Complexity versus scatter in fatigue modelling

Fraunhofer ITWM
CHALMERS
Research Centre
Industrial Mathematics

Thomas Svensson

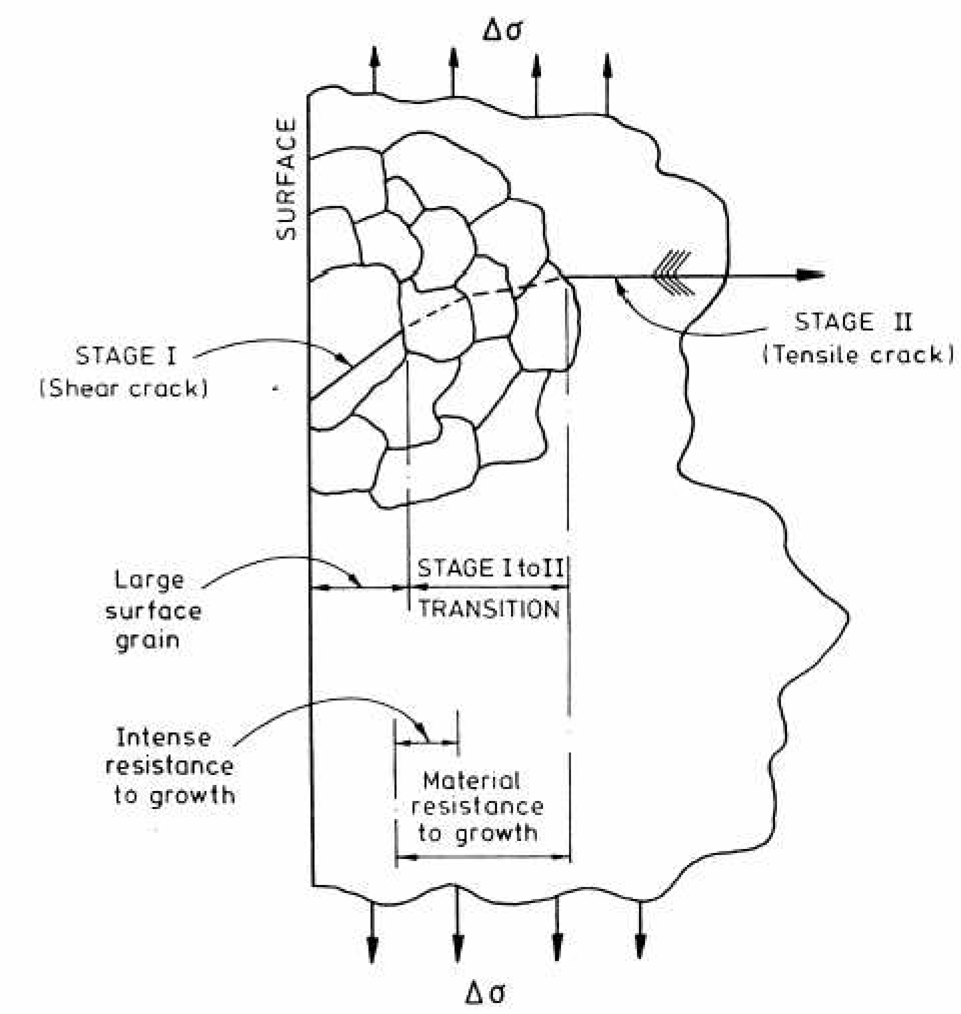
2004-08-16

Fatigue design versus scientific development

- Wöhler(1858) plotted log fatigue life against log stress and found a linear relationship, without any understanding about the cause of fatigue damage.
- Since then, the scientific community has made numerous investigations and now have a quite good understanding of the fatigue damaging mechanisms.
- Still, industry uses the Wöhler curve in design! Why?

- Engineers are oldfashioned ?
- Contacts between academy and industry is poor?
- **The knowledge about the variables needed in advanced models is limited.**

Fatigue, a complex phenomenon



- Crack initiation depends on local surface defects and chrystallographic orientation
- Initial short crack growth depends on local grain structure
- Long crack growth depends on ”mean resistance”

Fatigue, a complex phenomenon

Study of Interaction of Short Fatigue Cracks in 2024T351 Aluminium Alloy

H. Proudhon[1] and J-Y. Buffière[1]

[1] GEMPPM (INSA de Lyon), CNRS UMR 5510, 20 Avenue Albert Einstein 69621 Villeurbanne Cedex.
Henry.proudhon@insa-lyon.fr
Jean-yves.buffiere@insa-lyon.fr

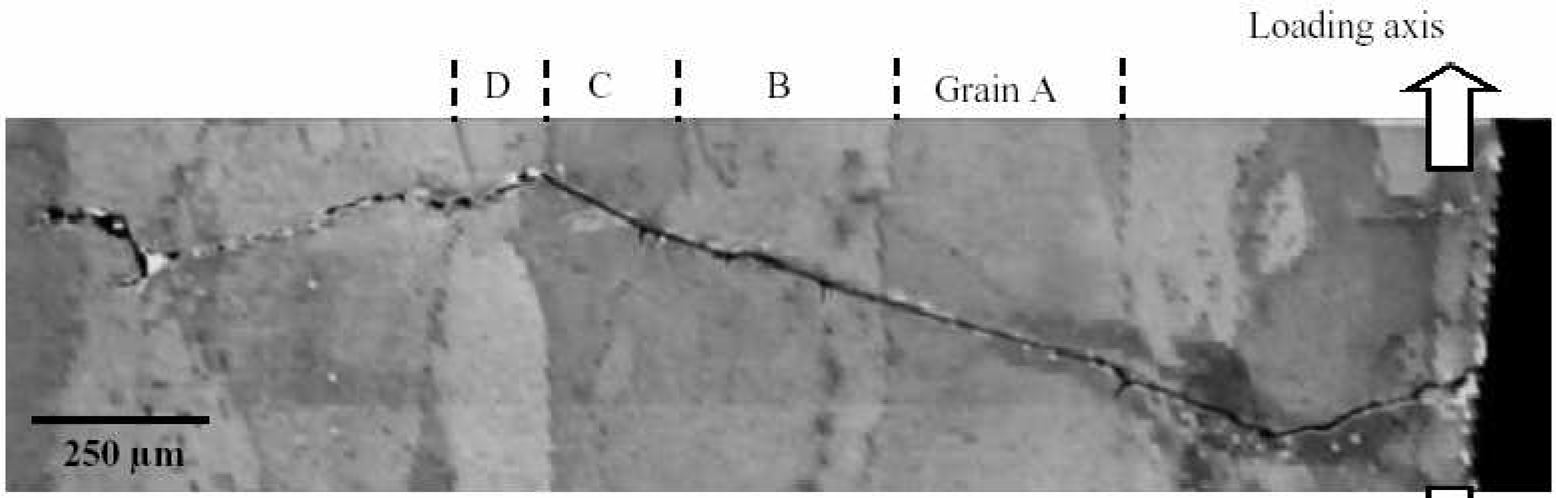


Figure 7. Superimposed EBSD acquisition of the crack surface and crack micrography after electropolishing, A B C D letters show grains where orientation has been analysed.

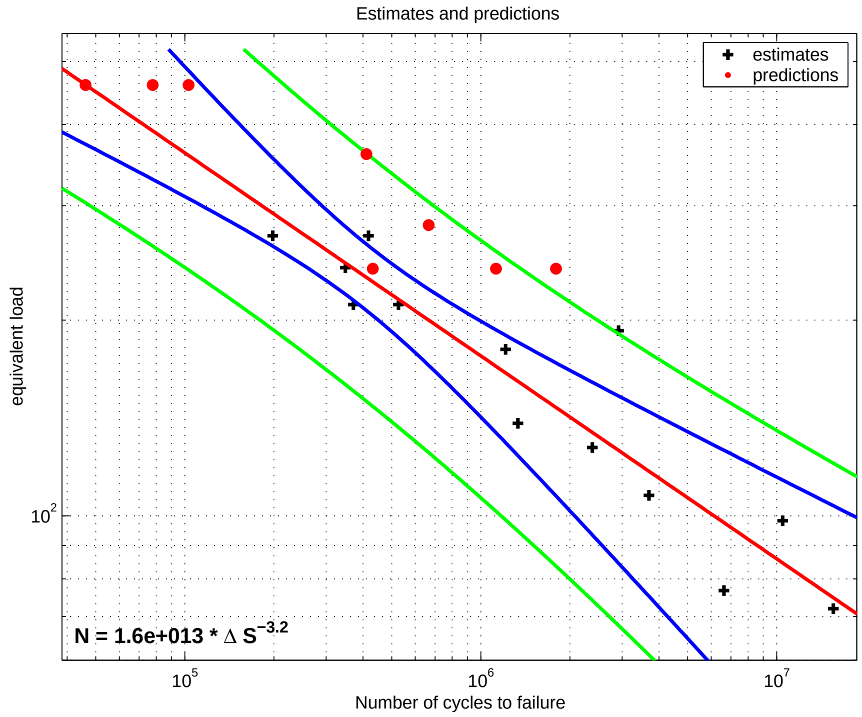
Fatigue life, the designers problem

- Physical models only exist at micro level
- The microstructure details are unknown
- The sizes of local defects are unknown at the design stage.
- The transitions between different mechanisms are unknown.

- Advanced calculation methods are not usable, since necessary variables and conditions are unknown.
- The engineer is forced to use empirical macroscopic models together with statistical modelling of scatter

Fraunhofer ITWM
CHALMERS
Research Centre
Industrial Mathematics

Fatigue design in practice



- The Basquin equation

$$N = \alpha \cdot \Delta S^{-\beta}$$

- The Palmgren-Miner rule

$$1 = \sum_i \frac{n_i}{N_i} = \frac{1}{\alpha} \sum_i n_i \Delta S_i^{\beta}$$

- The Paris´ law

$$\frac{da}{dN} = C' \cdot \Delta K^m = C \cdot \Delta S^m \cdot a^{m/2}$$

Model complexity and random scatter

For polynomial regression we have the following model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots + \beta_{p-1} x^{p-1} + \varepsilon,$$

$$\text{where} \quad \varepsilon \sim N\left(0, \sigma^2\right)$$

For estimating the parameters a number of measurements are performed and the least squares method gives the estimates:

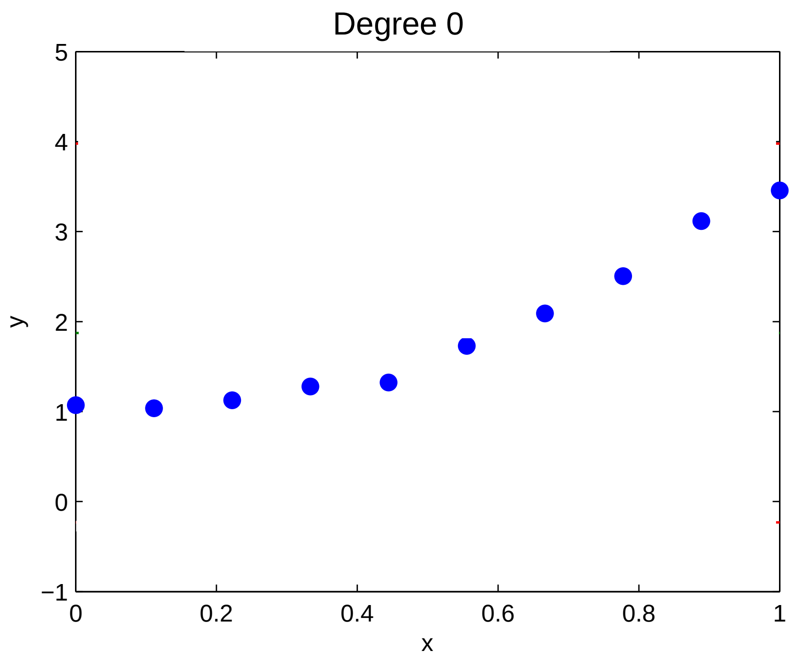
$$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots, \hat{\beta}_{p-1}, \hat{\sigma}$$

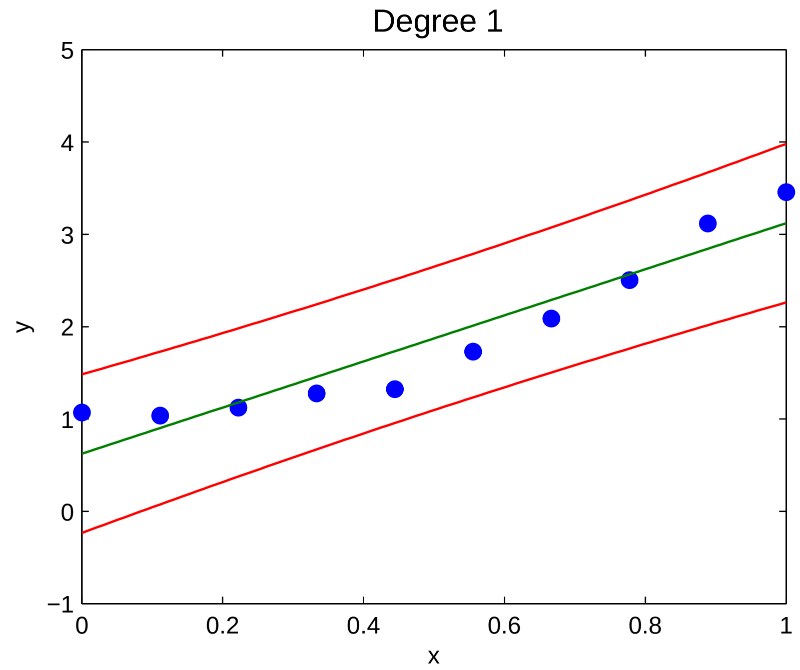
What polynomial degree should be chosen?

Model complexity and random scatter

Example: A data set with 10 measurements is fit to polynomials of different degrees:

The estimated functions with 95% **prediction intervals** are compared with the data points.

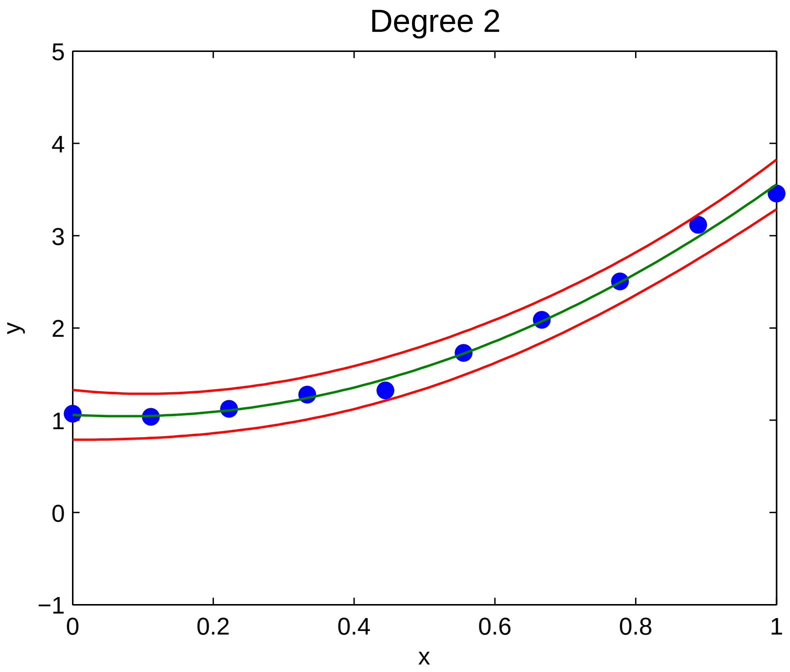


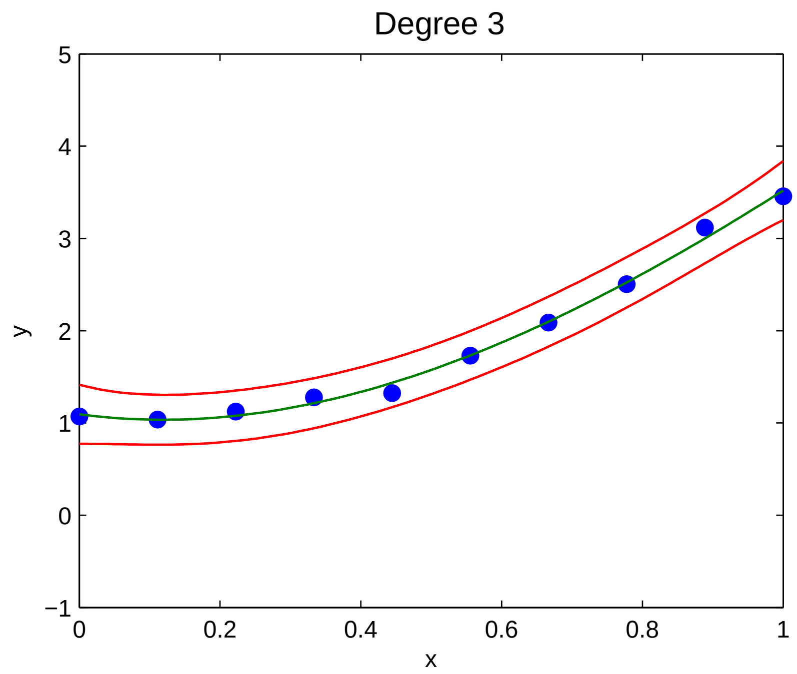


Fraunhofer ITWM
CHALMERS
Research Centre
Industrial Mathematics

Model complexity and random scatter

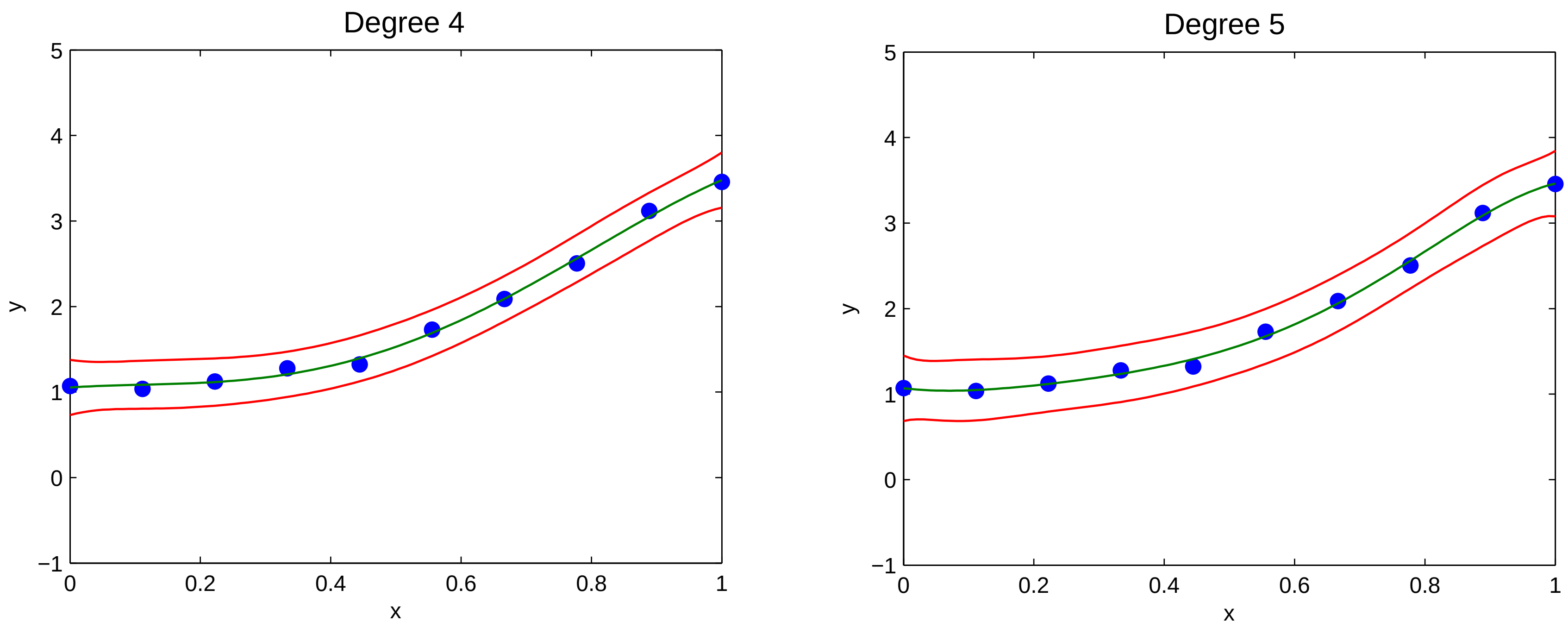
Increasing model complexity gives better fit and in the beginning also narrower prediction limits.





Model complexity and random scatter

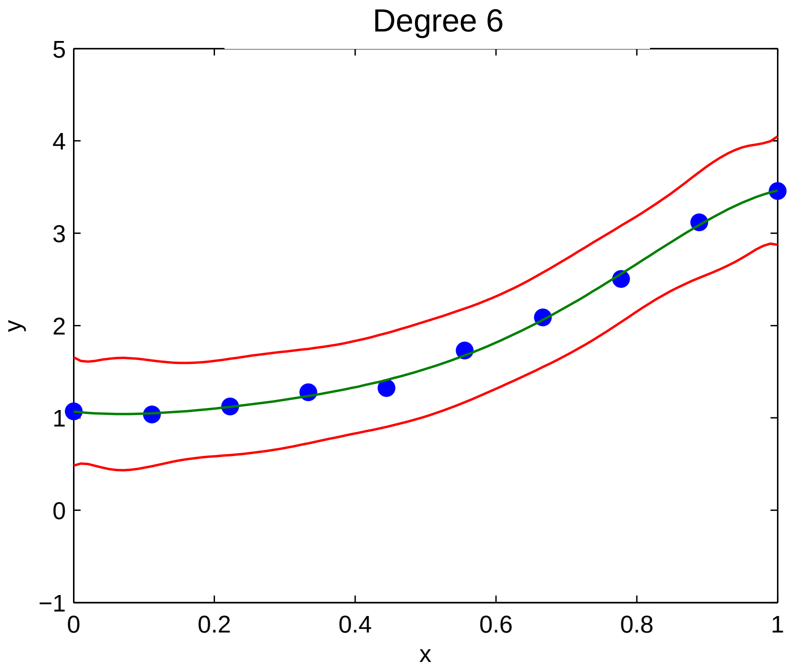
The fit becomes even better, but the prediction limits is not decreasing.

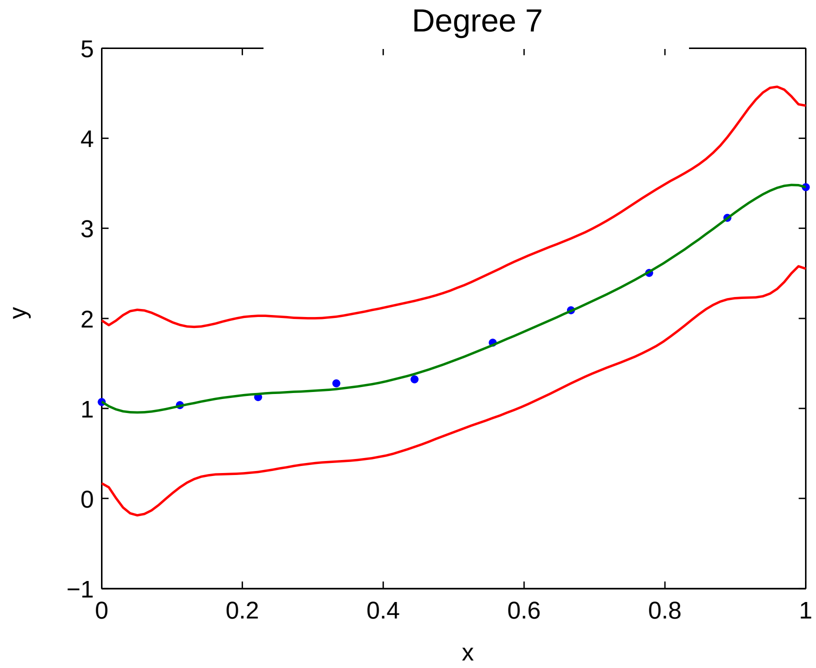


Model complexity and random scatter

Additional increase in complexity gives better fit, but the prediction ability become worse!

There exists an **optimal choice of model complexity** for a given data set size.





Model complexity and random scatter

Generally, in physical models, the interesting property may be a function of a large number of variables. A linear such model can be written:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon, \qquad \varepsilon \sim N\left(0, \sigma^2\right)$$

The Breiman/Freedman criterion for the optimal choice of model complexity is based on the expected prediction variance:

$$U_{n,p}^2 = s_{n,p}^2 \left(1 + \frac{p}{n - p - 1}\right)$$

where the expectation is taken over the ***random*** *x*-space

Model complexity and random scatter

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon, \qquad [\mathbf{X}, \varepsilon] \sim N([\boldsymbol{\mu}, 0], \boldsymbol{\Sigma})$$

When we have a physical phenomenon and choose a certain model complexity we actually partition this sum according to knowledge and usage:

$$Y = \sum_{i=1}^{p} \beta_i X_i + \underbrace{\sum_{j=p+1}^{m} \beta_j X_j + \underbrace{\sum_{k=m+1}^{\infty} \beta_k X_k}_{\varepsilon}}_{e}$$

where

$X_1 \ldots X_p$	are included
$X_{p+1} \ldots X_m$	are excluded
$X_{m+1} \ldots X_\infty$	are not measurable

Model complexity and random scatter

$$Y = \sum_{i=1}^{p} \beta_i X_i + \underbrace{\sum_{j=p+1}^{m} \beta_j X_j + \underbrace{\sum_{k=m+1}^{\infty} \beta_k X_k}_{\varepsilon}}_{e}$$

In fatigue design problems the logarithm of life can be seen as such a linear function of a large number of variables:

$$\begin{aligned}\ln N = &\ln \alpha + \beta_1 \ln \Delta S + \\ &+ \beta_{p+1} f_{p+1}(S_m) + \beta_{p+2} f_{p+2}(T) + \ldots + \beta_m f_m(freq) + \\ &+ \beta_{m+1} f_{m+1}(a_0) + \beta_{m+2} f_{m+2}(C) + \beta_{m+2} f_{m+2}(S_{op}) + \\ &+ \beta_{m+3} f_{m+4}(G_{konfig}) + \beta_{m+5} f_{m+5}(G_{orient}) + \beta_{m+6} f_{m+6}(HV_{local}) + \\ &+ \sum_{k=m+4}^{\infty} \beta_k X_k\end{aligned}$$

Included in the Wöhler curve concept

Excluded

Not measurable

Unknown

Model complexity and random scatter

$$Y = \sum_{i=1}^{p} \beta_i X_i + \underbrace{\sum_{j=p+1}^{m} \beta_j X_j + \underbrace{\sum_{k=m+1}^{\infty} \beta_k X_k}_{\varepsilon}}_{e}$$

The total errors in the model is the sum of model simplification errors and ”random” scatter (ε) .

The model simplification errors kan be decreased by a more precise model, but only variables with essential information should be included. **Optimal complexity depends on available data**.

The errors originating from unknown sources can be decreased by restriction of the application, i.e restricting to an area where the variation is small in

$$X_{p+1} \ldots X_{\infty}$$

Optimal complexity in fatigue modelling

$$\log N = f\left(a_0, a_C, \Delta\sigma, A, Y, \Delta\sigma_{w0}, \sigma_Y, k, \Delta K_{eff,th}, K_{op,\max}, K_{\min}\right)$$

A Taylor expansion gives an approximate linear function:

$$\log N - \log \mu_N \approx \sum_{i=1}^{p} \beta_i \left(X_i - \mu_{X_i}\right) + \varepsilon$$

Rough estimates of the variances of the X-variables and the use of the Breiman/Freedman criterion gives the following optimal model for the case with initiation&growth and around 30 reference tests:

$$N = \alpha_0 \cdot Y^{\beta_Y} \cdot \Delta\sigma^{\beta_{\Delta\sigma}} \cdot E$$

which is the Basquin equation extended with a geometry dependent variable.

Conclusions

The use of simple empirical models in industry can not be blamed on oldfasioned engineers, but rational arguments suggest that….

…the Wöhler curve, the Palmgren-Miner rule and the Paris law are often complex enough in engineering design!

Statistical methods like the Breiman/Freedman criterion may be very useful tools in engineering design and is also simple to use in each specific application.

Existant complex calculation methods in fatigue should be critically studied with respect to their prediction abilities.

Fraunhofer ITWM
CHALMERS
Research Centre
Industrial Mathematics

References

Breiman, L., Freedman, D., How many variables should be entered in a regression equation? *Journal of the American Statistical Association*, Vol. 78, No. 381, Theory and Methods Section, 1983, pp. 131-136.

T. Svensson, Model complexity versus scatter in fatigue, to appear in *Fatigue & Fracture of Engineering Materials and Structures*, 2004.