Statistical Methods for Helicopter Preliminary Design and Sizing

Max Lier

DLR (German Aerospace Center), Institute of Flight Systems
Lilienthalplatz 7, D-38108 Braunschweig, Germany
max.lier@dlr.de

The present paper focuses on the applicability of statistical methods to helicopter design. Firstly the structure and statistical dependencies of a database of 150 existing helicopters are investigated by means of principal component and correlation analysis. The multivariate regression method presented is capable of automated computation of regression functions for a variety of input and output parameters. Additionally a minimum degree of complexity of the regression function is estimated by hypothesis testing. In contrast to most of the approaches used in literature a polynomial regression model was chosen in this paper. The regression result can be improved by using a partial data set, which can be extracted using manually defined criteria or – statistically motivated and unsupervised – by clustering. Subject to the underlying database relative errors of less than 10% for certain design parameters are achievable – allowing for a suitable application in helicopter preliminary design.

INTRODUCTION

Building upon the knowledge gathered in recent years concerning fixed-wing aircraft preliminary design [1], DLR is currently developing an integrated and automated tool for helicopter preliminary design and evaluation. In this context the applicability of statistical methods on helicopter preliminary sizing and design has been studied.

The data available for the preliminary sizing of helicopters is generally very limited since only the mission and performance specifications are available at the beginning of the design process. In many cases conceptual studies are therefore only based on the experience of the design engineer. However, the results of the early design stages have a great influence on the subsequent design process.

The use of statistical methods in the context of helicopter design has infrequently been covered in literature. Recent contributions were made by Rand and Khromov in 2002 [2] or by Kim and Oh in 2007 [3], both suggesting potential functions for regression. Computational methods for helicopter design not using statistics exclusively are conventionally designed using iterative algorithms (e.g. [4], [5]).

This paper studies the applicability of various statistical methods to the early stages of helicopter design. Physics-based methods are left out deliberately in order to find out if fundamental physical relationships can be reproduced using sole statistical methods. Consequently the design process is based on the customer specification of maximum speed, range and payload only. At worst this is the only information available to the design engineer at the beginning of the design task.

HELICOPTER DATABASE

An expansive database of existing helicopters is a cardinal prerequisite for the successful application of statistical methods. In fact data collection likely makes up the largest part of the amount of work required to utilize statistics in preliminary design. The database used in this paper contains a variety of different parameters of 159 (conventional) helicopters, among them main and tail rotor characteristics and dimensions as well as engine, performance and mass data. It was compiled using various sources, mainly [6] and [7]. The studies described in this paper use a subset of the database with 16 design parameters. As some of the methods used for the statistical evaluation require fully populated matrices the number of helicopters is reduced to 81 in this subset. The range of parameters is shown in Table 1.

Correlation Analysis

The aim of correlation analysis is to measure associations and dependencies between two measured variables. Those dependencies are usually quantified by correlation coefficients, the most common of which is named by its developer Karl Pearson. Alas, one major drawback of the Pearson correlation coefficient is its ability to only detect linear relations between variables. Therefore the commonly known Kendall Tau rank correlation coefficient [8] is used here, which only

parameter	symbol	unit	minimum	maximum	median
main rotor radius	r_{MR}	m	3.6	17.5	6.4
number of main rotor blades	b_{MR}		2	8	4
main rotor rotational velocity	Ω_{MR}	rad/s	12.6	55.5	36.0
tail rotor radius	r_{TR}	m	0.3	3.8	1.1
number of tail rotor blades	b_{TR}		2	13	2
length (fuselage)	l	m	6.3	33.7	12.2
width (landing gear)	w	m	1.6	7.6	2.5
height (overall)	h	m	2.4	11.6	3.8
number of engines	N		1	3	2
takeoff power	P_{TO}	kW	108	14914	962
empty mass	m_E	kg	383	28200	2204
fuel mass	m_F	kg	58	9600	590
maximum takeoff mass	m_{TO}	kg	621	49600	3561
maximum speed	V	km/h	139	365	259
maximum fuel range	R	km	213	1204	604
payload	m_{PL}	kg	139	20000	826

Table 1: *Range of parameters in the data set*

r_{MR}															
0,37	b_{MR}														
-0,82	-0,35	Ω_{MR}													
0,80	0,40	-0,71	r_{TR}												
0,36	0,43	-0,37	0,27	b_{TR}											
0,83	0,42	-0,72	0,77	0,38	l										
0,64	0,35	-0,63	0,66	0,25	0,57	w									
0,76	0,45	-0,66	0,73	0,41	0,78	0,61	h								
0,49	0,51	-0,41	0,51	0,35	0,55	0,44	0,55	N							
0,73	0,49	-0,61	0,73	0,40	**0,80**	0,55	0,75	0,64	P_{TO}						
0,78	0,49	-0,68	0,77	0,40	**0,87**	0,59	**0,81**	0,60	**0,86**	m_E					
0,72	0,48	-0,63	0,72	0,44	0,80	0,54	0,77	0,64	**0,86**	**0,84**	m_F				
0,77	0,50	-0,67	0,77	0,41	**0,86**	0,59	**0,82**	0,60	**0,88**	**0,94**	**0,87**	m_{TO}			
0,25	0,29	-0,21	0,26	0,29	0,38	0,10	0,32	0,43	0,42	0,39	0,43	0,41	V		
0,05	0,11	-0,05	0,02	0,27	0,12	0,05	0,08	0,24	0,13	0,12	0,20	0,12	0,31	R	
0,71	0,47	-0,59	0,66	0,42	0,74	0,54	0,77	0,56	0,79	0,79	0,77	**0,84**	0,37	0,10	m_{PL}

strong correlation	$|\tau| \geq 0.8$
moderate correlation	$0.5 \leq |\tau| < 0.8$
weak correlation	$|\tau| < 0.5$

Table 2: *Kendall Tau Rank Correlation Coefficients*

measures the tendency of a variable to increase if another variable does likewise increase. It can adopt values between -1 and 1. Considering n observations of two variables φ and ψ as well as all possible (different) pairs of observations $(\varphi_i, \psi_i), (\varphi_j, \psi_j)$ the Kendall tau correlation coefficient is defined as

$$\tau = \frac{N_C - N_D}{\frac{1}{2} n(n-1)}$$

where N_C is the number of concordant pairs, for which

$$\left(\varphi_i < \varphi_j \wedge \psi_i < \psi_j\right) \vee \left(\varphi_i > \varphi_j \wedge \psi_i > \psi_j\right)$$

and N_D is the number of discordant pairs, for which in contrast

$$\left(\varphi_i < \varphi_j \wedge \psi_i > \psi_j\right) \vee \left(\varphi_i > \varphi_j \wedge \psi_i < \psi_j\right).$$

Table 2 shows the correlation coefficients of the data set described above. Fundamental physical relations are clearly mirrored by the correlation analysis. There is a strong inversely varying relationship between main rotor radius and rotational speed leading to a bounded blade tip velocity in order to reduce compressibility effects. The physical dimensions are related to each other, the predominant relationship between main rotor radius and fuselage length is very plausible. The mass variables are related among themselves and form the major component of the takeoff power, which can as easily be comprehended.

In regard to helicopter design it is clearly visible that speed and range are ill-suited for statistical evaluation as they show only weak correlation values to all other variables. Therefore it has to be kept in mind that the regression analysis following is also based on a poor statistical basis for these two input variables.

Principal Component Analysis

The Principal Component Analysis (PCA, e.g. [9]) is in general used to transform a data set of (possibly) dependent variables into a set of independent ones. Those independent variables (principal components) are linear combinations of the original variables. Under certain circumstances it is also possible to use PCA to reduce the complexity of the data set. In particular, if few principal components describe the majority of the variance of the data set and those components are linear combinations of again only few of the original variables, it would be possible to describe the main features of the data set by using only

those variables thus simplifying the data structure significantly.

The covariance matrix Σ_X of any data set X holds the variances of the data set at its main diagonal. The remainder of the entries can be interpreted as redundancies of the data set. Hence a diagonalization of Σ_X will lead to an optimal representation of the total variance. Given a standardized data matrix $\hat{X}$ an orthonormal basis P is required, such that the covariance matrix of the transformed data set $Y = \hat{X}P$ is diagonal. Now,

$$\underline{\underline{\Sigma}}_Y = \frac{1}{n-1} Y^T Y = \frac{1}{n-1} \left(\hat{X}P\right)^T \left(\hat{X}P\right)$$

$$= \frac{1}{n-1} P^T \hat{X}^T \hat{X}P = P^T \left(\frac{1}{n-1} \hat{X}^T \hat{X}\right) P$$

and therefore

$$\Sigma_Y = P^T \Sigma_{\hat{X}} P .$$

Thus P is the matrix of eigenvectors of $\Sigma_{\hat{X}}$, which in turn are the linear coefficient vectors of the principal components. Furthermore it can be shown that the ratio of the specific eigenvalue and the sum of all eigenvalues of the principal component concerned is equivalent to the fraction of the total variance described by this component.
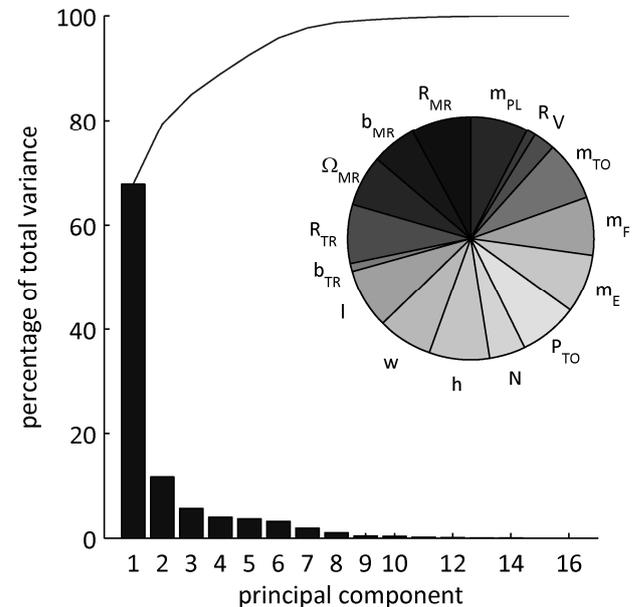
Figure 1: *Pareto diagram of the percentage of total variance described by the principal components and pie chart of the composition of the first principal component*

Figure 1 shows the percentage of the total variance described by the 16 principal components. It is

clearly visible that the first component accounts for about two thirds of the total variance. Unfortunately this cannot be exploited to reduce the complexity of the data set as this very component is nearly equally dependent on all original variables. In accordance with the result of the correlation analysis the performance values and blade numbers represent the smallest fractions of the first principal component. As a consequence a multitude of variables is necessary to universally describe a helicopter. The use of the maximum takeoff mass as a single design parameter will, albeit widely used, statistically lead to suboptimal results.

MULTIVARIATE REGRESSION

The studies presented in this paper are based on a regression algorithm which is capable of automated computation of regression functions for arbitrary input and output parameters. Additionally a minimum degree of complexity of the regression function is estimated by hypothesis testing. In contrast to most of the approaches used in literature a polynomial regression model was chosen in this paper. Polynomial regression can be carried out without iterative calculations and provides a high level of flexibility.

Methodology

The well-known method of least squares estimates an approximate solution of overdetermined systems and is widely used for data fitting problems. Given a data set

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

of n observations of a dependent variable y and n observations x_{ij} of k independent variables $\xi_1 \dots \xi_k$ and considering a linear regression function with k coefficients b_i

$$\tilde{y}(\xi_1, \dots, \xi_k) = b_0 + b_1\xi_1 + \dots + b_k\xi_k$$

the sum of squared residuals is

$$S_\varepsilon = \sum_{i=1}^{n} (y_i - \tilde{y}(x_{i1}, \dots, x_{ik}))^2 .$$

By finding the roots of the partial derivatives of S_ε with respect to the dependent variables the minimum sum of squared residuals can be obtained. Hence, the vector of coefficients can be calculated with

$$b = (Z^T Z)^{-1} Z^T y$$

where

$$Z = \begin{bmatrix} 1 \\ \vdots & X \\ 1 \end{bmatrix} .$$

Polynomial regression models can be implemented by substituting every monomial by a single independent variable and expanding the data matrix accordingly. For polynomial regression models (allowing the coupling of input variables) of the degree g and k independent variables of the form

$$\tilde{y}(\xi_1, \dots, \xi_k) = b_0 + \sum_{e=1}^{g} \sum_{j=1}^{k} b_i\xi_j^e + \sum_{p=1}^{k} \sum_{\substack{q=2 \\ q>p}}^{k} \prod_{\substack{a,b>0 \\ a+b\leq g}} b_i\xi_p^a\xi_q^b$$

the number of monomials (and thus the number of input variables for the linear regression problem) is

$$m = \binom{g+k}{g} = \frac{(g+k)!}{g!\,k!} .$$

Hence the applicability of the algorithm is affected by the size of the data set available. If the number of observations (here: helicopters) is lower than the number of monomials the system is not overdetermined, $Z^T Z$ becomes singular and can therefore not be inverted.

The maximum degree of the regression model that can be obtained depends on the number of observations and the number of independent variables. Figure 2 shows that the number of observations needed is rapidly rising with an increasing number of independent variables. Thus higher order relationships cannot be taken into account if the size of the data set is limited.
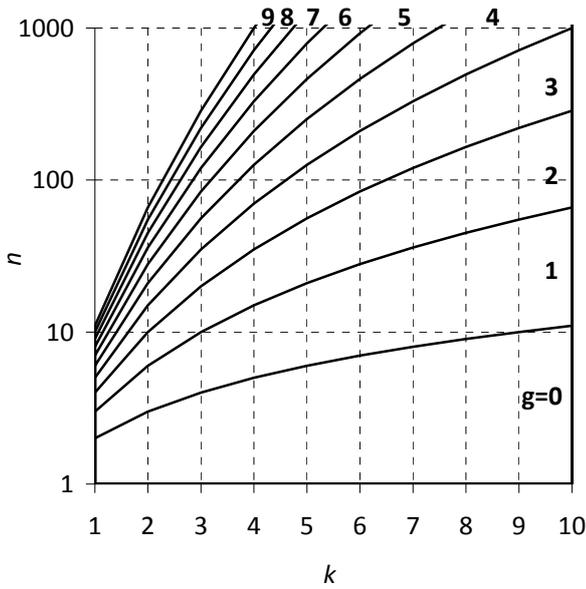
Figure 2: *Maximum degree of the regression polynomial* ***g*** *depending on the number of observations* ***n*** *and independent variables* ***k***

There are a number of requirements for the regression function. The data set should be described by the regression function as accurate as possible. For the sake of calculation effort a minimum complexity of the regression function leading to acceptable results could be desirable. This becomes even more relevant if the regression function is used for exploratory studies and must hence be evaluated countless times.

Regression analysis can lead to very small coefficients which do not have a substantial influence to the overall result but can considerably increase the complexity of the regression function. By neglecting those coefficients a compromise between the conflicting objectives (accuracy and simplicity) can be found.

Hypothesis testing is a common way to determine if a coefficient is zero with respect to a defined significance level and can thus be neglected. For every coefficient the null hypothesis (the coefficient is zero)

$$H_0 : b_i = 0.$$

is tested against the alternative hypothesis (the coefficient if not zero)

$$H_1 : b_i \neq 0$$

using the test statistic

$$t_0 = \frac{b_i}{\sqrt{\mathrm{var}(b_i)}}$$

where the variance of the coefficient is the corresponding element of the main diagonal of the data set's covariance matrix

$$\mathrm{var}(b_i) = \Sigma_{ii} = \left[\frac{S_\varepsilon}{v} \left(Z^T Z \right)^{-1} \right]_{ii} .$$

The number of degrees of freedom v of the regression problem can easily be calculated as the difference of the number of observations (helicopters) n and the number of coefficients m

$$v = n - m .$$

Given a significance level (probability of incorrectly rejecting the null hypothesis) α the null hypothesis is rejected if the test statistic is bigger than the critical value of the $\left(1 - \frac{\alpha}{2}\right)$-quantile of Student's t-distribution [10] $t_{\alpha/2}$ (see figure 3).
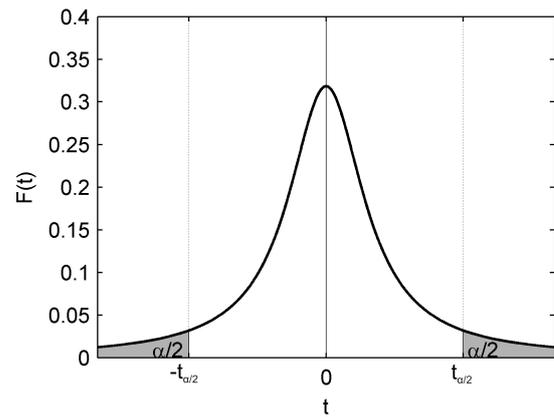
Figure 3: *Student's t-distribution with significance level* ***α***

This method is remarkably useful for the chosen polynomial regression model. There is no a-priori-knowledge about the natural degree of the polynomial best describing the data set. Eventually a data set with $p + 1$ observations can always be exactly modelled by a p-th order polynomial (assuming an injective relation).

The algorithm used here gradually increments the degree of the polynomial model and uses the t-test described above to afterwards simplify the coefficient vector by neglecting the coefficients for which the null hypothesis is not rejected. If all coefficients belonging to monomials of the highest and second highest order have been set to zero the algorithm stops since the last two increases of the degree of the polynomial did not improve the result. The regression function resulting in the

smallest sum of residuals so far is considered the best regression function found.

Results

The application of the described algorithm to the helicopter data set results in 13 regression functions for the dependent variables, which are summarized in table 3. Although the number of helicopters in the data set (81) would allow for a fifth order polynomial with 56 monomials, the null hypothesis approach effectively reduces the number of coefficients resulting in first and second order polynomials with two to five monomials. The significance level was set to 0.1 for all subsequent calculations. Almost every variable is dependent on the payload being the only input variable with second order relations as well. Speed appears in all but 4 equations while range occurs in less than half of the regression functions.

variable	constant term	dependent on			number of monomials
		V	R	m_{PL}	
r_{MR}	✓			✓	2
b_{MR}	✓	✓		✓	3
Ω_{MR}	✓			✓	2
r_{TR}	✓	✓		✓	3
b_{TR}		✓	✓		2
l	✓	✓		✓	3
w	✓	✓	✓	✓	5
h	✓	✓	✓	✓	5
N		✓	✓	✓	3
P_{TO}			✓	✓	3
m_E		✓		✓	2
m_F		✓	✓	✓	3
m_{TO}				✓	2

Table 3: *Summary of the regression function obtained using the data set of 81 helicopters*

In order to quantify how well the data set is represented by the regression functions the mean absolute relative error over the whole data set was calculated by comparing the values computed using the regression functions with the original helicopter data. Figure 4 (black bars) shows that the mean error ranges from 10 to 45 per cent. The physical dimensions are captured best while the estimated values of masses and takeoff power are less accurate. The large deviation of the number of tail rotor blades can easily be explained by the presence of helicopters with conventional tail rotors as well as fenestrons in the data set. Fenestrons consist of eight to thirteen blades while blade numbers of conventional tail rotors do not exceed five.
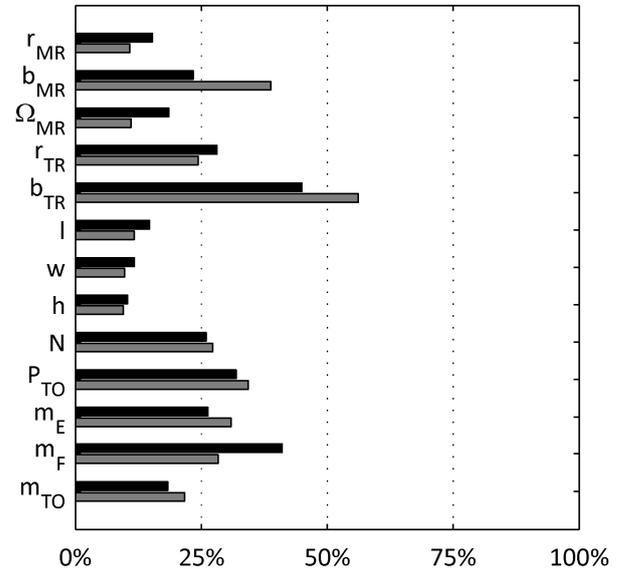
Figure 4: *Mean absolute relative errors of the regression function obtained using the full data set (81 helicopters, black bars) and the manually reduced data set (36 helicopters, grey bars)*

As the range of helicopters in the data set is still widespread in terms of the dependent variables the result could probably be improved by extracting a smaller data set, which contains helicopters of a similar class. As an initial approach the reduction of the data set can be done manually. Exemplarily a subset was extracted containing helicopters with a payload centred around the Eurocopter EC 135 payload of 785 kg. The upper and lower bounds of the subset payload where determined such that 36 helicopters (being the minimum number of observations to allow for 4th order polynomials) remain in the subset. The regression functions were obtained for this subset using the same algorithm. The mean absolute errors are shown in figure 4 (grey bars). Compared to the global regression the errors are of a similar magnitude. The estimations of the physical dimensions are slightly better whereas the discrete variables (number of blades and engines) and most of the mass parameters show even larger errors. Using payload as the sole datum for the extraction of a data subset does not account for any differences in structural or system weights. Using

the maximum takeoff weight instead would obviously produce better results (as can also be seen by comparison of the correlation coefficients for payload and takeoff mass, see table 2). However, as the takeoff weight is not known by a minimal design specification (consisting of payload, range and speed only) it can not be used in this case.

DATA CLUSTERING

Based on the results described in the previous section statistical methods for the extraction of a suitable subset were studied.

Clustering Method

The k-means algorithm [11] was selected for clustering the data due to its straight-forward implementation and well-established use in science. Given a data set of n observations x_i the algorithm tries to partition the data set into k clusters by minimizing the distance between the data points and the corresponding cluster centroids μ_j:

$$\operatorname*{argmin}_{S_j} \sum_{j=1}^{k} \sum_{x_i \in S_j} d(x_i, \mu_j).$$

A variety of metrics can be used to obtain the distances and cluster centroids, which in turn have a considerable influence on the result. Four different metrics have been compared in this study:

a) L$_2$ (Euclidean) distance
The most common distance metric is the Euclidean distance

$$d = \|x_i - \mu_j\|.$$

b) L$_1$ (Manhattan) distance
Using the absolute differences along the coordinate axes as a distance measure results in the L$_1$ metric
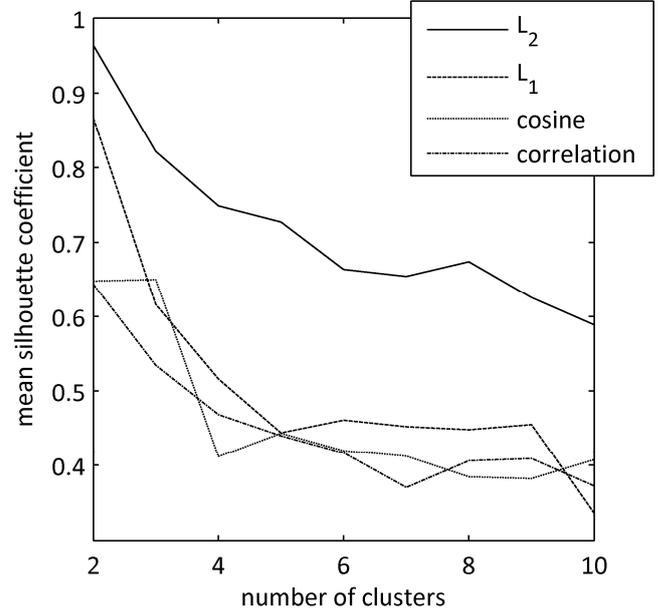
$$d = |x_i - \mu_j|.$$

c) Cosine distance
The cosine distance metric is based on the angle between two observations which can be calculated using the dot product

$$d = \cos(\angle(x_i - \mu_j)) = \frac{x_i \cdot \mu_j}{\|x_i\| \, \|\mu_j\|}.$$

d) Correlation distance
The correlation distance uses the (Pearson) correlation coefficient r to define the distance

$$d = 1 - r(x_i, \mu_j).$$

Figure 5: *Mean silhouette coefficients for different numbers of clusters and distance metrics*

The k-means algorithm requires the number of clusters to be preset. The silhouette coefficient is an appropriate measure to determine the natural number of clusters of a given data set. It can be determined for every observation and is defined as

$$s = \begin{cases} 1 - \frac{a}{b} & \textit{if } a < b \\ \frac{b}{a} - 1 & \textit{if } a \geq b \end{cases}$$

where a is the average distance of the observation to the other points of the cluster and b is the minimum average distance to the points of the other clusters. Thereby the silhouette coefficient combines cluster cohesion (the similarity of the cluster and the observations within, a) and the cluster separation (the dissimilarity of the clusters to each other, b). The mean value of s can be used to measure how appropriate the data has been clustered. Figure 5 shows that the mean silhouette value is mostly decreasing with an increasing number of clusters indicating that only few clusters are existent in the data set. Using only two clusters is nonetheless inappropriate due to the fact that the algorithm plainly sorts out the three or four heaviest helicopters leaving the vast majority of the data set in the second cluster. However, the number of clusters cannot be
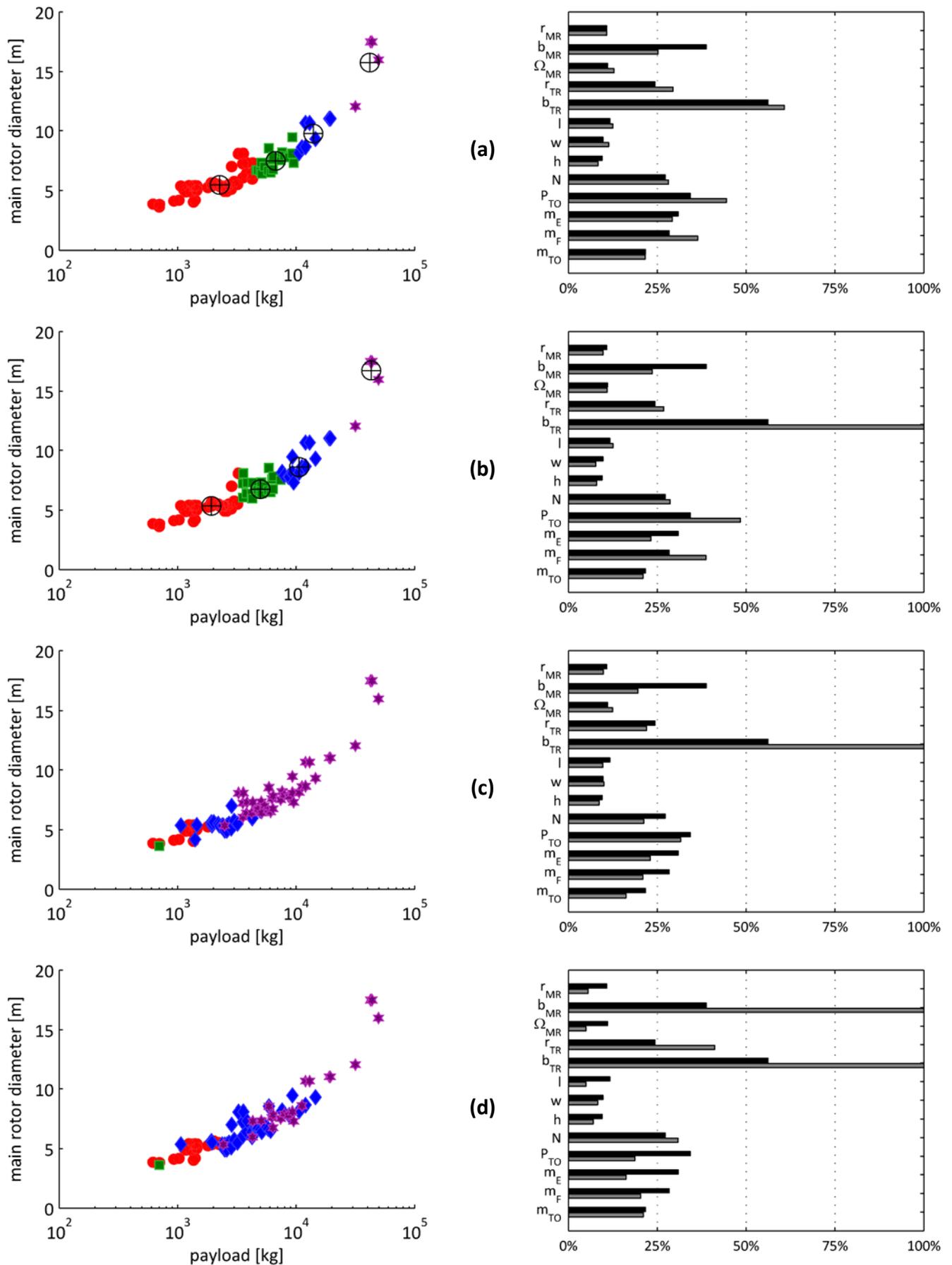
Figure 6: *Clusters obtained using L₂ (Euclidean) distance (a), L₁ (Manhattan) distance (b), cosine distance (c) and correlation distance (d) metrics with centroids where applicable (crosshairs) as well as the corresponding mean errors of the regression functions determined within the cluster containing the Eurocopter EC 135 (grey bars) compared to the manual EC135-centred subset (black bars)*

increased freely, because the number of observations within each cluster is rapidly decreasing, eventually rendering the regression algorithm useless. Comparing the distance metrics, the Euclidean distance yields a much better grouping of the clusters than the other metrics.

Results

These considerations in mind the data set has been clustered into four clusters leading to average cluster sizes of about twenty helicopters. Figure 6 shows the distribution of the clusters (by way of example on a plane of payload and main rotor diameter). The L_1 and L_2 metrics mainly divide the data set into clusters of helicopters with similar mass parameters, whereas the other metrics incorporate other variables to a much higher degree. Nevertheless, none of the metrics is capable of grouping the different anti-torque devices together as one would probably prefer if the clustering would be done by hand.

Comparing the resulting regression functions for the clusters containing the Eurocopter EC 135 as an example (see figure 6 on the right) to the ones obtained by the manually extracted (payload-bounded) subset there are no significant improvements. The cosine distance metric leads to slightly better estimations only, although of almost all variables. All but the Euclidean metric cause the regression algorithm to eliminate all coefficients of some regression functions resulting in a constant zero estimation and thus logically generating an error of 100%. This is mainly attributable to the small size of the resulting clusters. The significance level should be gradually increased with decreasing size of the data set in order to obtain regression functions with reasonable levels of complexity. Yet this is avoided here for the sake of comparability.

APPLICATION

Although the effect on the average errors over a subset is small, clustering can lead to significant improvements in regard of a single helicopter. Figure 7 shows the errors for the estimation of the Eurocopter EC 135 data with the regression functions obtained using different data sets. For most of the design variables the results obtained using the clustered or reduced subsets show a significant improvement. The majority of variables is estimated with an error of about 10% or less using the cluster determined by the k-means
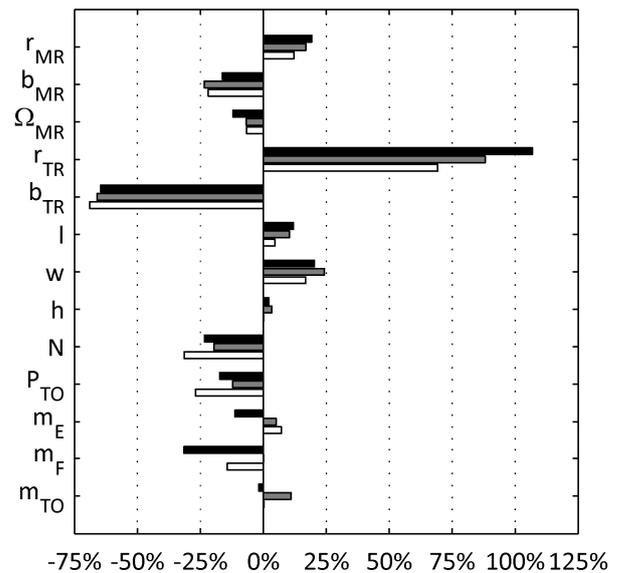
Figure 7: *Relative errors of the estimated values obtained by the regression functions for the whole data set (black bars), the manually reduced data set (grey bars) and the cluster determined by the k-means algorithm in conjunction with the Euclidean distance metric (white bars), each compared to the real Eurocopter EC 135 data*

algorithm. Height and maximum takeoff mass even show errors of less than 1%. The major discrepancy regarding the tail rotor cannot be solved as there are still helicopters with conventional tail rotors and fenestrons present.

SUMMARY AND CONCLUDING REMARKS

A data set of 81 helicopters was studied. Speed, range and payload have been selected as independent variables being the essential part of the customer specification. The 13 design variables selected show considerable differences in their statistical properties. The mass properties and physical dimensions show the strongest correlation to other variables. The performance parameters (speed and range) are only weakly related to the data set. This leads to a poor basis for the estimation of design parameters using the customer specification.

The regression algorithm presented is able to automatically determine regression functions of an appropriate complexity. The polynomial regression model proves advantageous. The results show minimum errors of about 10%, although rising to more than 40% for certain parameters. The extraction of a suitable subset as a basis for the regression analysis can improve the result to some degree, although it is important to maintain a minimum cluster size to achieve acceptable results.

Applied to a specific helicopter the significant improvements can be demonstrated for the majority of parameters.

Statistical methods are well suited for the early stages of helicopter design if a sufficiently large database is available to base the calculations on. Especially for unconventional configurations this poses a problem as only few of those helicopter models exist.

Concerning DLR research activities the method will be used to obtain initial values for the helicopter geometry and mass data, whereas simple physics-based methods will be favoured for performance and power calculations.

REFERENCES

[1]   Böhnke, D.; Nagel, B.; Gollnick, V.: *An approach to multi-fidelity in conceptual aircraft design in distributed design environments*. 2011 IEEE Aerospace Conference, Big Sky, USA, 2011.

[2]   Rand, O. ; Khromov, V.: *Helicopter Sizing by Statistics*. 58th Annual Forum of the American Helicopter Society, Montreal, Canada, 2002.

[3]   Kim, J.-M.; Oh, W.-S.: *A Study of Rotorcraft Initial Design Using Statistics*. Rotor Korea, AHS Specialists' International Conference, Seoul, South Korea, 2007.

[4]   Davis, S. J.; Rosenstein, H.; Stanzione, K. A.; Wisniewski, J. S.: *User's Manual for HESCOMP, The Helicopter Sizing and Performance Computer Program*. Naval Air Development Office, Warminster, 1979. (NADC-78265-60)

[5]   Johnson, W.; Sinsay, J. D.: *Rotorcraft Conceptual Design Environment*. 2nd International Forum on Rotorcraft Multidisciplinary Technology, Seoul, South Korea, 2009.

[6]   Jackson, P. (ed.): *Jane's All The World's Aircraft 2009-2010*. IHS Jane's, Coulsdon, 2009. (and older editions)

[7]   Oliver, D. (ed.): *Jane's Helicopter Markets and Systems*, Issue 29. IHS Jane's, Coulsdon, 2009.

[8]   Kendall, M: *A New Measure of Rank Correlation*. Biometrika 30 (1938), no 1-2, pp. 81-89.

[9]   Joliffe, I. T.: *Principal Component Analysis*. 2nd ed., Springer, New York, 2002.

[10]   Gosset, W. S. (Student): *The probable error of a mean*. Biometrika 6 (1908), no 1, pp. 1-25.

[11]   Lloyd, S. P.: *Least squares quantization in PCM*. IEEE Transactions on Information Theory 28 (1982), no 2, pp. 129–137.