# Evolutionary Bands for the Expected Response in Nonlinear Regression *

Przemysław Prętki[1] and Andrzej Obuchowicz [2]

[1,2] University of Zielona Góra,
Institute of Control and Computation Engineering,
Zielona Góra, Poland,
email: {P.Pretki,A.Obuchowicz}@issi.uz.zgora.pl

**Abstract** In this paper evolutionary algorithms are applied to computation of confidence intervals for the expected response of nonlinear models. A simple phenotypic evolutionary algorithm was adapted to deal with nonlinear constraints and utilized to find the maximum and minimum value of a nonlinear model responses inside a confidence region. Moreover, the adequacy of the proposed approach is tested in a series of numerical simulations, and compared with the commonly applied linearization technique.

## 1 Introduction

Linear as well as nonlinear regression analysis finds a number of successful applications in many fields of modern science e.g. economics, physics, control theory and computer science [1, 2]. In this paper an application of evolutionary algorithms (EAs) to the model-based fault diagnostic systems [5, 8, 10] is presented. The main purpose of such systems is to detect any unwanted behaviors in industrial processes. The observed increasing complexity of modern systems has resulted in many of the so-called Soft Computing methods in this field. The procedure of building a model-based diagnostic system, usually consists of a few steps. First, the Experimental Design Theory can be applied to collect a data set which would contain sufficient information about the system of interest. Next, on the basis of input-output dependencies, the parametric model is constructed in such way to imitate the true system as precisely as possible. In this way the correctness of the industrial process is monitored on the basis of residual signals i.e. a difference between model and the true system.
Among many parametric models, the approach based on artificial neural networks (ANNs) has received considerable research attention in recent years. One of the fundamental advantages of neural networks is that they can approximate any smooth function with an arbitrary degree of accuracy. For a known structure of a neural network, training boils down to the estimation of its parameters. From this point of view neural models can be seen as a tool of nonlinear regression analysis. To increase reliability and sensitivity of

*The work was supported by Ministry of Education an Science in Poland under the grant 3T11A01530

neural models to faulty conditions, the approach in this paper is to compute adaptive thresholds for the residual signal. The paper is organized as follows: The second and the third sections are devoted to the introduction to general concepts of linear and nonlinear regression respectively. In the fourth section the problem of computation the bands for the expected response function is formulated. Moreover, the scheme of phenotypic evolutionary algorithm dedicated to solving the problem is given, along with the procedure that allows for maintenance of the population of individuals inside the feasible set. The fifth section includes results of simulation experiments. Finally, the last section concludes the paper.

## 2 Linear regression

Firstly, let us briefly introduce the main concepts of linear and nonlinear regression. Linear in parameters models can be generally represented in the following way

$$y_{m,k} = \boldsymbol{\theta}^T \boldsymbol{x}_k + \epsilon_k, \tag{1}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$ denotes a parameters vector, $\boldsymbol{x} \in \mathbb{R}^{n_u}$ is an input vector ($n_u = n_p$), and $\epsilon_k$ stands for an additive noise. As one can see, the above model consists of the deterministic part $\boldsymbol{\theta}^T \boldsymbol{x}_k$ and stochastic part $\epsilon_k$. A regression is usually posed as an optimization problem – we are attempting to find the unknown parameters vector $\hat{\boldsymbol{\theta}}$ in such a way as to minimize the distance between observable data and model responses:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{n_p}} S(\boldsymbol{\theta}) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{n_p}} ||\boldsymbol{Y} - \boldsymbol{\theta}^T \boldsymbol{X}||, \tag{2}$$

where $\boldsymbol{X} \in \mathbb{R}^{n_t \times n_u}$ is the matrix of regressor variables ($n_t$ denotes number of observation), $\boldsymbol{Y} \in R^{n_t}$ stands for the vector of random variables representing the given data. It is worth noting that the norm definition $||\cdot||$ in (2) depends on nature of the noise. Usually the unknown parameters $\boldsymbol{\theta}$ are determined using the maximum likelihood estimation technique which poses very attractive asymptotic properties:

- for the linear model and gaussian, uncorrelated disturbances, the estimates are determined explicitly by the well-known formula:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}, \tag{3}$$

- $1 - \alpha$ join confidence regions for $\hat{\boldsymbol{\theta}}$ is the ellipsoid:

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{X}^T \boldsymbol{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq s^2 \, n_p \, F_{(n_p, n_t - n_p; \alpha)}, \tag{4}$$

  where $F_{(n_p, n_t - n_p; \alpha)}$ denotes the upper $\alpha$ quantile for Fisher's distribution with $n_p$ and $n_t - n_p$ degrees of freedom, and $s^2 = S(\hat{\boldsymbol{\theta}})/(n_t - n_p)$ is variance estimate.
- $1 - \alpha$ confidence band for the response function is given by:

$$\hat{\boldsymbol{\theta}}^T \boldsymbol{x} \pm s \sqrt{\boldsymbol{x}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}} \sqrt{n_p \, F_{(n_p, n_t - n_p; \alpha)}} \tag{5}$$

As one can observe, linear models provides inference regions which are easy to calculate no matter how many parameters one include in the model.

## 3 Nonlinear regression

Sometimes, the complexity of dependencies between observed variables does not allow the use of linear models. In such situations the nonlinear regression models should be employed:

$$y_{m,k} = f(\boldsymbol{x}_k, \boldsymbol{\theta}^*) + \epsilon_k, \tag{6}$$

where $f(\cdot)$ stands for a known non-linear function, $y_k$ is a scalar endogenous variable, $\boldsymbol{x}_k \in \mathbb{R}^{n_u}$ denotes a vector of exogenous variables, $\boldsymbol{\theta}^* \in \mathbb{R}^{n_p}$ is a vector of true (unknown) parameters, and $\epsilon_k$ are unobservable scalar i.i.d. random variables with $\mathrm{E}[\epsilon_k] = 0$, and $\mathrm{Var}[\epsilon_k] = \sigma^2$, another unknown parameter. Due to the nonlinearity of $f(\cdot)$, generally, it is impossible to deliver explicit formula which guaranties optimality of (2) and many of the maximum likelihood estimates lose some of their desirable properties. Thus, in order to obtain the parameters estimate $\hat{\boldsymbol{\theta}}$, some numerical algorithms must be applied. The statistical inference about confidence regions and confidence intervals is even more complicated. whether there exists a vector $\boldsymbol{\theta}_0 \in \Theta$ such that $f(\boldsymbol{u}_k, \boldsymbol{\theta}_0) = z$.
One way out of such difficulties, very often applied in practice, is to make use of a linearized version of the model (6) in the vicinity of the current estimates $\hat{\boldsymbol{\theta}}$ [1, 2, 3]:

$$y_{m,k} = f(\boldsymbol{u}_k, \hat{\boldsymbol{\theta}}) + \boldsymbol{V}_k^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{7}$$

where $\boldsymbol{V}_k = \frac{\partial f(\boldsymbol{u}_k, \boldsymbol{\theta})}{\partial \theta}\big|_{\hat{\boldsymbol{\theta}}}$. Based on the linear approximation (7) it is possible to determine the lower $y_{m,k}^L$ and upper $y_{m,k}^U$ adaptive thresholds for the predicted response of the model (6), [1]:

$$y_{m,k}^L = f(\boldsymbol{u}_k, \hat{\boldsymbol{\theta}}) - s\sqrt{\boldsymbol{V}_k^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{V}_k}\sqrt{n_p\, F_{(n_p, n_t - n_p; \alpha)}}, \tag{8}$$

$$y_{m,k}^U = f(\boldsymbol{u}_k, \hat{\boldsymbol{\theta}}) + s\sqrt{\boldsymbol{V}_k^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{V}_k}\sqrt{n_p\, F_{(n_p, n_t - n_p; \alpha)}}. \tag{9}$$

Even though the above approach is willingly used in many practical applications [2], one has to be cautious about the results obtained in this way. In fact, the linearization method (7) leads to the reliable results only if the nonlinear model can be accurately approximated by a linear one at certain operating conditions. This however is not necessarily fulfilled, and many examples of such models can be found in [1].

## 4 Evolutionary bands for the expected response

In order to avoid linearization approach or other extremely computational intensive method used in regression analysis [1, 2, 3], in this section an alternative way of handling with confidence intervals is presented. The lower $y_{m,k}^L$ and upper $y_{m,k}^U$ adaptive thresholds for the predicted response of the nonlinear model (6) can be computed by solving the following optimization problems:

$$y_{m,k}^L = \arg\min_{\boldsymbol{\theta} \in \Theta} f(\boldsymbol{x}_k, \boldsymbol{\theta}) \tag{10}$$

$$y_{m,k}^U = \arg\max_{\boldsymbol{\theta} \in \Theta} f(\boldsymbol{x}_k, \boldsymbol{\theta}) \tag{11}$$

The feasible parameters set $\Theta$, which for linear models is an ellipsoid (4), in the case of nonlinear models, has much more complex shape [1]:

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{n_p} \middle| S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) \leq s^2 \, n_p \, F_{(n_p, n_t - n_p; \alpha)} \right\}, \tag{12}$$

The nonlinear function $f(\cdot)$ may possess more than one local optima, thus it seems reasonable to use some of global optimization techniques to solve (10) and (11). In this paper the $ESSS$ algorithm (Evolutionary Search with Soft Selection) [4] is utilized to solve (10),(11).The evolution process can be described in a few words.
At the very beginning, the population of solutions is randomly generated by adding an isotropic $\alpha$-stable vector $S_\alpha S$ [7] to the estimate point (2), which naturally belongs to the feasible parameters set $\Theta$. In succeeding steps algorithm performs alternately evaluation, selection and mutation operators after which the correction procedure 1 must be carried out. Since in considered problem the matter of obeying constrains plays a major role, it is not recommended to use such approaches as penalty functions [6]. Instead of this, in the Tab. 1 a very simple and efficient method for the projection of an individual into border of $\Theta$ is presented. The evolutionary process stops after a maximum number generation is reached. In order to prevent the correctness algorithms 1 from a burden of computation it is not recommended to choose the small value for the stable index $\alpha$ [7] because heavier tails distributions generates more of the so-called macro-mutations that may lay outside the feasible set $\Theta$.

## 5 Simulations experiment

In order to show that the proposed approach provides more accurate adaptive thresholds for residual signals, let us introduce the following experiment. Lets consider one input, one output neural network consisted of three hidden neurons with hyperbolic tangent transfer function and one linear output unit. The total number of parameters is equal to $n_p = 14$. It must to be stress that in the experiment parameters $\boldsymbol{\theta}^*$ of the network were chosen randomly. The neural model prepared in this way had served as a deterministic part in (6). The training set $\{x_k, y_k\}_{k=1}^{18}$ was collected by equally divide interval $[0, 10]$ - $\{x_k\}$. In order to simulate the noise $\epsilon_k$, the sequence of non-correlated Normal pseudo-numbers $\xi_k \sim N(0, 0.1^2)$ was added e.g. $y_k = f(x_k, \boldsymbol{\theta}^*) + \xi_k$. Next, the training set was used to obtain estimates $\hat{\boldsymbol{\theta}}$ for the neuron network of the same structure (for this purpose the well-known Levenberg-Marquardt method was used). For test purposes, the set of twenty points was chosen in such a way to uniformly cover the interval $[0, 10]$. In Fig. 1 (a) the comparison of adaptive thresholds (at $1 - \alpha = 0.9$ confidence level) for residual signals is presented. In order to check the adequacy of the bands obtained for both methods, the set of two hundred systems responses for each tested point was generated. The percentage of the residuals that lay inside adaptive thresholds can be observed in Fig. 1 (b).

**Remark 1.** In the case of nonlinear models, the linearization method may overestimate the value of the bands for expected response. In Fig. 1 (b) one can observe that the bands obtained for linearized model, cover nearly 100% percent of all system responses. It may cause serious problems in many engineering application e.g. model-based fault detection systems [5, 8] where sensitivity to the faulty conditions plays a key role.

**Table 1.** The outline of the projection algorithm

---

*Input data*

$Eps$ – absolute accuracy of localization of the border;

$\boldsymbol{\theta}, \boldsymbol{\theta}'$ – parent and individual after mutation respectively;

$R = s^2\, n_p\, F_{(n_p, n_t - n_p; \alpha)}$ – radius of the feasible set;

*Output data*

$\boldsymbol{\theta}^*$ – the individual projected onto border;

*Algorithm*

$\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$

$h \leftarrow ||\boldsymbol{\theta}^* - \boldsymbol{\theta}'||_2$

Repeat

  If $S(\frac{1}{2}(\boldsymbol{\theta}^* + \boldsymbol{\theta}')) - S(\boldsymbol{\theta}) > R$ then

    $\boldsymbol{\theta}' \leftarrow \frac{1}{2}(\boldsymbol{\theta}^* + \boldsymbol{\theta}')$

  else

    $\boldsymbol{\theta}^* \leftarrow \frac{1}{2}(\boldsymbol{\theta}^* + \boldsymbol{\theta}')$

  end if

  $h \leftarrow h/2$

Until $(h > Eps)$

---

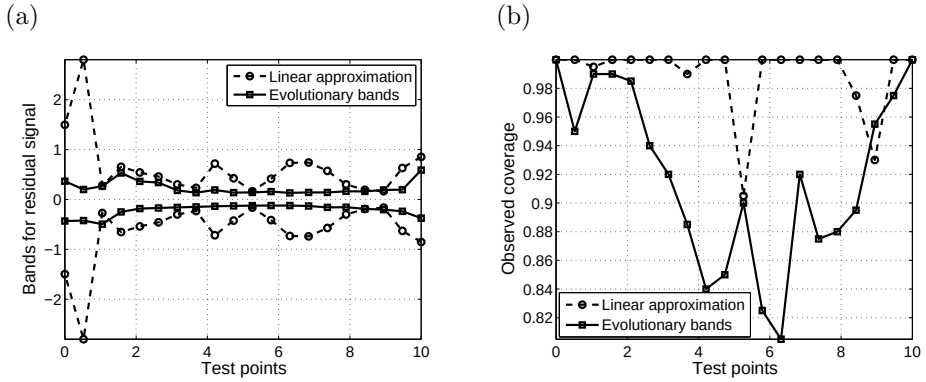


**Figure 1. (a)** – $1 - \alpha = 0.9$ bands for the expected responses of the neural model, and **(b)** – their observed coverage.

## 6 Concluding remarks

This paper constitutes an attempt to meet the needs created by practical application of nonlinear regression analysis, through the development of new techniques of calculating confidence intervals for predicted model response. It is important to note that approaches based on the linearization method and estimation covariance matrix, may not give an adequate description of the parametric uncertainty of nonlinear models. Such situations occur very often in the case of neural networks which are very often applied to the model-based fault detection systems. Numerical simulations showed that for models which are characterized by high degree of nonlinearity, the linearization approach may lead to misleading conclusions. Thus, in this paper, making use of evolutionary algorithms, more reliable and precise description of confidence intervals is proposed. An application of evolutionary algorithms to this problem allow one to deal with nonlinearity of regression models directly, without any other approximation and exempts the user from utilization of other highly computational intensive methods.

## Bibliography


[1] Bates, D.M. and Watts, D.G. *Nonlinear Regression Analysis and Its Application.* Willey & Sons,New York, 1988.

[2] Chryssolouris G., Lee M. and Ramsey A.*Confidence interval prediction for neural network models.* – IEEE Trans. Neural Networks, Vol.7, No.1, pp.229–232, 1996.

[3] J. R. Donaldson, R. B. Schnabel .*Computational Experience With Confidence Regions and Confidence Intervals for Nonlinear Least Squares.* – Technometrics , Vol.29, No.1, pp.67–82, 1987.

[4] R. Galar, *Evolutionary search with soft selection.* – Biological Cybernetics ,Vol.60, 1989, pp.357–364.

[5] Korbicz J., Kościelny J.M., Kowalczuk Z. and Cholewa W. (Eds.) (2004): *Fault Diagnosis. Models, Artificial Intelligence, Applications.* – Berlin: Springer-Verlag.

[6] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer–Verlag, Berlin Heidelberg, 1996.

[7] A. Obuchowicz,P.Prętki ,*Phenotypic evolution with mutation based on symmetric $\alpha$-stable distribution*, – International Journal of Applied Mathematics And Computer Science, 2004, vol. 14, No. 3,289-316.

[8] P.Pretki, M.Witczak, *Assessment and minimization of parametric uncertainty for multi-output neural networks - application to fault diagnosis* In: Recent Developments in Artificial Intelligence Methods - AI-METH 2005. Gliwice, Polska , pp. 155–160.

[9] R.D. De Veaux, J. Schweinsberg, J. Schumi, L. H. Ungar .*Prediction Intervals for Neural Networks via Nonlinear Regression.* – Technometrics , Vol.40, No.4, pp.273–282, 1998.

[10] M. Witczak, *Advances in model-based fault diagnosis with evolutionary algorithms and neural networks.* – International Journal of Applied Mathematics And Computer Science, 2006, vol. 16, No. 1, 85–99.