19th AIAA/CEAS Aeroacoustics Conference
May 27-29, 2013, Berlin, Germany AIAA 2013-2012

Airfoil noise analysis using symbolic regression

Ennes Sarradj*, Thomas Geyer†
Chair of Technical Acoustics, Brandenburg University of Technology Cottbus, 03046 Cottbus, Germany

Due to the physical complexity of the problem and the large number of parameters that have an influence empirical or semi-empirical models for airfoil noise generation are used in many applications. These models are constructed from sets of observed data. In contrast to classical regression, where the data is used to find the coefficients of the regression function, the method of symbolic regression does symbolically identify the function itself. The present study applies this method to establish models for the noise generation at an airfoil. The process is exemplified for the noise generation at porous airfoils, where data sets from previous experimental studies are available. For both the self noise from the interaction of a turbulent boundary layer with the trailing edge of an porous airfoil and the noise generated at the leading edge due to turbulent inflow models of different accuracy and complexity are proposed. It is shown that symbolic regression can be used as a tool to provide insight into the noise generation process and the influencing parameters.

I. Introduction

The noise generated at an airfoil in flow is of great importance in many technical applications. In order to identify possible noise problems and to find mitigation measures, methods for the assessment of this noise are needed. Some efforts have lead to successful analytical and numerical prediction of airfoil noise. However, the physical complexity of the problem, the large number of parameters that have an influence in practical cases and the computational cost that is connected with numerical computations have lead to a certain popularity of empirical or semi-empirical models.

These models are based on data sets of observations of the noise generated and a tuple of the respective controlling parameters. The data is either generated through experimental tests or by a series of numerical computations. Then, a plausible regression model that is based on analytical models is used to find the coefficients of the regression function. This function then produces a relationship between controlling parameters and the sound generation that gives a minimal error when compared to the observed data set. Examples for such models are discussed and used in the literature.[1,2] While these models apply to a wide range of conventional technical airfoils they can not be used for airfoils with modifications such a serrations and porosity. In such cases the prediction of sound generation either relies on analytical and numerical methods or new empirical models need to be found.

Semi-empirical models require to assume a model for the regression function, which may be not straightforward to find. This is especially true when there are parameters involved that are not considered in analytical models and when the underlying physics are not fully understood. In such cases, an analysis may help that does not need to assume a model for the regression function, but instead finds this from the data. A special form of such approach is the symbolic regression, where the goal is not to find coefficients for a certain model, but to create a functional relation which produces the observed data when given the respective parameter data sets as input.

The present study applies the method of symbolic regression to establish models for the noise generation at an airfoil. The process is exemplified for the noise generation at porous airfoils. The paper is organized as follows: First, basic analytical models for conventional airfoils are shortly reviewed and the data on porous airfoils is summarized. Second, genetic programming as a tool for symbolic regression is briefly explained and the setup and techniques for the noise measurements as well as the choice of non-dimensional quantities

*Professor, Chair of Technical Acoustics, Brandenburg University of Technology Cottbus
†Research Assistant, Chair of Technical Acoustics, Brandenburg University of Technology Cottbus

for the characterization are discussed. After the data from two different experimental surveys is presented, the results for models of two different noise generation mechanisms are presented and discussed.

II. Airfoil noise models

A number of different mechanisms may result in sound generation at airfoils in flow. In most cases the interaction of the flow with the surface of the airfoil leads to fluctuating forces acting on the airfoil. These forces then cause sound radiation from the airfoil. In general, two possible reasons for the fluctuating forces were identified. First, effects in the flow boundary layer can produce such forces that in turn generate airfoil 'self noise'. In this case, the noise is controlled by the properties of the fluid, the flow velocity and a possibly complex set of airfoil properties. Secondly, turbulent flow disturbances that are convected in the oncoming flow may generate unsteady load on the airfoil. This is termed 'turbulence interaction' noise, and, in addition to the fluid and airfoil properties, is controlled by the characteristics of the turbulent inflow.

Here, only two special cases shall be considered: the self noise from the interaction of a turbulent boundary layer with the trailing edge of the airfoil and the noise generated at the leading edge due to turbulent inflow. A classical analytical model for the first mechanism was derived by Ffowcs Williams and Hall.[3] It considers the noise from a flow past the trailing edge of a semi-infinite plate. One conclusion from the model is the dependency of the farfield intensity (and thus the rms sound pressure $\tilde{p}$) on certain parameters. For a fixed observation point in the farfield,

$$\tilde{p}^2 \sim \tilde{u}^2 U M^2 L^2 \tag{1}$$

with the root mean square turbulent velocity fluctuation $\tilde{u}$, the flow velocity U, the Mach number $M = U/c_0$, the speed of sound c_0 and a characteristic dimension L of the eddies in the turbulent boundary layer. As both $\tilde{u}$ and M are proportional to U, an overall dependence of the farfield intensity on U^5 can be concluded. This important result is the basis for a number of refined models on trailing noise. It is also a basis for the semi-empirical model by Brooks et al.[1] that was derived from data measured on different NACA0012 airfoils and uses the boundary layer displacement thickness δ^* as a characteristic dimension. Other, more simplified airfoil noise models that include the noise generated at the trailing edge based on the work by Ffowcs Williams and Hall were for example presented by Schlinker and Amiet,[4] by Grosveld[5] and Lowson.[6] All of those models contain the dependency of the far field sound pressure level on the fifth power of the flow velocity.

For turbulent inflow noise, in his analytical model for the interaction of turbulence with the leading edge of a flat plate, Sharland[7] finds an U^6 dependency for the total sound power. Later, Amiet[8] presented an analytical model for the interaction of isotropic turbulence with the leading edge of a thin airfoil, which was verified using experimental results by Fink.[9] This analysis leads to a dependency on U^5 and $(\tilde{u}/U)^2$ for the far field intensity for a fixed Strouhal number $St = f\Lambda/U$, where f is the frequency and Λ is a characteristic dimension of the turbulent eddies in the inflow. Later, this theory was extended[10] to include the thickness of the airfoil edge. Several experimental studies confirmed the general validity of these models and find that a velocity dependency of $U^{5\dots6}$ fits the results best.[11–18]

However, in both the trailing and leading edge cases analytical models and derived empirical formulations account only for standard airfoils or plates. They do not take into account any modifications that are made to the material of the airfoil or the shape of the edges that aim at reducing the noise generation. The case of porous airfoils may serve as an example. Such airfoils – or airfoils with porous patches – may have a considerably reduced noise generation. A preliminary attempt[19] to establish a semi-empirical model for the trailing edge noise based on turbulent velocity spectra from the boundary layer had only limited success. There is no model at all for the leading edge noise generation at porous airfoils. On the other hand side, data from a large number of measurements on porous airfoils is available to build models directly from the data and test them on it.

III. Methods and Data

A. Symbolic regression using genetic programming

Symbolic regression handles the problem to identify a function symbolically that fits a given set of data in some way. The function then gives a relationship between a dependent variable (e.g. sound pressure) and one or more independent variables (e.g. flow velocity, frequency). The data set used to identify the function consists of values for the dependent variable and assigned tuples with values of the independent variables. Different strategies exist for solving the problem of symbolic regression. One possible approach, that is quite versatile, is genetic programming. This means that the function to find is represented as a computer program. The problem is thus converted into the problem of finding a program instead of a function in symbolic form. The program is part of a set (the population) of such programs that is bred using algorithms that resemble the course of evolution including the principle of survival of the fittest and recombination of individual programs from the population.

In order to be able to apply genetic programming for symbolic regression, the model or functional relation sought must be represented by a composition of primitive functions and terminals. Such model may be quite effectively represented as a rooted tree the branches of which being the functions and the leaves being the terminals. The tree may be interpreted as a program to compute the function. Another important feature is that it can be manipulated by adding, changing or deleting subtrees. Thus, the function can be altered using such manipulations.

The course of genetic programming for symbolic regression can be summarized by the following steps. First, create a population of programs (trees) from the primitive functions and the terminals using some random process. Second, estimate the fitness of each program by executing it. The fitness is simply some defined measure that estimates how good the function coded by the program fits the sought relation. Third, select one or two individual programs from the population. The higher the fitness of an individual, the more probable its selection should be. The fourth step is to create new programs by applying genetic operations on the selected individuals (the parents). The possible genetic operations include: simple reproduction (copy), crossover (recombination of randomly chosen parts of the parents), mutation (randomly altering a part of a parent). The third and fourth step are repeated until a new population (the next generation) is built. Now, if no individual in the population has a required minimal fitness nor any other stopping criterion is fulfilled, everything is repeated beginning with the second step and using the new generation as the population. The effect of this approach is that the fitness of the population is ever increasing.

The application of genetic programming usually requires some preparatory steps. These steps are[20] (i) selecting the set of terminals as well as (ii) the set of primitive functions that should be used, (iii) defining a fitness measure, (iv) setting the parameters that control the run and finally (v) to decide what should be the result and possibly what termination criterion is to be used.

For symbolic regression the terminal variables are the independent variables. In data-generating experiments on airfoil noise these variables originally represent quantities that have dimensions. A problem with this is that in order to get a physically meaningful model, only some functions may be allowed on certain quantities. For instance, it is not allowed to add a length to a frequency. Thus, it is a good idea to use the independent quantities not directly, but to derive dimensionless quantities from the input data and to use them as terminals instead. Consequently, also the dependent quantity needs to be dimensionless. Additionally, the use of constants as terminals is possible. As no numerical operations are performed to estimate the value of these constants in genetic programming, use is made of ephemeral random constants that are initially introduced by random choice and are not changed afterwards. In the course of genetic programming, these constants can be combined using the primitive functions available, thus making the selection of those constants possible that lead to higher fitness.

The set of primitive functions governs the possible types of functional relationships that may be derived by symbolic regression. For a certain physical process it is often clear that only a limited number of functional building blocks can be expected in the relation. In the case of airfoil noise it seems sensible to include basic operations such as addition, subtraction, multiplication, division and possibly power. In contrast, it seems not sensible to have functions such as sine and cosine in the relation unless the independent variables do also encompass any angles or quantities that may be combined into angles.

The fitness measure must consider how good the individual function reproduces the data used for the regression. This can be achieved by observing the error the function f makes in predicting the observed values of the dependent variable y for the set of independent variables $\mathbf{x}$. A number of possible choices exist for the error metric. The error metric that was used in the present analysis is the mean squared error:

$$\overline{e^2} = \frac{1}{N} \sum_{i=1}^{N} \left(y(\mathbf{x}_i) - f(\mathbf{x}_i) \right)^2 . \tag{2}$$

All errors are weighted equally in the sum, regardless of the value of $y(\mathbf{x}_i)$. If y is the root mean square sound pressure, both larger and smaller values can be expected. Then, it is preferable to consider the error in normalized form. This can be done by replacing y and f by their logarithms or by applying a logarithmic model for the fit.

The parameters to choose for the run of genetic programming are the size of the population, the method of choosing the parent individuals and the probabilities of the genetic operations that govern the breeding of a new generation. Two strategies were used in the present research. First, a fully configurable simple evolutionary algorithm[21] was implemented using the open source DEAP framework[22] implemented in Python. Following the literature,[20,23] the parameters that were used are: population size 500, size 3 tournament selection, mating probability 0.9, mutation probability 0.01. The second strategy was the use of algorithms and the parameter sets internal to the Formulize software.[24] Due to faster performance, the latter was chosen for all results reported here.

Finally, the result are the best individual functions from all generations together. Besides accuracy, parsimony is also important to select an individual function. A good heuristic is to prefer simple and clear structured functions as a model. Functions of a more complicated structure, however, have a greater potential to fit the data. Thus, a good solution to the symbolic regression problem is both parsimonious and accurate. Both measures have to be considered in the selection process. Thus, the best individual functions are those on a Pareto front in accuracy-parsimony-space. The prediction accuracy of a function is measured by the error also used for the fitness, but on another subset of the data as was used for the estimation of fitness during the genetic programming run. A complexity measure is used for parsimony. It is based on the number of terms in the function. Specifically, each primitive function branch and each terminal is given a specific weight and the total weight

$$w_t = \sum_i w_i \tag{3}$$

of the function is small for a parsimonious function and large for a function that has a complex structure. The sum is taken over all branches and terminals.

B. Experimental setup

Two experiments on airfoil noise were performed in the aeroacoustic wind tunnel at Cottbus university.[25] The first experiment concerned turbulent boundary layer - trailing edge noise, while the second experiment focuses on the noise generation at the leading edge due to turbulent inflow. Both experiments were described elsewhere[19,26,27] in detail. Thus, only a brief summary will be given here.

Both experiments used a very similar setup (Figure 1). An airfoil is placed in the measurement section of the tunnel in front of the nozzle. The nozzle and thus the jet have a diameter of 0.2 m. The airfoils used in the experiments have a span that is approximately 0.4 m and are thus wider than the jet. For the leading edge noise measurements, different grids were attached to the nozzle outlet to generate a turbulent flow. Hot-wire measurements were used for the estimation of the turbulent flow characteristics. A microphone array was applied to separately measure the individual contribution of the leading and the trailing edge to the noise.

The same set of porous airfoils was applied in both experiments. All airfoils had the same, slightly modified semi-symmetrical SD7003 shape[28] with a chord length of 0.235 m, but were made of different materials. These materials are characterized by the flow resistivity[29,30] r and the porosity σ, see Table 1.

All acoustic measurements were done using a planar microphone array with 56 1/4 inch microphone capsules flush mounted into an aluminum plate. The array has an aperture of 1.3 m and was mounted out of the flow in the measurement section at a distance of either 0.68 m (for the trailing edge noise measurement) or 0.72 m (for the leading edge noise measurements) above the airfoils. The duration of the individual measurements was 40 s and the microphone signals were sampled at 51.2 kHz. The data was partitioned into 999 blocks of 4096 samples each with 50% overlap, transformed using FFT and a von Hann window and assembled into a cross spectral matrix. This was then used as a basis for processing using beamforming and deconvolution algorithms. Based on a detailed comparison[19] of different deconvolution algorithms, orthogonal beamforming[31] was chosen in case of trailing edge measurements while CLEAN-SC[32] was chosen for the processing of the da ta from the leading edge noise measurements. In both cases, a three-dimensional source map was computed and the contributions from the leading and trailing edges were integrated over appropriate volumes located well within the core jet. Thus, the noise from the wind tunnel shear layers, and in case of the leading edge noise experiment also the noise from the turbulence generating grid, was filtered

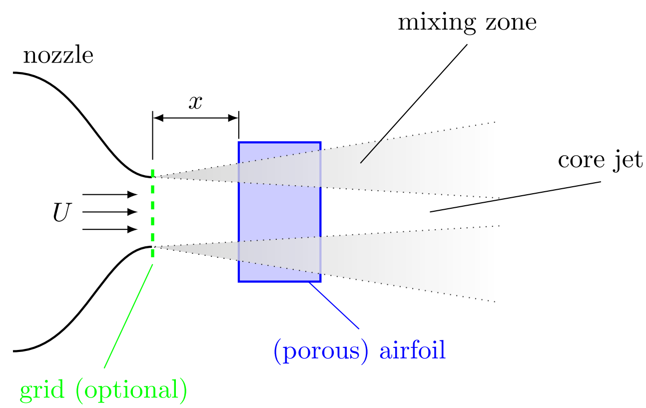


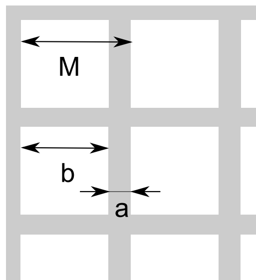
Figure 1. Schematic display of the measurement setup (top view)

Table 1. Materials used for the manufacturing of the airfoils (given are the air flow resistivity r and the open volume porosity σ_v)

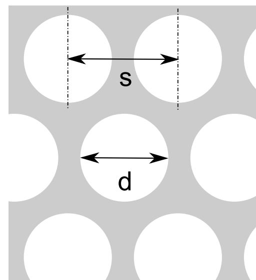
No	Name	Material	r [Pa s/m^2]	σ_v
	Reference	non–porous	∞	0
	M&K felt, 0.36 g/cm^3	woolen felt	506,400	0.73
	M&K felt, 0.22 g/cm^3	woolen felt	164,800	0.82
	Needlona felt, SO 2002	synthetic felt	130,200	$\approx$ 0.86
	ArmaFoam Sound	elastomer foam	112,100	0.85 - 0.9
	Needlona felt, WO–PE 1958	woolen / synthetic felt	40,100	$\approx$ 0.89
	Basotect	melamine resin–foam	9,800	>0.99
	Recemat	metal foam	8,200	>0.95
	Balzer RG 3550	polyurethane foam	4,400	>0.99
	Panacell 90 ppi	polyurethane foam	4,000	>0.99
	Panacell 60 ppi	polyurethane foam	3,600	>0.99
	M–Pore PU 45 ppi	polyurethane foam	1,500	0.86
	M–Pore Al 45 ppi	metal foam	1,000	0.90
	Panacell 45 ppi	polyurethane foam	700	>0.99

out. The span of the leading edge that contributed to the results was 0.1 m, while for the trailing edge it was 0.12 m. Finally, the data was converted into third octave band sound pressure levels.

A total of seven different grids were used to generate turbulence at the outlet of the wind tunnel nozzle. Six of these grids have a square mesh design and one of the grids is a perforated plate with round holes. The different dimensions of the grids lead to different properties of the turbulence that is generated. Table 2 gives an overview of the grids used for the experiments and the according grid parameters (see Figure 2(a) and 2(b)). The parameter t denotes the thickness of the grids. According to Roach,[33] turbulence grids with a porosity greater than 0.5 should be used in order to avoid flow instabilities. With the exception of the perforated plate with round holes (porosity 0.42), all grids fulfill this criterion.



(a) square mesh



(b) round holes

Figure 2. Definition of grid parameters for different designs

Table 2. Turbulence grids used in the experiments, grid parameters according to Figure 2(a) for perforated plates, square holes (*PPS*) and square mesh grids, round bars (*SMR*) and Figure 2(a) for perforated plates, round holes (*PPR*)

Grid	description	M (d) [mm]	a (s) [mm]	t [mm]
PPS 12/2	perforated plate, square holes	mesh width 12	bar width 2	1
PPS 14/4	perforated plate, square holes	mesh width 14	bar width 4	1
SMR 5/1	square mesh grid, round bars	mesh width 5	rod diameter 1	2
SMR 16/1.2	square mesh grid, round bars	mesh width 16	rod diameter 1.2	2.4
SMR 12.5/0.8	square mesh grid, round bars	mesh width 12.5	rod diameter 0.8	1.6
SMR 10.6/0.9	square mesh grid, round bars	mesh width 12.5	rod diameter 0.8	1.8
PPR 8/11	perforated plate, round holes	hole diameter 8	hole distance 11	1

The turbulent flow used for the leading edge noise experiments was characterized using measurements with a single-wire Dantec 55P14 hot-wire probe in a constant temperature anemometry setting. Using the probe, measurements were performed at four different flow velocities between 12 and 45 m/s and in three different distances to the grid (0.1 m, 0.2 m, 0.3 m) for each of the seven grids. For each distance, the velocity was measured at 15 different positions in a plane parallel to the grid.[27] The signal from the hot-wire was sampled at 25.6 kHz and recorded for 10 s at each measurement position. Then, an FFT with a von Hann window was applied on blocks of 4096 samples each with an overlap of 50%. The resulting autospectra of the velocity fluctuations were averaged over the 187 blocks per location and over the 15 locations, thus giving a good assessment of the actual autospectrum.

Isotropic turbulence can be characterized using the root mean square turbulent velocity fluctuation $\tilde{u}$ and the streamwise integral length scale Λ_x. In this case, the one-sided autospectrum of u can be fitted to[34]

$$G_{uu}(f) = \frac{4\tilde{u}^2\Lambda_x}{U\left(1+\left(\frac{2\pi f\Lambda_x}{U}\right)^2\right)} = \frac{4\tilde{u}^2 t_0}{1+(2\pi f t_0)^2}. \quad (4)$$

Using Taylors frozen turbulence hypothesis, the second form replaces the length scale by a time scale t_0 for the turbulence. Unlike in previous studies,[27] where the autocorrelation function was used to assess t_0, the present study uses (4) to assess both $\tilde{u}$ and t_0 from a least-squares fit of the measured autospectra for each combination of grid, flow velocity and distance. From this data, interpolating functions were constructed using a two-dimensional Clough-Tocher scheme[35] that give $\tilde{u}$ and t_0 as functions of the flow velocity and the distance to the grid. Example results are shown in Figure 3.

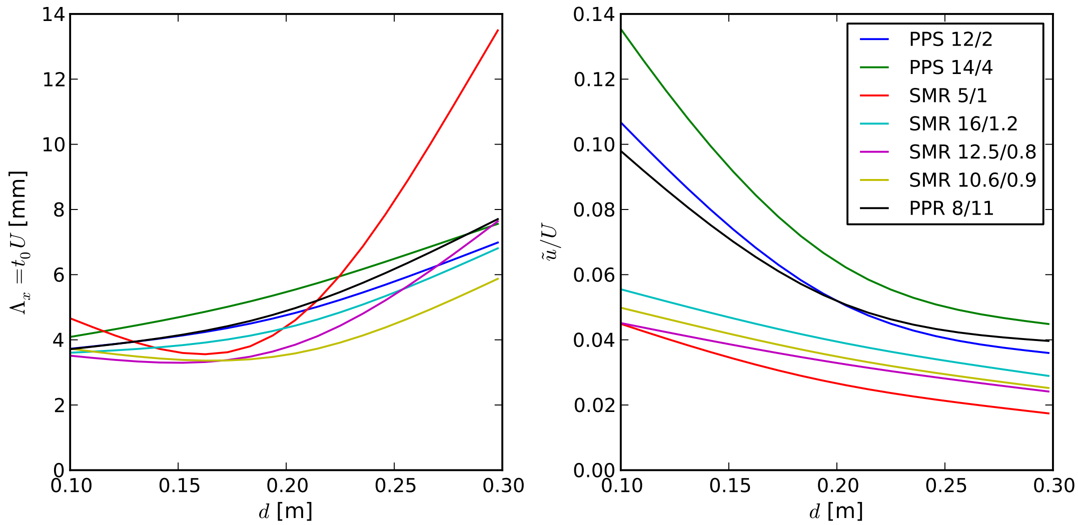


Figure 3. Characteristic length scale and turbulence intensity for seven grids, $U \approx 40\ m/s$

C. Dimensionless quantities

It is of advantage if the functional relation to find does not need to consider units. Thus, it is necessary to find dimensionless numbers that describe the problem at hand. Using the tool of dimensional analysis, the first step is to find all parameters that govern the problem. The second step is then to identify the number of dimensionless quantities using the Buckingham π theorem.[36]

The sound pressure $\tilde{p}$ of the noise generated by the turbulent boundary layer at the trailing edge of a porous airfoil is governed by the following parameters: the speed of sound c, the density ρ and kinematic viscosity ν of the fluid, the chord s, the span b, the angle of attack α and the shape of the airfoil, the flow velocity U, some boundary layer scale parameter or characteristic dimension of the boundary layer turbulence δ, the observer distance z (when observed under a constant angle), the frequency f, and for the porous material the flow resistivity r, porosity σ and tortuosity τ. From this 14 quantities, dimensional analysis leads to 11 dimensionless quantities, if the airfoil shape is not regarded. These were chosen to be:

- the normalized mean square sound pressure $\tilde{p}^2/(\rho c^2)^2$,
- the chord-based Reynolds number $Re = Us/\nu$,
- the chord-based Strouhal number $Sr = fs/U$,
- the Mach number $Ma = U/c$,
- the acoustical Rayleigh number[37] $Ra_a = f\rho/r$,
- the porosity σ,
- the tortuosity τ,
- the length ratios δ/s, z/s, b/s and

- the angle of attack α.

In the experiments, the fluid, the airfoil shape, size and angle of attack and the observer position were not altered. For the porous materials, the tortuosity was considered to be constant (≈ 1). Thus, only the first six quantities need to be considered and the sought function g_{TE} depends on five quantities:

$$\frac{\tilde{p}^2}{(\rho c^2)^2} = g_{TE}(Re, Sr, Ma, Ra_a, \sigma). \tag{5}$$

In the leading edge noise case, a very similar analysis can be performed. Instead of δ, parameters for the incoming turbulence must be included: t_0 and $\tilde{u}$. This leads to 12 dimensionless quantities. This time,

- the normalized mean square sound pressure $\tilde{p}^2/(\rho c^2)^2$,
- the chord-based Reynolds number $Re = Us/\nu$,
- the chord-based Helmholtz number $He = \dfrac{fs}{c}$,
- the turbulence-based Strouhal number $Sr = ft_0$,
- turbulence intensity $Tu = \tilde{u}/U$,
- the Mach number $Ma = U/c$,
- a length ratio between characteristic lengths for the porous structure and the turbulence $L = \sqrt{\nu\rho/r}/(Ut_0)$,
- the porosity σ,
- the tortuosity τ,
- the length ratios z/s, b/s and,
- the angle of attack α

were used. Again, the same parameters as for the trailing edge case remained constant. Consequently, the sought function g_{LE} depends on seven quantities:

$$\frac{\tilde{p}^2}{(\rho c^2)^2} = g_{LE}(Re, He, Sr, Tu, Ma, L, \sigma). \tag{6}$$

D. Data sets

The measurements for the trailing edge noise experiment were performed for a large number of combinations of different airfoils, flow velocities and angles of attack.[26] A subset from the available data was selected for 14 different airfoils and 0°angle of attack. For each airfoil measurements were taken at 15 to 17 different flow velocities between 25 m/s and 50 m/s. In order to include only the case of a non-compact aeroacoustic source at the airfoil, frequencies below 2 kHz were excluded from the analysis and only third octave frequency bands between 2 kHz and 20 kHz were included. Thus, the data set used for the regression consists of 2420 tuples of numbers. Each tuple includes all quantities that are used in Equation (5).

In order to do a successful regression, the independent variables need to span an adequate range, and the individual values of a certain variable should be distributed over that range. This is the case for the Reynolds number ($4{\cdot}10^5 \ldots\ 7.8{\cdot}10^5$), the Mach number (0.07... 0.14), the Strouhal number (9.4... 180) and the acoustical Rayleigh number (0... 32.7). However, for the porosity the range is very limited (0.73... 1, see Table 1) and many materials have a porosity close to 1. Thus, the porosity was not considered in the data set for the regression.

Figure 4 shows the data set for the trailing edge noise. The sound pressure level was scaled according to the U^5 hypothesis of Ffowcs Williams and Hall.[3] The spectra are plotted using the Strouhal number as non-dimensional frequency. Different flow resistivities are shown color-coded. The black dots represent a standard, non-porous airfoil. The figure allows the conclusion that certain flow resistivities result in noticeable sound reductions compared to others, but no definite relation can be read out of it.

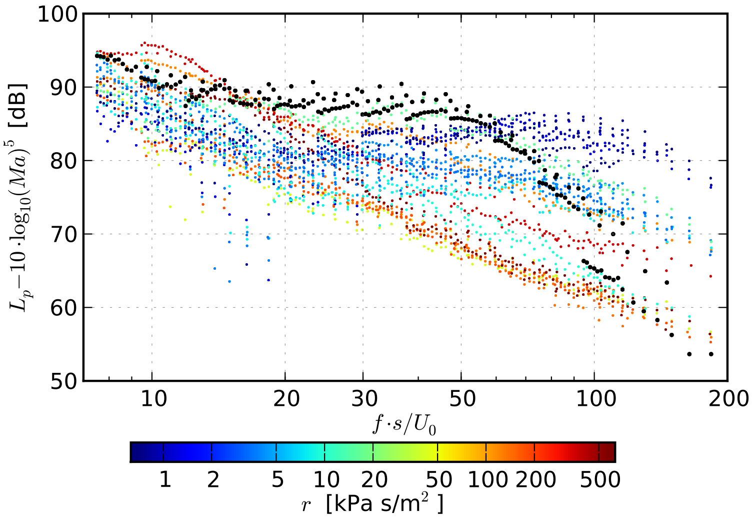


Figure 4. Scaled sound pressure levels for the trailing edge noise as a function of the chord-based Strouhal number and the flow resistivity (color-coded)

In case of the leading edge noise experiment, the inflow turbulence was varied by using different grids and by placing the airfoils in different distances from the grid. Seven grids and five different distances were used in the experiment. If all possible combinations would have been tried, the total number of measurement cases would have increased 35-fold compared to the trailing edge noise measurement. Thus, instead of a full-factorial design of the experiment, a D-optimal design[38] with some extra measurement cases was used. Altogether, this resulted in manageable 386 different parameter combinations or, equivalently, individual acoustic measurements. The D-optimal design assures an adequate distribution of each parameter over the realizable range of values.

Noise from turbulence interaction at the leading edge is known to have less energy at high frequencies.[39] Thus, only frequency bands up to 10 kHz were included in this case. The final data set for the leading edge noise experiments consists of 4139 tuples of the numbers used in Equation (6). Figure 5 shows the data set for the leading edge noise. The sound pressure level was scaled according to an assumed U^6 dependency and plotted over the Strouhal number. Again, it can be concluded that in some cases the noise generation is lower than for the non-porous reference airfoil. However, no obvious relation appears that can be used to predict the influence of the flow resistivity or other parameters on the noise generation.

IV. Results and Discussion

For the trailing edge noise, results from two different genetic programming runs are reported here. Both runs delivered a number of functions that establish a relation between the independent and the dependent quantities. Because the dependent quantity $\tilde{p}^2/(\rho c^2)^2$ span several orders of magnitude, instead of directly applying the model according to Equation (5) as a basis, its logarithm was used and a function g_{TE} was sought that fits

$$10 \lg \left(\frac{\tilde{p}^2}{(\rho c^2)^2} \right) = 10 \lg \left(g_{TE}(Re, Sr, Ma, Ra_a) \right). \tag{7}$$

This way, the deviation between model function and measured data is expressed in decibel.

The set of primitive functions for the first run was $\{+, -, \cdot, /, \hat{}\,\}$, the basic algebraic operations and the power function. The terminals that were allowed are $\{Re, Sr, Ma, Ra_a, C\}$, were C stands for an ephemeral constant. The mean squared error metric according to Equation (2) was used and the run was performed using the algorithms and parameters internal to the Formulize software. The run time spent for the first run was 50 hours on a personal computer. From all approximately 10^6 generations, a small set of candidate solutions were produced on an error-complexity-Pareto front (Figure 6). The weights used for the complexity

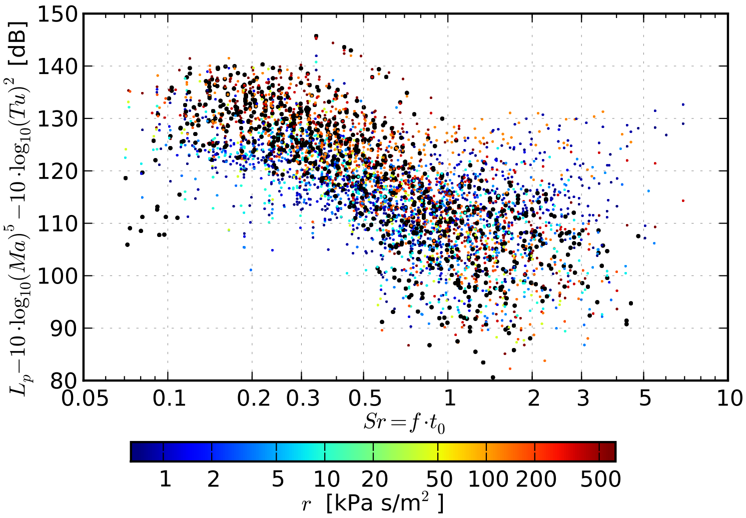


Figure 5. Scaled sound pressure levels for the leading edge noise as a function of the turbulence-based Strouhal number and the flow resistivity (color-coded)

measure were equal except for the division operator, were a double weight was applied. This means a slight extra penalty for the use of division.

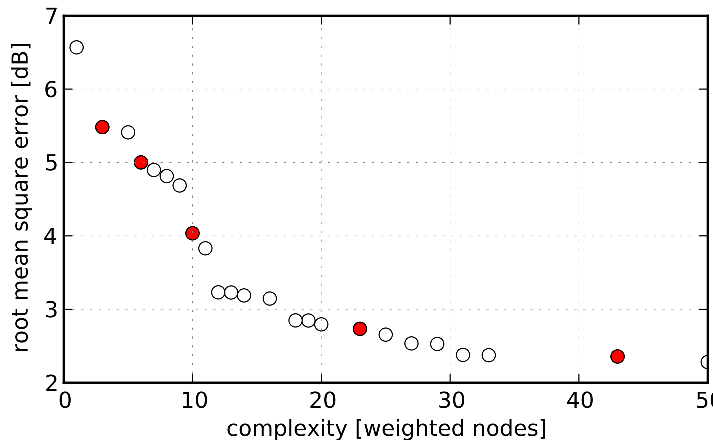


Figure 6. Error and complexity Pareto front for the first run (red markers denote solutions according to Equations (8) **through** (12)**)**

The five candidate solutions marked in Figure 6 and ordered from parsimonious to complex are:

$$\tilde{p}^2 = C_T \cdot Ma^{5.88}, \tag{8}$$

$$\tilde{p}^2 = C_T \cdot \frac{Ma^{4.32}}{Sr}, \tag{9}$$

$$\tilde{p}^2 = C_T \cdot \left(0.00062 Ra_a + \frac{Ma}{Sr}\right)^{2.49}, \tag{10}$$

$$\tilde{p}^2 = C_T \cdot Ma^{4.81} Sr \frac{17.8 + (0.00041 + Ra_a)^{-0.75}}{Sr + 1.05 Ra_a - 0.2}, \tag{11}$$

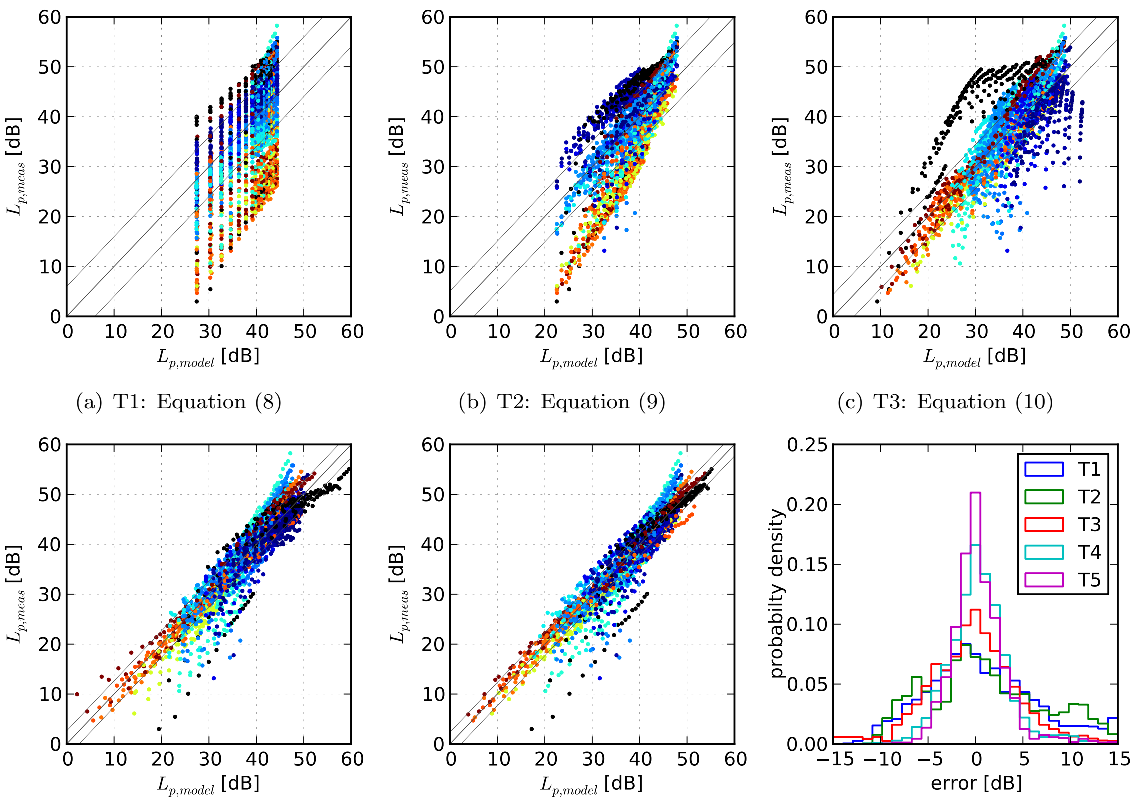
$$\tilde{p}^2 = C_T \cdot Ma^{5.02} \left((0.9 + 0.004 Sr + Ra_a Sr)^{-9.34} + 221 \left(\frac{0.75 + Ra_a}{Sr + 12 Ra_a} \right)^{2.41} \right). \tag{12}$$

As the data set referred to a constant span $b = 0.12$ m and observer distance $r = 0.68$ m, the candidate solutions were multiplied by a factor

$$C_T = \frac{17b}{3r}\,\mathrm{Pa}^2. \tag{13}$$

In the present analysis $C = 1$ Pa^2. However, this factor considers theoretical dependencies $\tilde{p}^2 \propto b$ and $\tilde{p}^2 \propto 1/r$.

The first candidate solution (8) depends on Ma only. It is far to simple to give a small error, but the exponent for Ma gives a dependency not very different from the theoretical U^5. However, Figure 7(a) shows that this candidate solution is virtually useless to predict the trailing edge noise generation. The second candidate solution (9) includes an additional dependency on Sr. Again the overall $U^{5.32}$ dependency is close to the theoretical value. In addition, the general trend of lower sound pressure for higher Sr agrees with the data set. The candidate solution (10) is the most parsimonious solution that exhibits a dependency on Ra_a. However, the Ra_a-dependency in the data set is clearly not as simple as in this solution. While the error is smaller (see Figure 7(c)) than for the other solutions, the dependency on U is more complicated and a comparison with theory is not easily possible. This is different for the fourth candidate solution (11), where the term $Ma^{4.81}$ can be factored out from the rest of the solution. The solution gives a relatively low error, approximates the theoretical U^5-dependency and exhibits a less complex combined dependency on Sr and Ra_a. The fifth and final candidate solution (12) has a nearly perfect agreement with the theoretical U^5-dependency, but has a very complex term for the Sr- and Ra_a-dependencies.



(a) T1: Equation (8) (b) T2: Equation (9) (c) T3: Equation (10)

(d) T4: Equation (11) (e) T5: Equation (12) (f) error histogram

Figure 7. Models for trailing edge noise from Equation (8)...(12) (corresponding to T1...T5), color scale for flow resitivity see Figure 4, the lines mark the mean and the mean ± root mean square error

Despite the use of the model according to Equation (7), that includes also Re, only Sr, Ma, and Ra_a appear in the candidate solutions. One possible explanation is that only one unique chord length was used in the experiments and the speed of sound varied only slightly due to temperature variations in the order of a few Kelvin. Consequently, Re and Ma are basically proportional, because the only quantity that was varied was the flow velocity U.

Figure 7(f) compares the error distribution for each of the five solutions. The error is small enough to have a significant amount of the data tuples with errors less than 3 dB for the last two solutions. These solutions could be used to model the dependency of the sound generation on flow velocity, porous material flow resistivity and frequency. However, both solutions are not perfectly suited. Solution (11) has a larger error and solution (12) has an unlikely complicated dependency on Sr and Ra_a. This could possibly be cured by extending the run time by a large amount.

Another approach was taken in the second run. The first run showed that fractional exponents are not necessary to model the Ma- or U-dependency. Moreover, the form of the frequency spectrum cannot be expected to yield a simple power law that includes Ra_a or Sr. Thus, the power function was removed from the set of primitive functions for the second run to prevent any fractional powers to appear in the solution. To include possible dependencies on integer powers, the terminal set was extended to $\{Re, Re^2, Re^3, Re^4, Re^5, Re^6, Sr, Sr^2, Sr^3, Sr^4, Ma, Ma^2, Ma^3, Ma^4, Ma^5, Ma^6, Ra_a, Ra_a^2, Ra_a^3, Ra_a^4, C\}$. The same error metric and parameters were used as in the first run.

After only $3{\cdot}10^4$ generations (runtime approximately 2 hours), the algorithm proposed a model with the same low error as solution (12), but with a less complex dependency on Sr and Ra_a:

$$\tilde{p}^2 = C_T \cdot Ma^5 \frac{500 + 47600 Ra_a^2}{79 + Sr^2 + 81000 Ra_a^2 + 425 Sr^2 Ra_a}. \tag{14}$$

This convenient model can be used to gain some insight into the reduced sound generation at the trailing edge of porous airfoils. To this end, the model can be used to predict the sound pressure level difference

$$\Delta L = L_{\text{porous}} - L_{\text{reference}} = L(Ma, Sr, Ra_a) - L(Ma, Sr, 0). \tag{15}$$

In Figure 8, this difference is plotted for arbitrary Mach numbers as a function of Strouhal number and acoustical Rayleigh number. The range of values for the plot is the same as for the experiments. The plot shows that the maximum reduction appears for $Ra_a \approx 0.1$ and that the reduction is less for larger Ra_a and Sr. For $Ra_a > 4$ and $Sr > 30$, the porous airfoil may produce more sound at the trailing edge than the non-porous reference airfoil.

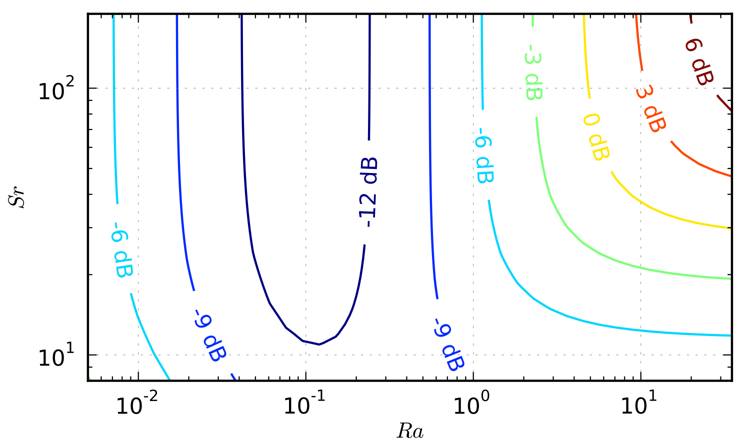


Figure 8. Trailing edge noise generation at a fixed Mach number compared to a non-porous reference airfoil ($Ra_a = 0$) as level difference using the model according to Equation (14)

The analysis run for the leading edge based on a data set with even wider span of the dependent quantity. Thus, with the same intentions as in the trailing edge case, the symbolic regression was performed using the logarithmic model:

$$10 \lg \left(\frac{\tilde{p}^2}{(\rho c^2)^2} \right) = 10 \lg \left(g_{LE}(Re, He, Sr, Tu, Ma, L) \right). \tag{16}$$

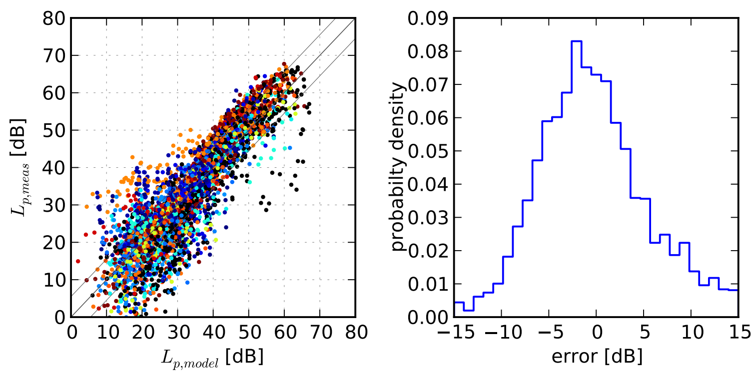
This time, the same set of primitive functions was used as in the second run for the trailing edge. The set of possible terminals included the first to the fourth power of He, Sr, Tu and L as well as the first to sixth power of Ma and Re and an ephemeral constant. The same error metric and parameters were used as in the trailing edge noise model runs.

Considering the number of independent quantities, the task of finding a relation that models leading edge noise is somewhat more difficult than it is for trailing edge noise. Consequently a much longer run time of 100 hours was used to allow the algorithm selecting results with small error and low complexity. However, it was not possible to produce results with same quality as in the trailing edge noise case. The relation

$$\tilde{p}^2 = C_L \cdot Ma^5\, Tu^2 \left(L + \frac{2.5Ma}{0.737 + (1 + 144SrL)\, He^2 - 1.66He} \right) \tag{17}$$

$$C_L = 7200 \frac{b}{r}\, \mathrm{Pa}^2 \tag{18}$$

is one of those with the smallest error found, but is also not to complex. Again, the result was multiplied by a factor that reflects the theoretical dependency on b and r. Here, this factor also includes an constant factor that is a result from the regression. The result has a root mean square error of 5.1 dB, much more than the best trailing edge models (Figure 9).



(a) Equation (17) (b) error histogram

Figure 9. Model for leading edge noise, color scale for flow resitivity see Figure 5, the lines mark the mean and the mean ± root mean square error

The result (17) depends on all independent quantities except the Reynolds number. Interestingly, it exhibits a dependency on Tu^2, which is also the case in theoretical models such as that of Amiet.[8] The Ma^5 factor hints at an U^5 dependency that can also be expected from theoretical models. However, the last term that depends on L, He and Sr has an additional factor Ma. Thus the overall dependency on U depends itself on the two summands in the parenthesis. If the second is much larger than the first, an overall U^6 dependency applies. If (17) is going to be applied to analyze possible sound reductions through the use of porous material, besides the ratio L, also He, Sr and Ma must be considered.

The decision which dimensionless quantities to use for symbolic regression is based on intuition and prior knowledge of theoretical models that leads to certain expectations on what quantities will appear in a model. Equation (17) shows that Sr was not the best possible choice because the turbulence based Strouhal number appears only in a product with L. This product

$$LSr = \frac{f\sqrt{\nu\rho/r}}{U} \tag{19}$$

is a Strouhal number itself, but based on a characteristic length for the porous structure and shows that the noise spectrum is less influenced by the turbulence time scale alone but by the ratio L and the characteristic length for the porous structure.

Both examples for the application of symbolic regression for the analysis of airfoil noise data have shown that it is indeed possible to establish meaningful models for the noise generation at an airfoil. These models can be used for prediction as long the range of values for the influencing parameters is in the same order as in the dataset used for the regression. They may also be useful to gain insight on the impact of individual parameters in a scenario where experiments do not allow the independent adjustment of these parameters.

V. Conclusion

The application of the method of symbolic regression to establish models for the noise generation at an airfoil was exemplified for the noise generation at porous airfoils. Both the self noise from the interaction of a turbulent boundary layer with the trailing edge of the airfoil and the noise generated at the leading edge due to turbulent inflow were considered. Based on two extensive sets of data from previous experimental studies, several models of different accuracy and complexity were found. These models were shown to provide insight into the physical process of noise generation and can also be used for the prediction of noise generation in certain situations. Consequently, the tool of symbolic regression can be used to complement other existing modeling strategies for airfoil noise.

References

[1]Brooks, T. F., Pope, D. S., and Marcolini, M. A., "Airfoil Self-Noise and Prediction," NASA Reference Publication 1218, 1989.

[2]Moriarty, P. and Migliore, P., "Semi-empirical aeroacoustic noise prediction code for wind turbines," Tech. rep., National Renewable Energy Laboratory, December 2003.

[3]Ffowcs Williams, J. E. and Hall, L. H., "Aerodynamic sound generation by turbulent flow in the vicinity of a scattering halfplane," *Journal Fluid Mechanics*, Vol. 40, No. 4, 1970, pp. 657 – 670.

[4]Schlinker, R. H. and Amiet, R. K., "Helicopter Rotor Trailing Edge Noise," NASA Contractor Report 3470, 1981.

[5]Grosveld, F. W., "Prediction of Broadband Noise from Horizontal Axis Wind Turbines," *Journal of Propulsion*, Vol. 1, 1985, pp. 292 – 299.

[6]Lowson, M. V., "Assessment and prediction of wind turbine noise," *Flow Solutions Report*, Vol. 92/19, 1992, pp. 1–59.

[7]Sharland, I. J., "Sources of Noise in Axial Flow Fans," *Journal of Sound and Vibration*, Vol. 3, 1964, pp. 302–322.

[8]Amiet, R. K., "Acoustic radiation from an airfoil in a turbulent stream," *Journal of Sound and Vibration*, Vol. 41, No. 4, February 1975, pp. 407–420.

[9]Fink, M. R., "Experimental Evaluation of Trailing Edge and Incidence Fluctuation Noise Theories," *13th AIAA Aerospace Sciences Meeting, AIAA-paper 75-206*, 1975.

[10]Gershfeld, J., "Leading edge noise from thick foils in turbulent flows," *Journal of the Acoustic Society of America*, Vol. 116(3), 2004, pp. 1416–1426.

[11]Olsen, W. A., "Noise Generated by Impingement of Turbulent Flow on Airfoils of Varied Chord, Cylinders, and Other Flow Obstructions," NASA Technical Memorandum X-73464, 1976.

[12]Paterson, R. W. and Amiet, R. K., "Acoustic Radiation and Surface Pressure Characteristics of an Airfoil due to Incident Turbulence," NASA Contractor Report CR-2733, 1976.

[13]Oerlemans, S. and Migliore, P., "Aeroacoustic Wind Tunnel Tests of Wind Turbine Airfoils," *10th AIAA/CEAS Aeroacoustics Conference, AIAA-Paper 2004-3042*, 2004.

[14]Moreau, S. and Roger, M., "Effect of Angle of Attack and Airfoil Shape on Turbulence-Interaction Noise," *11th AIAA/CEAS Aeroacoustics Conference Meeting and Exhibit, May 23-25, Monterey, USA*, 2004.

[15]Staubs, J. K., *Real Airfoil Effects on Leading Edge Noise*, Ph.D. thesis, Virginia Polytechnic Institute and State University, 2008.

[16]Glegg, S., Devenport, W., and Staubs, J., "Sound Radiation from Three Dimensional Airfoils in a Turbulent Flow," *46th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA-paper 2008-52*, 2008.

[17]Devenport, W. J., Staubs, J. K., and Glegg, S. A. L., "Sound Radiation from Real Airfoils in Turbulence," *Journal of Sound and Vibration*, Vol. 329, 2010, pp. 3470 – 3483.

[18]Hutcheson, F. V., Brooks, T. F., Burley, C. L., and Stead, D. J., "Measurement of the Noise Resulting from the Interaction of Turbulence with a Lifting Surface," *17th AIAA/CEAS Aeroacoustics Conference, AIAA-paper 2011-2907*, 2011.

[19]Geyer, T., *Trailing Edge Noise Generation of Porous Airfoils*, Ph.D. thesis, Brandenburgische Technische Universität (BTU), Cottbus, 2011.

[20]Poli, R., Langdon, W. W. B., and McPhee, N. F., *Field Guide to Genetic Programming*, Lulu Enterprises Uk Limited, 2008.

[21]Bäck, T., Fogel, D. B., and Michalewicz, Z. E., editors, *Evolutionary computation 2: advanced algorithms and operators (Vol. 2)*, Taylor & Francis, 2000.

[22]Fortin, F. A., De Rainville, F. M., Gardner, M. A., Parizeau, M., and Gagn, C., "Deap: Evolutionary algorithms made easy," *Journal of Machine Learning Research*, Vol. 13, 2012, pp. 2171 – 2175.

[23]Koza, J. R., *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT press, 1993.

[24]Schmidt, M. and Lipson, H., "Distilling free-form natural laws from experimental data," *Science*, Vol. 324(5923), 2009, pp. 81 – 85.

[25]Sarradj, E., Fritzsche, C., Geyer, T., and Giesler, J., "Acoustic and Aerodynamic Design and Characterization of a Small-Scale Aeroacoustic Wind Tunnel," *Applied Acoustics*, Vol. 70, 2009, pp. 1073 – 1080.

[26]Geyer, T., Sarradj, E., and Fritzsche, C., "Measurement of the noise generation at the trailing edge of porous airfoils," *Experiments in Fluids*, Vol. 48 (2), 2010, pp. 291 – 308.

[27]Geyer, T., Sarradj, E., and Giesler, J., "Application of a Beamforming Technique to the Measurement of Airfoil Leading Edge Noise," *Advances in Acoustics and Vibration*, Vol. Vol. 2012, 2012.

[28]Selig, M. S., Donovan, J., and Fraser, D., *Airfoils at Low Speeds*, SoarTech Aero Publications, 1989.

[29]Mechel, F., *Formulas of Acoustics, Second Edition*, Springer, 2008.

[30]"ISO 9053 Acoustics – Materials for acoustical applications – Determination of airflow resistance," 1993.

[31]Sarradj, E., "A Fast Signal Subspace Approach for the Determination of Absolute Levels from Phased Microphone Array Measurements," *Journal of Sound and Vibration*, Vol. 329, 2010, pp. 1553 – 1569.

[32]Sijtsma, P., "CLEAN based on Spatial Source Coherence," *13th AIAA/CEAS Aeroacoustics Conference, AIAA paper 2007-3436*, 2007.

[33]Roach, P. W., "The Generation of Nearly Isotropic Turbulence by Means of Grids," *Journal of Heat and Fluid Flow*, Vol. 8 (2), 1987, pp. 82 – 92.

[34]Hinze, J. O., *Turbulence, Second Edition*, McGraw-Hill, New York, 1975.

[35]Alfeld, P., "A trivariate clough-tocher scheme for tetrahedral data." *Computer Aided Geometric Design*, Vol. 1(2), 1984, pp. 169 – 181.

[36]Gibbings, J., *Dimensional analysis*, Springer, 2011.

[37]Cremer, L., Müller, H. A., and Schaultz, T. J., *Principles and applications of room acoustics*, Vol. 1, Applied Science London, 1982.

[38]NIST/SEMATECH, *e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/.

[39]Blake, W. K., *Mechanics of Flow-Induced Sound and Vibration, Volume II: Complex Flow-Structure Interactions*, Academic Press, Inc., 1986.