COMPARISON OF BAYESIAN AND OTHER APPROACHES TO THE ESTIMATION OF FATIGUE CRACK GROWTH RATE FROM 2D TEXTURAL FEATURES

Matej Mojzeš, Jaromír Kukal

Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, Praque, Czech Republic

e-mail: mojzemat@fjfi.cvut.cz

Hynek Lauschmann

Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Materials, Praque, Czech Republic

The fatigue crack growth rate can be explained using features of the surface of a structure. Among other methods, linear regression can be used to explain crack growth velocity. Non-linear transformations of fracture surface texture features may be useful as explanatory variables. Nonetheless, the number of derived explanatory variables increases very quickly, and it is very important to select only few of the best performing ones and prevent overfitting at the same time. To perform selection of the explanatory variables, it is necessary to assess quality of the given sub-model. We use fractographic data to study performance of different information criteria and statistical tests as means of the sub-model quality measurement. Furthermore, to address overfitting, we provide recommendations based on a cross-validation analysis. Among other conclusions, we suggest the Bayesian Information Criterion, which favours sub-models fitting the data considerably well and does not lose the capability to generalize at the same time.

Keywords: quantitative fractography, optimization, heuristic, linear regression, sub-model selection

1. Introduction

One of the tasks of quantitative fractography consists of modelling the relation between the fatigue crack growth rate (CGR) and textural features of images of fatigue fracture surfaces as explained in (Lauschmann and Siska, 2012; Nadbal *et al.*, 2008; Lauschmann *et al.*, 2006). For this purpose, e.g. a multilinear regression model (Sekeresova and Lauschmann, 2008; Kunz *et al.*, 2010) or a neural network may be used. Of these two, the neural network allows us to analyze and describe the structure of the obtained modeland better imagine the textural subset which is mutually related with the CGR (Lauschmann and Goldsmith, 2009).

The parameters of the respective regression model may be estimated using the least squares method. However, in real-world applications, the basic linear model is not flexible enough to fit the data. This can be solved by adding terms defined by non-linear functions of the basic features, e.g. logarithm, square root, etc. However, adding such features is soon limited by the given number of images.

According to Lauschmann and Goldsmith (2009), one possible way around this limitation is a two-phase stepwise regression with the first stage being bottom-up stepwise regression beginning with a constant model and terminating at a given overfitting level p_0. In each iteration, a new explanatory variable is included – the one which maximally decreases the sum of squares of residui. The second stage is top-down stepwise regression beginning with the final sub-model

from the first stage and terminating at the given final overfitting level p_F. In this procedure, an explanatory variable is selected for the elimination via the Wald test on a selected critical level.

While keeping in mind the relevant motivation for this problem, we suggest that instead of the stepwise regression, an alternative statistical approach based on the method of sub--model multiple testing may provide better results (Mojžeš *et al.*, 2012). There is a vast set of possible criteria that evaluate the quality of a given sub-model and are to be minimized. Further in the paper, we elaborate on the selection and assessment of some of the criteria. They are are interesting in the fractographic context, but may be applied generally to multi-parametric recognition as well.

2. Material and methods

2.1. Linear model

Denote by v_j the crack growth rate assigned to the j-th image of the fracture surface, and by f_{uj} the set of image features. The multilinear model in its basic form is based on the formula

$$\log_{10} v_j \approx c_0 + \sum_u c_u f_{uj} \tag{2.1}$$

Parameters c_u can be estimated by the least squares method. Since the linear model is not flexible enough to fit the data, we may add different non-linear functions of basic features and, therefore, modify the model to the following form

$$\log_{10} v_j \approx c_0 + \sum_q c_q h_q \tag{2.2}$$

where h are selected from an extended set of features containing the features f_u and a selection of their basic non-linear functions, e.g.

$$\{h_v\} \subset \mathbf{P} \cup \mathbf{Q} \cup \mathbf{R} \tag{2.3}$$

where

$$\mathbf{P} = \{f_u\} \qquad \mathbf{Q} = \{\log_{10} f_u, f_u^{-1}, f_u^{1/2}, f_u^2\}, \qquad \mathbf{R} = \{F_{uv}, F_{uv}^{-1}, F_{uv}^{1/2}, F_{uv}^2 \,|\, u > v\} \tag{2.4}$$

where $F_{uv} = f_u f_v$.

The next task will consist of defining a specific methodology on how to select and assess a distinct combination of explanatory variables, i.e. how to select the best sub-model from the extended feature set.

2.2. Sub-model selection

A sub-model should be considered a nested subset of the full model consisting of all the variables from the extended feature set. There are two extreme cases: the first is the full model and the second one corresponds to the constant model.

Let $n \in \mathbb{N}$ be the length of v (i.e. the number of observations), $m \in \mathbb{N}$ the extended feature set cardinality and $k \in \{0, 1, \ldots, m\}$ the number of explanatory extended feature set variables used in the sub-model. Also, let $\mathbf{c} = [c_0, c_1, \ldots, c_k]$ be the vector representing sub-model coefficients calculated solving Eq. (2.2). This vector can be divided into $\mathbf{c}_{red} = [c_1, c_2, \ldots, c_k]$ as its significant part and $\mathbf{c}_0 = [c_0]$ as the constant term coefficient.

Furthermore, we may express the sum of squares for the optimum $\mathbf{c}$ of a given sub-model as SSQ and SSQ_0 as the sum of squares for $\mathbf{c}_0$. Lastly, we will use the sub-model error defined as

$$s_e^2 = \frac{SSQ}{n - k - 1} \tag{2.5}$$

2.3. Data description

The methods developed in this paper will be applied to fatigue fracture surfaces of three laboratory specimens of the heat-resistant steel P92. Compact tension specimens (Fig. 1) (Lauschmann *et al.*, 2011) were loaded in air at 20°C by constant sinusoidal cycles with parameters of the external force according to Table 1. The loading frequency was lowered in steps from 13 Hz to 4 Hz in the final stage.
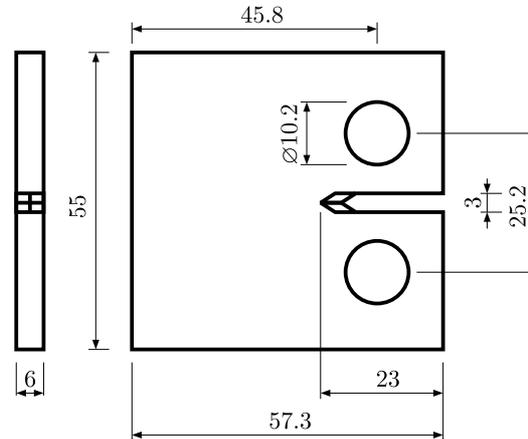

Fig. 1. Compact tension specimen

Table 1. External force parameters

Specimen	F_{min} [N]	F_{max} [N]
1	140	3300
2	2000	4800
3	3300	5500

The fatigue crack surfaces were documented using SEM with magnification 200×, real field of view was 0.6 mm×0.45 mm (examples in Fig. 2). The sequence of images was located in the middle of the crack surface along the same axis according to which the crack length was measured (Fig. 3). The recorded areas were mutually shifted by 0.4 mm. The direction of crack growth in the images was bottom-up. Digital representation in 1200×1600 pixels and 256 brightness values was used. The estimates of CGR were computed from frequently repeated records of the crack length. The course of the CGR related to the crack length was estimated, and every image was assigned a value of the CGR pertinent to its middle.
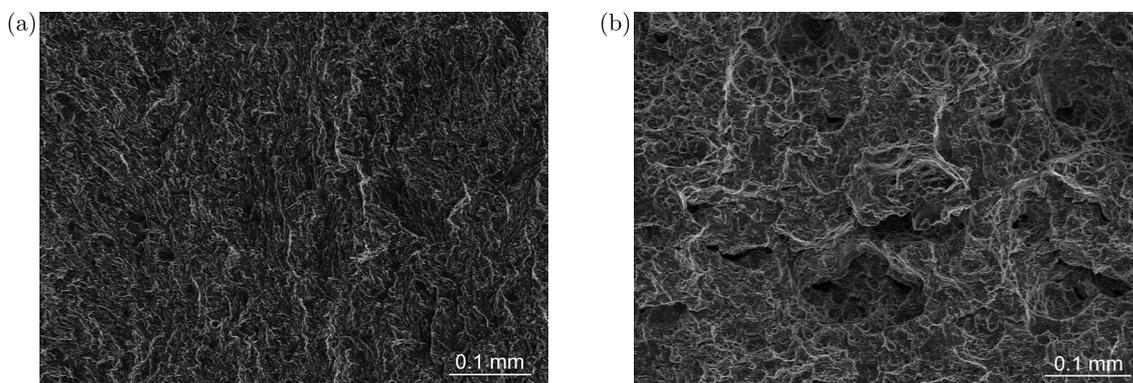

Fig. 2. Examples of SEM pictures of fracture surface; (a) low crack growth rate, (b) high crack growth rate
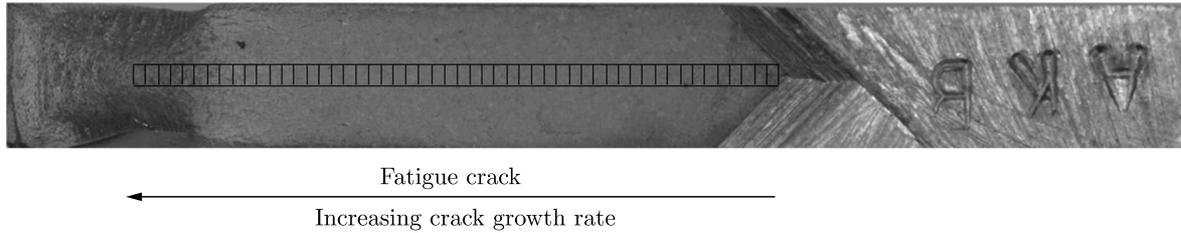
Fig. 3. Layout of snaps in crack surface (schematic plot)

For image textural features, energies of a 2D discrete wavelet transform were taken (Lausch-mann and Goldsmith, 2009). Decomposition using the Type 3 Daubechies wavelet at 8 levels was computed by Matlab function `wavedec2`. The energy is the mean square of wavelet coefficients for a given level and direction.

The basic sequence of features, $x_1, x_2, \ldots, x_{24}$, may be regarded as a set of $H_1, V_1, D_1, \ldots, H_8, V_8, D_8$ where H_j, V_j, D_j are wavelet decomposition energies at the j-th level in the horizontal, vertical and diagonal directions. The vector y represents decimal logarithm of the crack growth rate $y = \log_{10} v$.

The analysed data consisted of $n = 162$ observations and a total of 1224 features in the expanded feature set. It comprises:

- basic linear features $\mathbf{P}$, card($\mathbf{P}$) = 24,

- non-linear transformations of the basic features $\mathbf{Q}$, card($\mathbf{Q}$) = 96,

- dot product transformations of the basic features $\mathbf{R}$, card($\mathbf{R}$) = 1104,

as stated in (2.3).

To minimize potential numerical errors when working with the data, input data standardization was implemented as follows

$$x_k = \frac{h_k - \mathrm{E}h}{\sqrt{\mathrm{D}h}} \tag{2.6}$$

using $\mathrm{E}h$ and $\mathrm{D}h$ as the mean value and dispersion of the explanatory data.

Last, but not least – apart from the significance of the data, we can make use also of physical distribution of the data in the given data set, which is divided randomly into three separate groups. This will be especially useful when dealing with the cross-validation.

2.4. Selection heuristic

Searching for the best available sub-model is a binary optimization task that can be defined as minimization of the objective function $f \colon \mathbf{D} \to \mathbb{R}$ where

$$\mathbf{D} = \{\mathbf{x} \in \{0,1\}^m \mid \mathbf{0} \leqslant \mathbf{x} \leqslant \mathbf{1}\} \tag{2.7}$$

is the binary domain. Here, the binary vector $\mathbf{x}$ is directly representing utilization of the extended feature set, i.e. its components that are equal to "one" are included in the corresponding sub--model. Therefore, $\mathbf{0}$ refers to the constant model, and $\mathbf{1}$ to the full model.

Furthermore, suppose that we have an acceptable value of the objective function f^*. Then, we can define a set of solutions, the goal set, as

$$\mathbf{G} = \{\mathbf{x} \in \mathbf{D} \mid f(\mathbf{x}) \leqslant f^*\} \tag{2.8}$$

where

$$f^* \geqslant \min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{D}\} \tag{2.9}$$

For that purpose, we may utilize some of the well-known heuristic algorithms. We have chosen physically motivated Fast Simulated Annealing (FSA) (Kvasnička *et al.*, 2000) with reputable efficiency in the case of integer optimization tasks. FSA performs mutation on the ring neighbourhood

$$N(\mathbf{x}) = \{\mathbf{y} \in \mathbf{D} \,|\, \|\mathbf{y} - \mathbf{x}\|_1 = 1\} \tag{2.10}$$

Beginning with $k = 0$, $T_k > 0$ and the initial solution vector generated by uniform distribution $\mathbf{x}_0 \sim U(\mathbf{D})$, we perform FSA mutation as a uniformly generated random binary vector $\mathbf{y}_k \sim U(N(\mathbf{x}_k))$. Using $\eta_k \sim U([-1, +1])$, we set

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{y}_k & \text{for} \quad f(\mathbf{y}_k) < f(\mathbf{x}_k) + T_k \tan \dfrac{\pi\eta}{2} \\ \mathbf{x}_k & \text{for} \quad f(\mathbf{y}_k) \geqslant f(\mathbf{x}_k) + T_k \tan \dfrac{\pi\eta}{2} \end{cases} \tag{2.11}$$

until a solution from the goal set is found, or the pre-defined number of objective function evaluations is exhausted. The cooling strategy is represented by a non-increasing sequence of positive temperatures T_k.

We have been inspired by the increased efficiency of hybrid heuristics in the case of combination of a differential evolution and the steepest descent (Tvrdík J and Křivý, 2011). And since the previously defined set of optimization problems has many local minima, we have enhanced the FSA algorithm by a hybrid part – the steepest descent, which may increase the probability of reaching the global optimum.

In our approach to hybrid heuristic optimization, instead of $f(\mathbf{x})$ optimization, we optimize $g(\mathbf{x}) = f(\mathbf{h})$ where $\mathbf{x} = \mathbf{x}_0$, $\mathbf{h} = \mathbf{x}_H$ are the first and last members of any series $\{\mathbf{x}_k\}_{k=0}^{H}$ satisfying $\mathbf{x}_j \in N(\mathbf{x}_{j-1})$, $f(\mathbf{x}_j) < f(\mathbf{x}_{j-1})$ for $j = 1, \ldots, H$. This means $\mathbf{h}$ is the best solution in terms of the steepest descent heuristic. Before any problem solution vector is evaluated, its nearest local neighbourhood is iteratively searched for a better solution until no further advance in terms of the objective function can be made (or until a pre-defined maximum number of local evaluations is exceeded).

This way we were able to set a higher temperature T_0 and use a more benevolent cooling strategy. In other words, the algorithm is able to prevent getting stuck in a local minimum and still not lose the ability to fine-tune a given solution. Thus the FSA performance, on this specific task, has been improved.

2.5. Cross-validation

Having the data divided into three distinct groups allows us to perform strong cross--validation to assess how the results of a specific criterion will generalize to an independent data set.

We will perform the optimization on two out of three groups (training groups) and validate the analysis on the remaining third group (verification group). To improve the overall consistency, multiple rounds of cross-validation will be performed using different permutations of the data sets, and the verification results will be aggregated over the rounds.

As the goodness of fit measure we propose to use R as the correlation coefficient between the original data and the data proposed by the respective sub-model. However, during the optimization itself, we will use the original objective function based on the minimization of the *CRIT* value.

3. Theory of sub-model selection

At this point, we should choose some of many possibilities for testing sub-model quality. We have selected a few of them in accordance with (Mojžeš *et al.*, 2016) that can be divided in two sets, based on the concepts they are based on. The first one comprises traditional statistical tests and the criterion that will reflect the quality of a sub-model will be the logarithm of the p_{value}. On the other hand, the second set contains different statistical information criteria regarding selection of the model. In the latter case, we are simply minimizing the value of the selected information criterion.

3.1. Sub-model testing

Here, we test significance of a sub-model described by its $\mathbf{c}_{red}$. Respective hypotheses are defined as

$$H_0 : \quad \mathbf{c}_{red} = \mathbf{0}$$
$$H_1 : \quad \mathbf{c}_{red} \neq \mathbf{0}$$

and the R-square as well as the Wald test are used for their testing.

R-*square test*

In order to use R^2 for the analysis of the sub-model and constant model, according to the variance analysis (Wooldridge, 2002), we must define the stochastic variable F

$$F = \frac{SSQ_0 - SSQ}{SSQ} \frac{n - k - 1}{k} \tag{3.1}$$

Here F belongs to $F_{k,n-k-1}$ and the respective $p_{value} = 1 - F_{k,n-k-1}(F)$.

Wald test

Alternatively, to test sub-model hypotheses via the Wald test (Anděl, 1978), the variable Z has to be used

$$Z = \frac{1}{ks_e^2} \mathbf{c}^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{c} \tag{3.2}$$

The matrix $\mathbf{W}$ is a result of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ without the first row and column. Finally, Z has the distribution $F_{k,n-k-1}$ and $p_{value} = 1 - F_{k,n-k-1}(Z)$.

Lastly, for both tests, the resulting sub-model quality criterion can be defined as

$$CRIT = \log_{10} p_{value} \tag{3.3}$$

to be minimized. Due to the fact that values of p_{value} may get very close to zero, it is necessary to avoid potential numerical issues and express p_{value} in terms of an incomplete gamma distribution.

3.2. Information criteria

Another approach to the assessment of the sub-model quality is based on statistical information criteria. The selected criteria are presented in order from the least stringent one.

Wilks Information Criterion

Ralston and Rabinowitz (2001) according to Wilks (1962) recommend searching for a sub--model having the minimal error s_e^2. The corresponding logarithmic form, consistent with the following criteria, can be defined as

$$WIC = n \ln s_e^2 \tag{3.4}$$

In this basic criterion, k (i.e. the number of explanatory variables included in the sub-model) is already indirectly penalizing the information quality.

Akaike Information Criterion

Furthermore, an additional penalty for adding more explanatory variables is included in the Akaike criterion measuring relative goodness of the sub-model (Akaike, 1974) denoted as

$$AIC = WIC + 2k \tag{3.5}$$

Bayesian Information Criterion

Bayesian criterion (Schwarz, 1978) prevents overfitting even more by generating a stronger penalty for extra explanatory variables. Following the existing terminology, the criterion may be defined as

$$BIC = WIC + k \ln n \tag{3.6}$$

for $n \geqslant 8$.

As opposed to the logarithm of p_{value}, the final $CRIT$ to be minimized is directly equal to values of the information criteria.

4. Results and discussion

4.1. Heuristic optimization

In order to compare the results achieved with the hybrid heuristic to the stepwise alternative, we have implemented a traditional stepwise approach (Mojzeš *et al.*, 2012). Despite using only 125 variables (basic linear and non-linear transformations, **P** and **Q**) the heuristic sub-model optimization approach was superior to the stepwise approach based on the best correlation coefficient found.

Table 2 aggregates the best results achieved with the current, much more computationally demanding, data. In multiple runs of the heuristic optimization, we were able to obtain even better values of the correlation coefficient. Still, as the ultimate target should be the ability to generalize the independent data set, we will draw final conclusions based on the following cross-validation.

4.2. Cross-validation

The full data set has been divided into three groups of data, each having 59 (Group I), 53 (Group II) and 50 (Group III) observations. For each permutation of training and verification groups the hybrid heuristic optimize the sub-model to make the model fit the training data as well as possible according to the respective method. The same settings and conditions were used as in the case of the full data set without cross-validation. The detailed results are organized

Table 2. Optimal sub-model quality and features using hybrid heuristic

Method	$CRIT$	R	k_{opt}	Term								
				f_u	$f_u^{1/2}$	f_u^2	f_u^{-1}	$\log_{10} f_u$	F_{uv}	$F_{uv}^{1/2}$	F_{uv}^2	F_{uv}^{-1}
F-test	-117.92	0.9909	27	0	0	0	0	0	7	4	11	5
Wald test	-100.29	0.9839	15	0	0	0	1	0	1	3	7	3
WIC	-1099.00	0.9993	88	1	1	4	1	1	14	19	22	25
AIC	-927.03	0.9992	82	0	1	2	1	0	20	19	18	21
BIC	-626.86	0.9871	20	0	0	1	0	1	1	2	13	2

Table 3. Cross-validation results

Method	Training groups	$CRIT$	R_{train}	R_{verify}	k_{opt}
R^2 test	I+II	-86.72	0.9970	0.5937	34
R^2 test	II+III	-76.70	0.9900	0.5785	13
R^2 test	I+III	-79.29	0.9924	0.9594	22
Wald test	I+II	-74.29	0.9867	0.8624	10
Wald test	II+III	-71.34	0.9848	0.6661	5
Wald test	I+III	-66.24	0.9854	0.9476	13
WIC	I+II	-736.29	0.9991	0.4471	55
WIC	II+III	-804.83	0.9997	0.5214	61
WIC	I+III	-716.25	0.9993	0.8286	63
AIC	I+II	-619.41	0.9988	0.8292	48
AIC	II+III	-674.31	0.9996	0.1351	56
AIC	I+III	-584.22	0.9987	0.9229	48
BIC	I+II	-479.12	0.9984	0.6896	45
BIC	II+III	-500.70	0.9995	0.3800	53
BIC	I+III	-430.91	0.9926	0.9574	21

in Table 3. The most important results are in the column of the correlation coefficient R_{verify} which measures the quality of the fit on the verification data set.

These results are aggregated using the mean of respective methods and, furthermore, expanded by comparing the data composed from distinct verification data sets to the original one in Table 4. Also, the results of composed cross-validation are depicted in Fig. 4 as log-log plots of measured and predicted CGRs. The predicted data are distinguished by group symbols $(\times, +, \circ)$. The results of WIC and AIC are the worst ones, since the outliers had to be omitted to enable the plotting. As can be seen, the Wald approach and BIC criterion offer the best results.

Table 4. Cross-validation summary

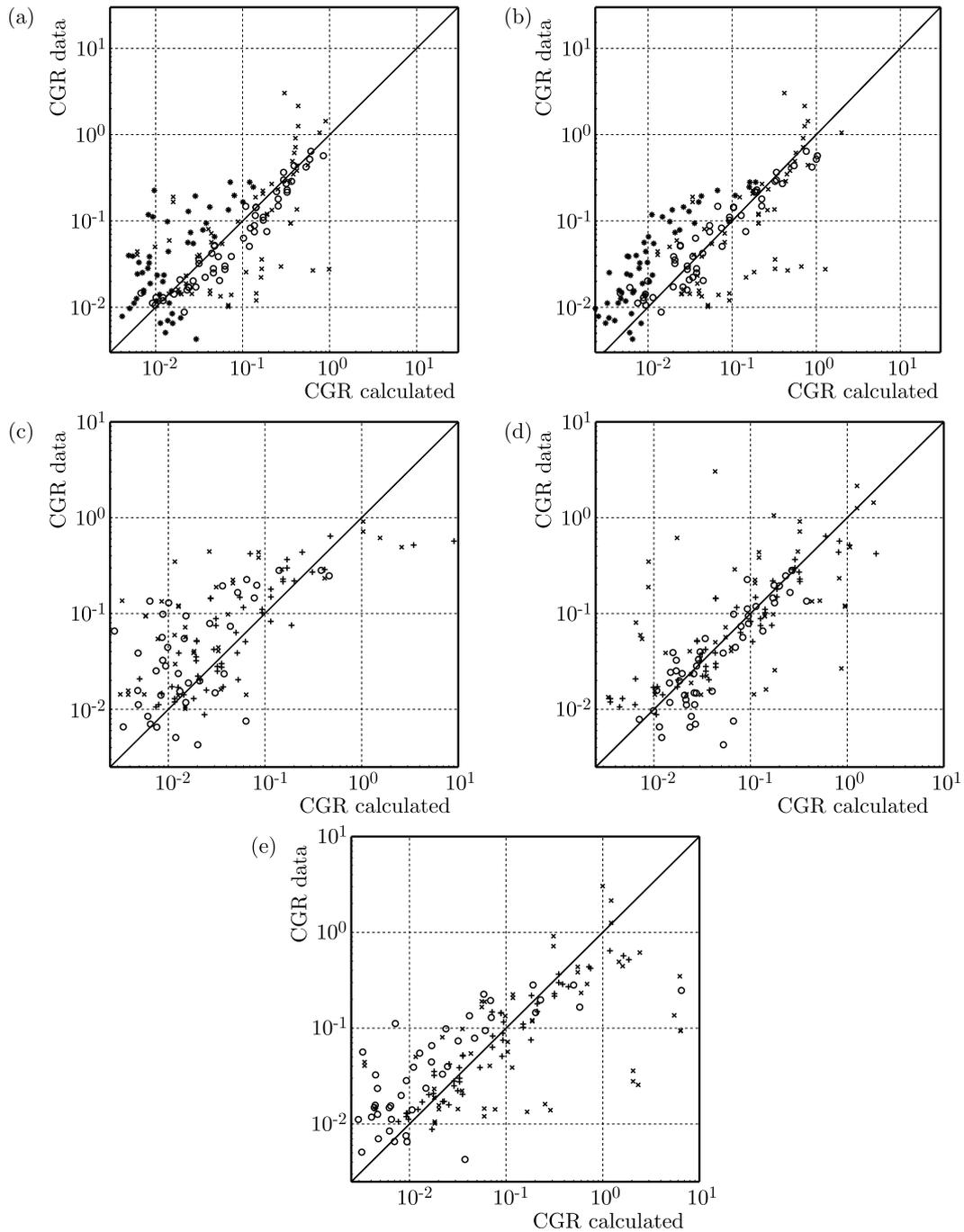| Method	Mean R	Composed R
R^2 test	0.7105	0.7043
Wald test	0.8254	0.7702
WIC	0.5990	0.5320
AIC	0.6291	0.3281
BIC	0.6757	0.6096

Fig. 4. Prediction of Crack Growth Rate (CGR) using various techniques of sub-model selection
($\times$ representing data in group I, $+$ in group II, and $\circ$ in group III)

5.   Conclusions

The benefits of the solution described above are considerable. Nearly an unlimited set of explanatory variables may be offered without any respect to the original number of observations in a given case. Very good models have been obtained also in previously unsolvable cases with a very small number of observations.

Of course, the final result is mostly dependent on the approach of sub-model selection. As it is apparent from the results of cross-validation, and also based on our experience, we are recommending BIC, Wald test and potentially also R^2 test and WIC. Nevertheless, there are significant differences between these four and, more specifically, we are suggesting:

- BIC as a universal criterion,
- Wald test as a well balanced criterion, similar to BIC,
- R^2 test as a legitimate criterion with respect to the variance analysis approach,
- WIC as a criterion that leads to considerable adherence to the data, however, as opposed to other criteria, lacks the ability of generalization.

References

1. AKAIKE H., 1974, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716-723

2. ANDĚL J., 1978, *Mathematical Statistics* (in Czech), SNTL/Alfa, Praha

3. KUNZ J., KOVARIK O., LAUSCHMANN H., SIEGL J., AUGUSTIN P., 2010, Fractographic reconstitution of fatigue crack growth in integrally stiffened panels, *FATIGUE 2010*, **2**, 1711-1720

4. KVASNIČKA V., POSPÍCHAL J., TIŇO P., 2000, *Evolutionary Algorithms* (in Slovak), STU Bratislava

5. LAUSCHMANN H., GOLDSMITH N., 2009, Textural fractography of fatigue fractures, [In:] *Fatigue Crack Growth: Mechanics, Behavior and Prediction*, A.F. Lignelli (Edit.), Nova Science Publishers, Inc., 125-166

6. LAUSCHMANN H., KUNZ J., KOVAŘÍK O., 2011, Analyzing crack growth rate in steel P92 (in Czech), Research report V-KMAT-839/11, FNSPE CTU Prague

7. LAUSCHMANN H., SIEGL J., SUMBERA J., SISKA F., NEDBAL I., 2006, An unifying concept for fatigue: The reference crack growth rate, *Materials Characterization*, **56**, 257-265

8. LAUSCHMANN H., SISKA F., 2012, The reference texture: A proposal of a physical explanation, *International Journal of Fatigue*, **43**, 120-127

9. MOJZEŠ M., KLIMT M., KUKAL J., 2016, Feature selection via competitive levy flights, *2016 International Joint Conference on Neural Networks (IJCNN)*

10. MOJZEŠ M., KUKAL J., LAUSCHMANN H., 2012, Sub-model testing in fractographic analysis, *Proceedings of Mendel 2012 Soft Computing Conference*, Brno University Technology Press, 350-355

11. NEDBAL I., LAUSCHMANN H., SIEGL J., KUNZ J., 2008, Fractographic reconstitution of fatigue crack history – Part II, *Fatigue and Fracture of Engineering Materials and Structures*, **31**, 177-183

12. RALSTON A., RABINOWITZ P., 2001, *A First Course in Numerical Analysis*, Courier Dover Publications

13. SEKERESOVA Z., LAUSCHMANN H., 2008, Multi-fractal features of fatigue crack surfaces in relation to crack growth rate, *Materials Structure and Micromechanics of Fracture V*, **567-568**, 129-132

14. SCHWARZ G., 1978, Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-464

15. TVRDÍK J., KŘIVÝ I., 2011, Hybrid adaptive differential evolution in partitional clustering, *Proceedings of 17th International Conference on Soft Computing, Mendel 2011*, VUT Brno, Brno, 1-8

16. WILKS S.S., 1962, *Mathematical Statistics*, John Wiley & Sons, Inc., New York

17. WOOLDRIDGE J.M., 2002, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press