration which are generally much smaller than the domain of calibration of the hard core hypotheses.

Lakatos called the replacement or modification of auxiliary hypotheses “problem shifts”. A research programme is said to be progressive if each problem shift improves the predictive performance of the model. In our case, which is typical of research programmes in the applied sciences, the research programme can only be progressive, that is, the predictive performance of a model is guaranteed to increase monotonically. This is because ideas that do not improve predictive performance are filtered out in the course of model development. This establishes the conditions for an evolutionary process in model development.

A schematic view of the model development process is shown in Figure 1. On the left side of the figure the steps taken in the particular model development project referenced in this paper are shown. The universal categories, common to all model development projects, are indicated on the right hand side.

2.1. The mathematical model. The mathematical model has three parts: The first part, in the terminology of MSRP, incorporates the hard core assumptions: The second and third parts incorporate the auxiliary hypotheses.

2.1.1. The hard core assumptions. It is assumed that damage initiation, a highly non-linear and irreversible process, is confined to small subdomains that are surrounded by an elastic stress field. The nonlinear process is controlled by the stress field that satisfies the assumptions of the theory of linear elasticity.

The hard core assumptions are not subject to investigation. Rather, the domain of calibration is defined so as to ensure that the model is operated within the limitations imposed by these assumptions. In the present instance the domain of calibration is defined so that the size of the inelastic zone is limited.

2.1.2. Auxiliary hypothesis 1: The predictor. The predictor is a functional defined on the elastic stress field. Its role is to generalize the S-N data¹ to triaxial stress fields. It characterizes the conditions that drive the formation of cracks in the neighborhood of stress raisers. The definition of a predictor is subject to the restriction that it has to be a finite number which is independent of the choice of the coordinate system. This suggests that the predictor should be defined in terms of the stress invariants. Any number of predictors can be defined. We will consider the predictor defined in reference [7]. This predictor, denoted by G_α , depends on three parameters that have to be determined by calibration. Its definition is given in Appendix A.

¹The name refers to the relationship between the stress S and the number of load cycles N at which failure occurred under constant-cycle loading at various mean stress levels.

2.1.3. Auxiliary hypothesis 2: The statistical model. The statistical model characterizes the random nature of failure caused by mechanical fatigue. Statistical models are calibrated on the basis of the available S-N data. Many plausible statistical models can be defined [12]. The statistical model used in the investigation discussed here has five parameters and belongs in the class of random fatigue limit models proposed by Pascual and Meeker [13]. Its definition is given in Appendix B.

2.2. Calibration. It is assumed that the material is elastic and isotropic, therefore its stress-strain relationship is characterized by two experimentally determined constants. These constants, the modulus of elasticity and Poisson's ratio, are random variables, however their dispersion is small in comparison with the dispersion of crack formation events therefore their statistical variation is neglected.

The predictor generalizes the S-N data, collected for notch-free specimens tested either under constant or mildly varying uniaxial stress conditions, to notched specimens where the stress gradient can be very large and the stress conditions are triaxial at the notch roots. A key assumption is that crack formation is completely determined by the surrounding elastic stress field, notwithstanding the fact it is a highly nonlinear process. It is reasonable to expect that this assumption will hold when the size of the plastic zone is sufficiently small in comparison with some measure of volume associated with the predictor. Therefore it is necessary to give a precise definition for the meaning of "sufficiently small" which then limits the domain of calibration of the predictor. This is one of the tasks of model development. Further discussion and details are available in [10].

The statistical model employed in the model development project represented by the flowchart in Figure 1 is a random fatigue limit model that has five parameters defined in Appendix B. Those parameters are determined by maximizing the likelihood function on the S-N data [12], [7].

As indicated in Figure 1, the statistical model is calibrated first and then the predictor is calibrated on the basis of fatigue data obtained for notched specimens. Once the calibration of the predictor is complete, it is possible to compare alternative predictors and evaluate their relative performance with reference to the calibration data. This was documented in [9]. The conclusion was that the predictor G_α performs substantially better than previously proposed predictors. This is evidenced by the facts that its domain of calibration is larger and the value of the likelihood function is greater.

2.2.1. The domain of calibration. The domain of calibration establishes limits on the admissible data $\mathbf{D}$ and defines intervals for the parameters $\mathbf{p}$. In the example discussed here, $\mathbf{D}$ must be consistent with the assumptions of linear elasticity, that is, the strains are infinitesimal, the relationship between the stress and strain is linear, with the exception of small subdomains in the vicinity of stress raisers, the size of which is clearly defined. Examples are available in references [10], [11]. If the model was calibrated under special conditions (for example, under uniaxial loading only) then the domain of calibration imposes commensurate limitations on $\mathbf{D}$.

Restrictions on the parameters $\mathbf{p}$ are imposed by the availability of data used for calibration. For example, the five-parameter statistical model defined in reference [7], was calibrated on the basis of S-N data in the range of 10^4 to 10^7 cycles. This defines the domain of calibration for the statistical model. Similarly, the calibration of the predictor is limited by the interval of highly stressed volume for which calibration records for notched specimens are available [9].

An often overlooked limit on the statistical model is that calibration is based on events that have high probabilities of occurrence. Therefore prediction of events that have low probabilities of occurrence involves extrapolation beyond the domain of calibration. An important corollary is that the most important quantity of interest in engineering design, the factor of safety, cannot be validated. Therefore subjective decisions cannot be avoided in establishing factors of safety [8].

2.3. Prediction. Having formulated and calibrated a mathematical model, its predictive performance is tested against new data which are not in the domain of calibration. Predictions are made on the basis of information on the conditions under which new data will be observed. Ideally, the predictions should be made before the new data became available.

The value of new data outside of the domain of calibration is that the predictive performance of the model can be tested and the domain of the calibration set can be enlarged. The value of new data within the domain of calibration is that the dataset on which calibration was based is enriched and therefore confidence in the calibrated parameters is increased.

A typical validation statement is formulated as follows: The probability is (say, 90%) that the observation x will be in the interval $(x_L < x < x_U)$. In the case of the mechanical fatigue experiments considered here, the prediction is that, given a notched test article and the mean and amplitude of the applied cyclic load, the probability is 90% that the number of cycles N_f at which failure will occur will fall between the limits N_L and N_U . Another way of saying this is that in 9 out of 10 experiments N_f will lie in the interval (N_L, N_U) where N_L (resp. N_U) correspond to the 5% (resp. 95%) survival probability, given the loading conditions.

The probability of n successes in m trials is given by the binomial distribution:

$$Pr(m, n | \theta) = \frac{m!}{n!(m-n)!} \theta^n (1-\theta)^{m-n} \quad (2.1)$$

where θ is the ratio n/m as $m \rightarrow \infty$. In our example $\theta = 0.90$.

Remark 1. The selection of the 90% probability threshold was based on the following consideration: Validation is typically based on events that have high probability of occurrence. Calibration data are generally scarce for survival probabilities of less than 5% and greater than 95% and hence those probabilities fall outside of the domain of calibration of the statistical model.

2.4. Validation. The primary purpose of validation is to seek evidence that would justify rejecting a model. The secondary purpose is to develop information on the basis of which the domain of calibration is defined. The prediction considered here is

that out of 10 experiments there will be 9 failures in the predicted intervals, more precisely it was predicted that $\theta = 0.90$ in equation (2.1).

Suppose that 12 experiments were performed and 9 outcomes fell in the predicted interval. Would that justify rejecting the model? - Assuming that the model is correct ($\theta = 0.90$), according to equation (2.1) the probability of observing 9 successes in 12 trials would be 8.5%. In testing statistical hypotheses a null hypothesis is usually not rejected unless the probability observing an outcome, given that the hypothesis is true, is less than 5%. For example, if only 7 outcomes fell in the predicted interval out of 12 experiments then the probability of outcome, assuming that $\theta = 0.90$, would be less than 0.4% and we would not have confidence in the model. However it is difficult to justify any fixed value of probability below which a model would be rejected.

Models are typically rejected not because their validity was refuted by observation, but because a better model was found. By objectively ranking competing models, predictive performance is progressively improved over time. Model development projects are open-ended: No one can claim to have the final word. New ideas and new observations present opportunities for continued improvement. The management of the process of model development in industrial and research organizations is guided by the principles of simulation governance [14].

2.5. Disposition. If no reasons were found to reject a model then the validation data are merged with the calibration data and the calibrated parameters, as well as the domain of calibration, are updated.

Using the updated calibration records, the likelihood function is evaluated for the entire set of data. The value of the log likelihood function (LL) is a metric suitable for measuring the relative merit of alternative models. For a given set of data, the highest ranking model has the highest LL value and/or the largest domain of calibration.

New data within the domain of calibration presents an opportunity to update the calibrations. Predictions outside of the domain of calibration are either validated or fail to be validated. If validated then the calibration records are updated and the domain of calibration is suitably modified. If fail to be validated then a point outside of the domain of calibration is obtained which provides information about the size of the domain. A mathematical model is a validated model within its domain of calibration.

Remark 2. Model development projects are open-ended. The selection of a model from among a set of candidate models depends on the available data. Both the set of candidate models and the available data change over time, therefore any conclusion as to which model to use is likely to change over time. The exercise of command and control over numerical simulation, which include the formulation, ranking and updating of mathematical models and management of the relevant data is called simulation governance [14].

3. RANKING

Assume that two models M_1 and M_2 have the same number of statistical parameters and both passed validation tests with the same number of successes. We have to

decide which model is better and why. The answer is provided by the posterior ratio defined by

$$\varrho \stackrel{\text{def}}{=} \frac{Pr(M_1 | \mathbb{D})}{Pr(M_2 | \mathbb{D})} = \frac{Pr(\mathbb{D} | M_1)Pr(M_1)}{Pr(\mathbb{D} | M_2)Pr(M_2)} \quad (3.1)$$

where $\mathbb{D}$ represents the entire set of available data, that is, the data available for calibration and validation. The probabilities are obtained from the statistical model.

Usually it is assumed that $Pr(M_1) \approx Pr(M_2)$ in which case ϱ is approximately equal to the ratio of likelihoods, called the Bayes factor, which can be easily computed. If ϱ is much greater than one then we prefer model M_1 , if it is much smaller than one then we prefer model M_2 . If this ratio is between 3 and $1/3$ then the two models are in a virtual tie. For additional discussion and examples we refer to [3].

In addition to considering the posterior ratio, we have to consider the size of the domain of calibration. How much weight should be assigned to the size of the domain of calibration depends on the scope of the intended application of the model. In other words, we choose the model that fits the available data best. In the philosophical literature the term “abduction” (inference to the best explanation) is used to refer to this approach.

Remark 3. If models M_1 and M_2 do not have the same number of statistical parameters then equation (3.1) has to be modified by a factor, called the Occam factor, which penalizes the model that has the larger number of parameters. See (for example) [15].

4. PHILOSOPHICAL POSITIONS

The question of how progress occurs in science is one of the central themes in the philosophy of science. The approach presented here is consistent with the philosophical idea called model-dependent realism [6]. In this view, aspects of reality are seen through mathematical models. Therefore progress in science means progress in formulating, calibrating, validating and updating mathematical models. It is essential therefore to have objective measures for ranking mathematical models according to their predictive performance, taking into account their domains of calibration.

Particularly relevant to our discussion was Lakatos’ methodology of scientific research programmes. An opposing point of view, first articulated by Paul Feyerabend, is that it is not possible to establish meaningful methodological rules that govern scientific progress. In Feyerabend’s view, “epistemological anarchy” exists. In his words; “anything goes” [16]. These are colourful ways of saying that complete creative freedom exists.

The conflict between these points of view is resolved if we distinguish between the methodologies in the applied and foundational sciences. Model development in the applied sciences fits into the framework of MSRP well because the focus is on the auxiliary hypotheses. The hard core assumptions are based on ideas, theorems and methods previously established in the foundational sciences. It is doubtful that a set of rules can be devised that will encompass the diverse lines of investigation at the forefront of research in physics where the focus is on the hard core itself.

Reference [17] provides a comprehensive overview of the state of model development in physics. In the applied sciences Lakatos' views, whereas at the forefront of physics Feyerabend's epistemological anarchy seem to provide the best description of how scientific models evolve.

5. SUMMARY AND CONCLUSIONS

Mathematical models are transformations of one set of data into another, as indicated by equation (1.1). The transformation operators are based on clearly defined assumptions and depend on physical properties that have to be determined by calibration. The domain of calibration is defined by the intervals for which measurements of the physical properties are available, and the limitations imposed by the assumptions incorporated in the operators.

Consideration of the size of the domain of calibration is essential. Without such consideration no model could be rejected because just about any model can be calibrated on a sufficiently small domain.

Model development is always open-ended. This is because new data may become available and new models may be proposed at any time that will justify replacing or updating the currently adopted model.

The methodology outlined in this paper establishes a framework for the evolutionary development of mathematical models.

Acknowledgement. The authors thank Drs. Ricardo Actis and Börje Andersson for their helpful comments on the manuscript.

REFERENCES

1. J. T. Oden, Babuška I., and D. Faghihi. "Predictive Computational Science: Computer Predictions in the Presence of Uncertainty." *Encyclopedia of Computational Mechanics*. Ed. by René de Borst Erwin Stein and Thomas J. R. Hughes. Vol. 1: Fundamentals. John Wiley & Sons Ltd., 2017. DOI: 10.1002/9781119176817.ecm2101.
2. I. Lakatos. "Falsification and the methodology of scientific research programmes." *Can theories be refuted?* Ed. by Sandra G. Harding. Dordrecht: Springer Netherlands, 1976, pp. 205–259. DOI: 10.1007/978-94-010-1863-0_14.
3. B. Szabó and Babuška I. *Finite Element Analysis: Method, Verification and Validation*. John Wiley & Sons Ltd., 2021.
4. S. Hawking and R. Penrose. *The Nature of Space and Time*. Princeton University Press, Princeton, New Jersey, 1998.
5. K. R. Popper. *The Two Fundamental Problems of the Theory of Knowledge*. Routledge, 2014.
6. S. Hawking and L. Mlodinow. *The Grand Design*. Bantam Books, New York, 2010.
7. B. Szabó, R. Actis, and D. Rusk. "Predictors of fatigue damage accumulation in the neighborhood of small notches." *International Journal of Fatigue*, **92**(1), (2016), pp. 52–60. DOI: 10.1016/j.ijfatigue.2016.06.017.

8. B. Szabó, R. Actis, and D. Rusk. “On the formulation and application of design rules.” *Computers and Mathematics with Applications*, **74**(9), (2017), pp. 2191–2202. DOI: 10.1016/j.camwa.2017.06.040.
9. B. Szabó, R. Actis, and D. Rusk. “Validation of notch sensitivity factors.” *Journal of Verification, Validation and Uncertainty Quantification*, **4**(1), (2019). 011004. DOI: 10.1115/1.4044236.
10. B. Szabó, R. Actis, and D. Rusk. “Validation of a predictor of fatigue failure in the high cycle range.” *Computers and Mathematics with Applications*, **80**, (2020), pp. 2451–2461. DOI: 10.1016/j.camwa.2020.07.026.
11. B. Szabó, R. Actis, and D. Rusk. “Validation test of a predictor of fatigue failure under biaxial out-of-phase loading conditions.” *Computer Methods in Applied Mechanics and Engineering*, **378**, (2021). 113718. DOI: 10.1016/j.cma.2021.113718.
12. I. Babuška, Z. Sawlan, M. Scavino, B. Szabó, and R. Tempone. “Bayesian inference and model comparison for metallic fatigue data.” *Computer Methods in Applied Mechanics and Engineering*, **304**(1), (2016), pp. 171–196. DOI: 10.1016/j.cma.2016.02.013.
13. F. G. Pascual and W. Q. Meeker. “Estimating fatigue curves with the random fatigue-limit model.” *Technometrics*, **41**(4), (1999), pp. 277–289. DOI: 10.2307/1271342.
14. B. Szabó and R. Actis. “Simulation governance: Technical requirements for mechanical design.” *Computer Methods in Applied Mechanics and Engineering*, **249–252**(1), (2012), pp. 158–168. DOI: 10.1016/j.cma.2012.02.008.
15. D. S. Sivia. *Data Analysis. A Bayesian Tutorial*. 2nd ed. Oxford: Oxford University Press, 2006.
16. I. Lakatos and P. Feyerabend. *For and against Method: Including Lakatos’s Lectures on Scientific Method and the Lakatos-Feyerabend Correspondence*. University of Chicago Press, 2016.
17. R. Penrose. *Fashion, Faith and Fantasy in the New Physics of the Universe*. Princeton: Princeton University Press, 1999.

APPENDIX A. THE PREDICTOR

The predictor defined in reference [7] was used in the present investigation. This predictor, denoted by G_α , is based on the assumption that the onset of fatigue failure can be correlated with the averaged volume integral of a linear combination of two stress invariants. It is defined as follows:

$$G_\alpha(\sigma_{ij}, R) = \frac{1}{V_c} \int_{\Omega_c} (\alpha I_1 + (1 - \alpha)\bar{\sigma}) dV \left(\frac{1 - R}{2} \right)^{1/2}, \quad 0 \leq \alpha \leq 1 \quad (\text{A.1})$$

where σ_{ij} is the stress tensor field, $I_1 = \sigma_{kk}$ is the first stress invariant, $\bar{\sigma}$ is the von Mises stress, R is the cycle ratio, defined as the minimum normal stress divided by the maximum normal stress, and V_c is the volume of the domain of integration defined by

$$\Omega_c = \{\mathbf{x} \mid \sigma_1 > \beta \sigma_{\max} > 0\} \quad (\text{A.2})$$

where σ_1 is the first principal stress and $\sigma_{\max} > 0$ is the maximum macroscopic stress. This is a generalization of the uniaxial stress to triaxial stress in the sense that G_α is defined for triaxial stress and, in the special case when constant uniaxial stress σ_1 is applied then it has the value of the equivalent stress σ_{eq} :

$$G_\alpha(\sigma_{ij}, R) = \sigma_{\text{eq}} \stackrel{\text{def}}{=} \sigma_1 \left(\frac{1 - R}{2} \right)^{1/2}. \quad (\text{A.3})$$

The parameter β depends on the material and a parameter that characterizes the stress gradient in the vicinity of stress raisers. In machine design this parameter is the notch radius. Since surface defects generally cannot be characterized by a single geometric parameter, we use the highly stressed volume V for that purpose. By definition,

$$V \stackrel{\text{def}}{=} \int_{\Omega_\varrho} y(\mathbf{x}) dV \quad \text{where} \quad y = \begin{cases} 1 & \text{when } \sigma_1(\mathbf{x}) > \gamma \sigma_{\max} \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.4})$$

The domain of integration Ω_ϱ is the neighborhood of a stress raiser. Denoting the location where σ_1 is maximum at a stress raiser by $\mathbf{x}_0$ we have $\Omega_\varrho = \{\mathbf{x} \mid |\mathbf{x} - \mathbf{x}_0| < \varrho\}$ where ϱ is chosen large enough to include all points where $\sigma_1(\mathbf{x}) > \gamma \sigma_{\max}$ and small enough to include only one stress raiser or one group of closely located stress raisers.

The three parameters of G_α are determined by calibration. Parameter α establishes a convex combination of the first stress invariant I_1 and the von Mises stress $\bar{\sigma}$, see equation(A.1). Parameter β defines the domain of integration which depends on the material properties and the stress field in the vicinity of stress raisers. The parameter γ is independent of the material properties. It defines the highly stressed volume V which depends on the stress distribution, and hence the type of loading, but is independent of the magnitude of the load. The predicted number of fatigue cycles is not sensitive to the choice of γ . In reference [7] $\gamma = 0.85$ was used.

APPENDIX B. THE STATISTICAL MODEL

The selection of a statistical models from among candidate models is discussed in [12]. Based on those results, we selected the random fatigue limit model proposed by Pascual and Meeker [13]. This model is based on the assumption that the random variable defined by $W = \log_{10} N$, where N is the number of cycles to failure, is normally distributed with mean $\mu(G_\alpha)$ and standard deviation s .

$$\mu(G_\alpha) = A_1 - A_2 \log_{10}(G_\alpha - A_3), \quad G_\alpha - A_3 > 0. \quad (\text{B.1})$$

The parameter A_3 , called the fatigue limit, is assumed to be a random variable. Specifically, $V \stackrel{\text{def}}{=} \log_{10} A_3$ is assumed to be normally distributed with mean μ_f and standard deviation s_f . Therefore the statistical model considered here has five parameters. The set of five parameters that maximize the likelihood function is denoted by $\boldsymbol{\theta}$.

We will assume that $V = \log_{10} A_3$ has normal distribution with mean μ_f and standard deviation s_f and write $A_3 = 10^v$. Therefore the random fatigue limit model

has five parameters: $\boldsymbol{\theta} \stackrel{\text{def}}{=} \{A_1 A_2 s \mu_f s_f\}$ and the probability density function of w , given G_α , v and $\boldsymbol{\theta}$ is

$$\phi(w | G_\alpha, v, \boldsymbol{\theta}) = \frac{\exp(-(w - \mu(G_\alpha, 10^v))^2 / (2s^2))}{s\sqrt{2\pi}}. \quad (\text{B.2})$$

The probability density of v , given G_α and $\boldsymbol{\theta}$, is:

$$f(v | G_\alpha, \boldsymbol{\theta}) = \frac{\exp(-(v - \mu_f)^2 / (2s_f^2))}{s_f\sqrt{2\pi}}. \quad (\text{B.3})$$

The marginal probability density of w is

$$\phi_M(w | G_\alpha, \boldsymbol{\theta}) = \int_{-\infty}^{\log_{10} G_\alpha} \phi(w | G_\alpha, v, \boldsymbol{\theta}) f(v | G_\alpha, \boldsymbol{\theta}) dv \quad (\text{B.4})$$

and the marginal cumulative distribution function is:

$$\begin{aligned} \Phi_M(w | G_\alpha, \boldsymbol{\theta}) &= \int_{-\infty}^w \phi_M(t | G_\alpha, \boldsymbol{\theta}) dt \\ &= \frac{1}{2} \int_{-\infty}^{\log_{10} G_\alpha} \left(1 + \operatorname{erf} \left(\frac{w - \mu(G_\alpha, 10^v)}{s\sqrt{2}} \right) \right) f(v) dv. \end{aligned} \quad (\text{B.5})$$

The functions ϕ_M and Φ_M have to be evaluated numerically.

Given a set of independent and identically distributed observations $(w_i, G_\alpha^{(i)})$, $(i = 1, 2, \dots, n)$ the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\phi_M(w_i | G_\alpha^{(i)}, \boldsymbol{\theta}) \right]^{1-\delta_i} \left[1 - \Phi_M(w_i | G_\alpha^{(i)}, \boldsymbol{\theta}) \right]^{\delta_i} \quad (\text{B.6})$$

where $\delta_i = 0$ (resp. $\delta_i = 1$) if the specimen failed (resp. did not fail) at $n_i = 10^{w_i}$ cycles.

The set of parameters $\boldsymbol{\theta}$ maximize $L(\boldsymbol{\theta})$, equivalently the log likelihood function $LL(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$, with respect to the S-N data.