Applied Soft Computing xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Benchmark problems for nonlinear system identification and control using Soft Computing methods: Need and overview

Andreas Kroll [a,*], Horst Schulte [b]

[a] University of Kassel, Measurement and Control Department, Mechanical Engineering Faculty, Germany
[b] HTW University of Applied Science Berlin, Department of Engineering I, Control Engineering, Germany

ARTICLE INFO

*Article history:*
Received 6 September 2013
Received in revised form 9 July 2014
Accepted 14 August 2014
Available online xxx

*Keywords:*
Benchmarking
Nonlinear control
Nonlinear system identification

ABSTRACT

Using benchmark problems to demonstrate and compare novel methods to the work of others could be more widely adopted by the Soft Computing community. This article contains a collection of several benchmark problems in nonlinear control and system identification, which are presented in a standardized format. Each problem is augmented by examples where it has been adopted for comparison. The selected examples range from component to plant level problems and originate mainly from the areas of mechatronics/drives and process systems. The authors hope that this overview contributes to a better adoption of benchmarking in method development, test and demonstration.



## 1. Introduction

Once new Soft Computing (SC) methods are developed, it is of key interest to compare their performance in relation to state-of-the-art methods in order to position them. However, such comparisons seldom take place due to the efforts of adopting sufficient insight in and control over methods other than the own research focus. This problem can be circumvented if benchmark problems are adopted more widely such that one can retrieve competing results from literature without having to master other methods. In fact, well-established benchmark problems are available for problems such as classification, control and modeling, to name a few. The objective of this article is to present selected (nonlinear) benchmark problems for identification and control, and to promote a wider adoption to provide a framework for comparing alternative SC methods. A word of caution is due: Methods typically work better on some and worse on other problems. Hence, good results on a single benchmark problem highlight advantages for respective problem types but should not be generalized.

Benchmarking is based on the principles of the ability to be validated, reproducibility and comparability. This requires exact specification of benchmark problems spanning from the process description over experiment design, test data to assessment criteria for obtained results. Unfortunately, even complete and commonly adopted "benchmark" problems often do not provide a complete, self-enclosed description. A typical situation is that a process model to be used is described well; however engineering details, experiment/test design and assessment criteria are incomplete or lacking.

In the following section, sets of assessment criteria are proposed for modeling and control. Section 3 contains the benchmark problems. The presented selection does not attempt to be complete but rather offer attractive problems for the Soft Computing community. The sequence was designed to start from simple component and ascend to plant level problems. This article is an extended version of the identification and control part of a position paper written for the German GMA technical committee on Computational Intelligence [33].

## 2. Assessment criteria

### 2.1. Model performance

Criteria to assess the results of modeling tasks can address approximation quality, model complexity and model interpretability [47]. Most commonly, the approximation/prediction error is used as assessment criterion. Most significant is the result for validation/test rather than for the training data. Many different criteria are proposed as e.g. sometimes the worst case and sometimes the average deviation may be more important. In case of benchmark problems, it is recommended to report a few widely accepted criteria such as a subset of the following ones: Given $N$ data sets where

* Corresponding author. Tel.: +49 5618043248.
*E-mail addresses:* andreas.kroll@mrt.uni-kassel.de (A. Kroll), horst.schulte@htw-berlin.de (H. Schulte).

http://dx.doi.org/10.1016/j.asoc.2014.08.034
1568-4946/

$y(k)$ is the output of a system and $\hat{y}(k)$ the corresponding output of the model, this could be the *maximum absolute error* (MAE)

$$J_{\text{MAE}} = J_{\max} = \max_{1\le k\le N} \left| y(k) - \hat{y}(k) \right| , \qquad (1)$$

the *sum of squared errors* (SSE)

$$J_{\text{SSE}} = \sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2 , \qquad (2)$$

the *mean squared error* (MSE)

$$J_{\text{MSE}} = \frac{1}{N} \sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2 , \qquad (3)$$

and/or the *root mean squared error* (RMSE)

$$J_{\text{RMSE}} = \sqrt{J_{\text{MSE}}}. \qquad (4)$$

The measure *variance accounting for* (VAF)

$$J_{\text{VAF}} = \left( 1 - \frac{\text{Var}(y(k) - \hat{y}(k))}{\text{Var}(y(k))} \right) \cdot 100\,\% , \qquad (5)$$

origins from linear regression where it provides the percentage of the variance of $y$ that can be explained by the used linear regression model.[1] $\text{Var}(\circ)$ gives the variance of $\circ$. $J_{\text{VAF}}$ can be estimated by the coefficient of determination $R^2$

$$R^2 = 1 - \frac{\sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2}{\sum_{k=1}^{N} (y(k) - \bar{y})^2} \quad \text{with} \quad \bar{y} = \frac{1}{N} \sum_{k=1}^{N} y(k). \qquad (6)$$

To admit comparing models with different numbers of parameters $R^2$ is adjusted to

$$R_a^2 = 1 - \frac{N-1}{N - (dim(O) + 1)} \cdot (1 - R^2). \qquad (7)$$

A related measure is the *normalized mean squared error* (NMSE)

$$J_{\text{NMSE}} = \frac{\sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2}{\sum_{k=1}^{N} (y(k) - \bar{y})^2} , \qquad (8)$$

and the best fit rate (BFR)

$$J_{\text{BFR}} = \left( 1 - \frac{\sqrt{\sum_{k=1}^{N} \left( y(k) - \hat{y}(k) \right)^2}}{\sqrt{\sum_{k=1}^{N} (y(k) - \bar{y})^2}} \right) \cdot 100 \ \% \qquad (9)$$

Note, that $J_{\text{VAF}}$ and $J_{\text{BFR}}$ can take negative value, which are typically replaced by 0. $k$ is the discrete time with $t = k\,T_0$, $T_0$: sampling time. Common practice is to additionally asses the frequency distribution of the residual $\epsilon(k) = y(k) - \hat{y}(k)$ wrt. to mean, shape and symmetry of the tails. This provides a qualitative indication whether the residuals are normally distributed.

It is important to differentiate between one-step-ahead and recursive model evaluation: In the first case measurements available until present time $k$ are used to predict the output $\hat{y}(k+1)$ one-step-ahead into the future

$$\hat{y}(k+1) = f(y(k), \ldots, y(k-n), \mathbf{u}(k-\tau), \ldots, \mathbf{u}(k-\tau-m)) , \qquad (10)$$

where $n$ and $m$, respectively, is the number of lagged terms considered and $\tau$ a discrete dead-time. In a second case, lagged predictions are used as model inputs instead of measured data:

$$\hat{y}(k+1) = f(\hat{y}(k), \ldots, \hat{y}(k-n), \mathbf{u}(k-\tau), \ldots, \mathbf{u}(k-\tau-m)). \qquad (11)$$

[1] Note, that it loses the original statistical interpretation when applied to nonlinear models.

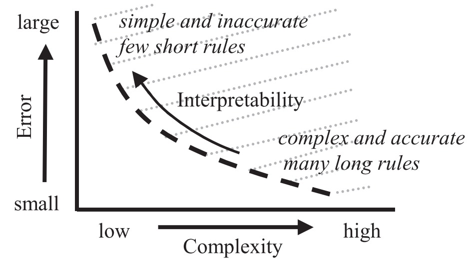


**Fig. 1.** Illustration of a typical relationship between approximation quality, complexity and interpretability of a model [47].

Good recursive model evaluation results are more difficult to achieve than good one-step-ahead predictions. In case models of different complexity yield similar approximation performance, the simpler model should be preferred ("Ockham's principle"). Model-complexity-oriented criteria take this into account by valuing the number of parameters (as a measure for model complexity) besides the approximation error. An example of such a criterion is Akaikes's Final Prediction Error (FPE)

$$J_{\text{FPE}} = \frac{1 + dim(\mathbf{\Theta})/N}{1 - dim(\mathbf{\Theta})/N} \; \frac{1}{N} \sum_{k=1}^{N} \left( \hat{y}(k, \mathbf{\Theta}) - y(k) \right)^2 . \qquad (12)$$

In case of multi-input multi-output systems above recorded, criteria can be assessed individually for the outputs. Alternatively, the quantities can be scaled and aggregated to a single metric as in Section 3.9.

Other criteria are recorded e.g. in [61]. Ease of model interpretability is a meaningful concept for comparing fuzzy and neuro-fuzzy rule-based models, see Fig. 1. However, interpretability is difficult to define as a metric criterion.

### 2.2. Control performance

Criteria to assess the control performance include the step response, reference tracking, disturbance rejection behavior and the control effort. This is analyzed for the nominal case and in some benchmark problems also with predefined structured model uncertainties. The latter is application-dependent. However, the general procedure is discussed in this section. Therefore, details will be given in relation to the individual benchmarks.

As previously described for the modeling task, widely accepted criteria and related measures will be analogously given in the following: The *step response* is used to characterize the accuracy, damping and speed of the closed loop system. To measure the steady state accuracy the integrated absolute error (IAE) is given by

$$J_{\text{IAE}} = \sum_{k=1}^{N} |y(k) - y_{ss}|\, T_0 \qquad (13)$$

with $y_{ss}$ as the steady-state response. Common practice is to measure the maximum percent overshoot (PO)

$$J_{\text{PO}} = \frac{\max_{1\le k\le N} (y(k) - y_{ss})}{y_{ss}} \cdot 100 \qquad (14)$$

and maximum percent undershoot (PU)

$$J_{\text{PU}} = \frac{\max_{1\le k\le N} (-y(k))}{y_{ss}} \cdot 100 \quad . \qquad (15)$$

The *reference tracking criteria* are concerned with the response of the closed-loop system to the time-variable reference $r(t)$ alone,