Input-output data sets for development and benchmarking in nonlinear identification

Torbjörn Wigren

Division of Systems and Control, Department of Information Technology, Uppsala University, PO box 337, SE-75105 Uppsala, Sweden. E-mail torbjorn.wigren@it.uu.se.

Abstract

This report presents two sets of data, suitable for development, testing and benchmarking of system identification algorithms for nonlinear processes. The first data set is recorded from a laboratory process that can be well described by a block oriented nonlinear model. The data set is challenging; it consists of only 500 samples, the nonlinear effect is large and the damping is not too good. The second data set is recorded from a laboratory process known to be governed by nonlinear differential equations.

1. Introduction

The field of non-linear system identification is developing rapidly today, with several powerful algorithms being available in standard software packages like the MATLAB System Identification Toolbox developed by L. Ljung (The Mathworks, 2010). That toolbox e.g. supports black-box methods based on NARX (Ljung, 1997), neural networks (Sjöberg and Ljung, 1992), block-oriented models (Billings and Fakhouri, 1982), as well as interfacing to grey-box models and algorithms (Bohlin, 1994).

However, the development and characterization of new algorithms for identification of non-linear dynamic systems are still challenging tasks. Simulated data and theoretical analysis are two standard tools used in this development. The use of measured data is reported more seldom, though. This is regrettable, since lack of testing against real data may result in a too poor robustness and algorithms that are not suitable for the operating conditions encountered in many applications.

One reason for the limited experimental testing is the limited availability of publicly available sets of data, suitable for experiments. Such data should resemble significant nonlinear effects. At the same time it is an advantage when large parts of the measured data can be explained by a not too complex mathematical model. In such cases the data can also often be used for development of both black-box and grey-box algorithms. In addition, the modeling problem then has a reasonable size, and validation against a well defined "truth" allows conclusions to be drawn on the qualities of the applied methods. The sampling process needs to be well defined, in order to allow for the use of both continuous time and discrete time models.

This report contributes by the presentation of two freely downloadable sets of input-output data for nonlinear identification The first set is recorded from the coupled electric drives (Wellstead, 1979; Wigren,1990). The speed control dynamics of that laboratory process can be accurately modeled by a third order block oriented nonlinear model. The second process consists of two cascaded tanks (Wigren, 2006) with free outlets. This laboratory process can be modeled with two coupled non-linear ordinary differential equations.

Section 2 and 3 describe the processes, while the recorded data is discussed in section 4. Previously obtained identification results using the data are summarized in section 5. The conclusions follow in section 6.

2. A third order block oriented nonlinear system

The CE8 coupled electric drives (Wellstead, 1979) consists of two electric motors that drive a pulley using a flexible belt. The system is depicted in Fig. 1. The pulley is held by a spring, resulting in a lightly damped mode. The electric drives can be individually controlled allowing the tension and the speed of the belt to be simultaneously controlled. The drive control is symmetric around zero, hence both clockwise and counter clockwise movement is possible. Here the focus in only on the speed control system. The reason is that the angular speed of the pulley is measured with a pulse counter and this sensor is *insensitive to the sign of the velocity*. Following the sensor, analogue low pass filtering and anti-aliasing filtering is applied. The dynamic effects are generated by the electric drive time constants, the spring and the analogue low pass filtering. The latter has a quite limited effect on the output and may be neglected.

By considering the sum of the voltages applied to the motors as the input, physical modeling results in a lightly damped linear third order system, as measured from voltages to pulley velocity. The pulley velocity is rectified by the pulse counter to give the speed, after which it is filtered. A tentative discrete time description of this system is

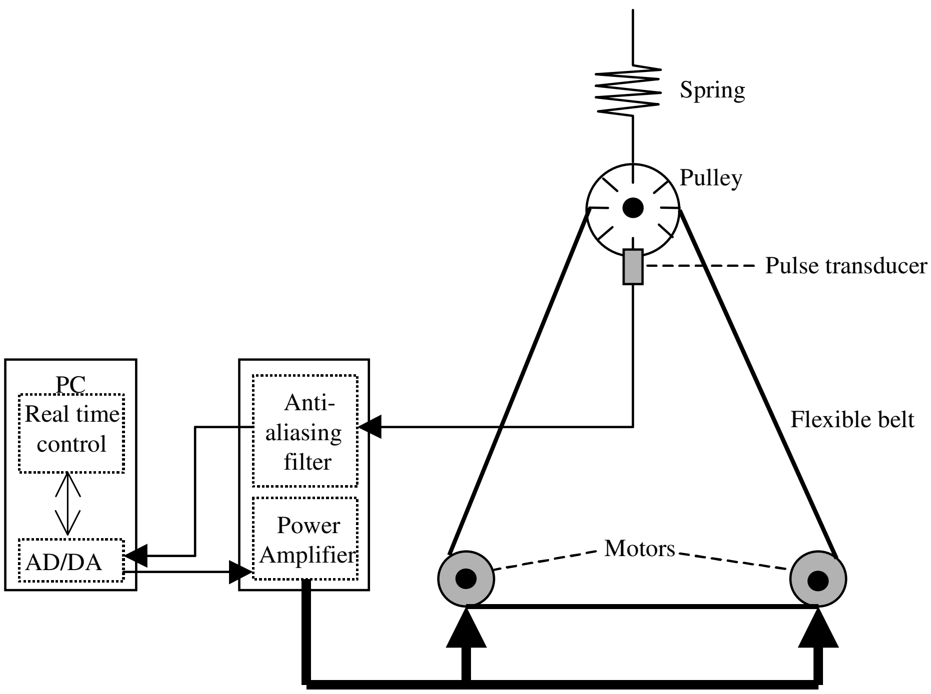

Fig. 1: The CE8 coupled electric drives.

$$v(t) = \frac{b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}}{1 + f_1 q^{-1} + f_2 q^{-2} + f_3 q^{-3}} u(t) + w(t) \tag{1}$$

$$s(t) = |v(t)| + e(t) \tag{2}$$

$$y(t) = \frac{c_1 q^{-1} + \ldots + c_n q^{-n}}{1 + d_1 q^{-1} + \ldots + d_n q^{-n}} s(t). \tag{3}$$

Here t is discrete time, $u(t)$ is the speed, $w(t)$ is the systems noise, $s(t)$ is the speed, $e(t)$ is pulse counter noise, and $y(t)$ is the measured filtered output. The remaining symbols represent the polynomial coefficients of the filter models. The exact model of the analogue filtering is not known, however the bandwidth of the anti-aliasing filter is 12 Hz. This is less than the sensor bandwidth. Further information of the process is available in Wellstead, (1979), and in Wigren, (1990).

3. A second order nonlinear state space system

The second process is a fluid level control system consisting of two cascaded tanks with free outlets fed by a pump. The water is transported by the pump to the upper of the two tanks. The process is depicted in Fig. 2. The input signal to the process is the voltage applied to the pump and the two output signals consist of measurements of the water level of the tanks. Since the outlets are open, the result is a dynamics that varies nonlinearly with the level of water. The process is controlled from a PC equipped with MATLAB interfaces to the A/D and D/A converters.

The laboratory process is suitable for physical modeling. Application of Bernoulli's principle and conservation of mass results in

$$\begin{pmatrix} \frac{dh_1}{dt} \\ \frac{dh_2}{dt} \end{pmatrix} = \begin{pmatrix} -\frac{a_1\sqrt{2g}}{A_1}\sqrt{h_1} + \frac{1}{A_1}ku(t) \\ -\frac{a_2\sqrt{2g}}{A_2}\sqrt{h_2} + \frac{a_1\sqrt{2g}}{A_2}\sqrt{h_1} \end{pmatrix} + \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} h_1(t) \\ h_2(t) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \end{pmatrix}. \tag{5}$$

Here h_1 and h_2 denote the levels of the upper and the lower tank, respectively. $w_1(t)$ and $w_2(t)$ are system noises. The outputs are given by $y_1(t)$ and $y_2(t)$, these are corrupted by the measurement disturbances $e_1(t)$ and $e_2(t)$. The areas of the tanks are A_1 and A_2 while the effluent areas are denoted a_1 and a_2. The gravity is denoted by g, the voltage to input flow conversion constant by k and the applied voltage to the pump by $u(t)$.

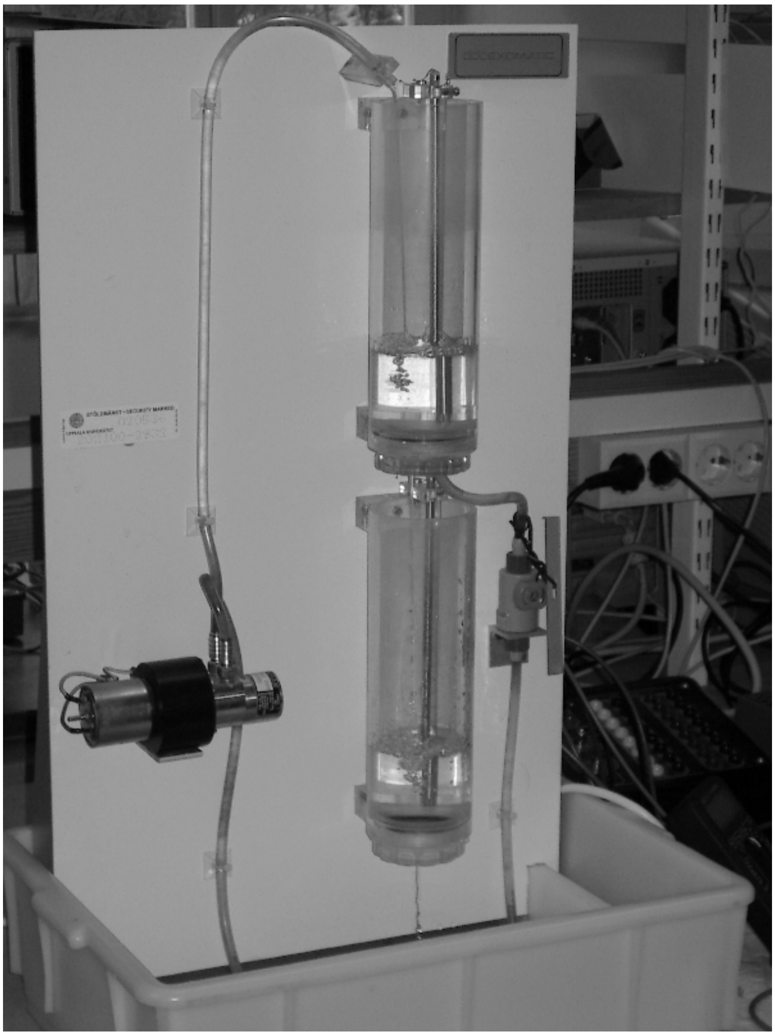

Fig. 2. The cascaded tanks.

4. Data

All data was collected in open loop experiments using zero-order hold (ZoH) sampling.

Two types of inputs were used for the coupled electric drives. The first input signal was a PRBS with a clock period of 5 times the sampling period. The signal was switching between $-u_{PRBS}$ V and $+u_{PRBS}$ V, resulting in the process changing the belt rotation direction frequently. Three realizations were recorded for $u_{PRBS} = 0.5, 1.0, 1.5$. The input-output data was collected with a sampling period of 20 ms. The second type of input signal was obtained from a PRBS with a clock period of 5 times the sampling period, switching between -1.0 V.and +1.5 V (first realization), as well as between -1.0 V and +3.0V (second realization). The signal in each clock interval of constant signal level and with a duration of 5 sampling periods, was then multiplied with a random number, uniformly distributed in amplitude between 0 and 1. The resulting input signal is then uniformly disributed in amplitutude. The reason for this is that when the system is nonlinear, both the frequency and amplitude contents of the input signal are important for identification (Wigren, 1990). The input-output data was again recorded with a sampling period of 20 ms. The input-output data obtained from the first realization of the input signal is depicted in Fig. 3.

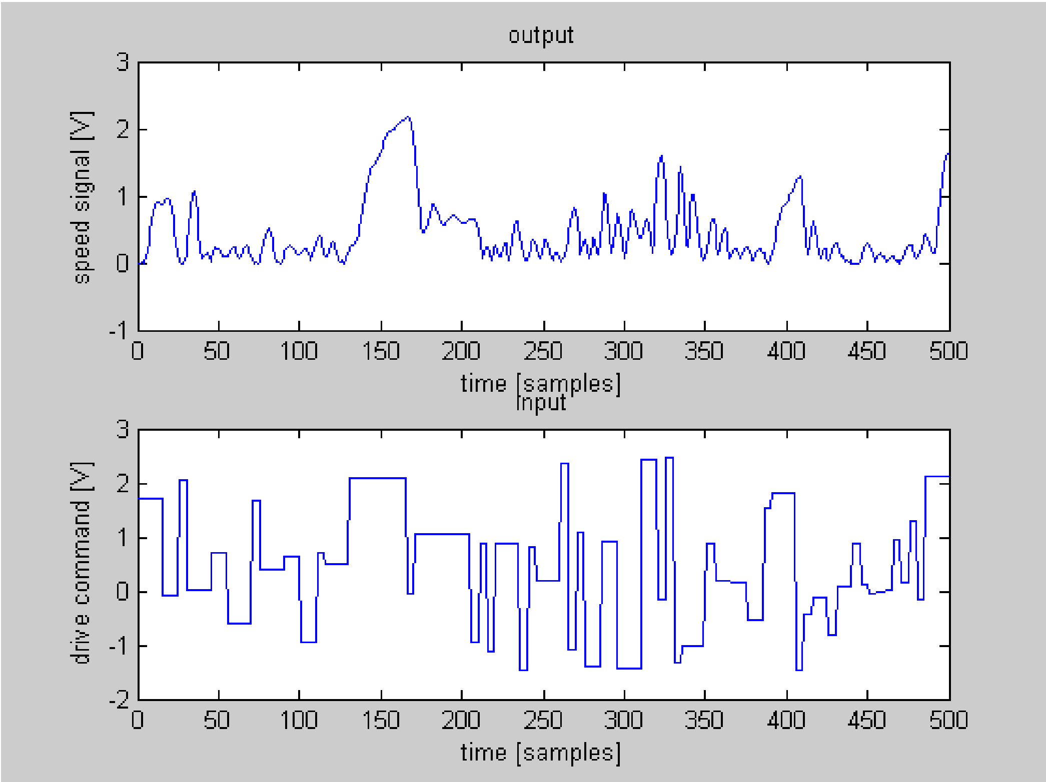

Fig. 3. Coupled electric drives input-output data.

The data that was recorded from the cascaded tanks also used an input signal that was generated as the uniformly distributed input signal above. The data of the file Tank1.mat used a clock period of 30, a sampling period of 5.0 s and provide 2500 samples of input-output data for both the upper and lower tank. In Tank2.mat a clock period of 15 samples was used, the sampling period was 4.0 s, and 7500 samples of data were recorded.

The data files available for download are summarized in Table 1.

System	File	Input	Output	Comment
Drives	DATAPRBS.mat	u1	z1	
Drives	DATAPRBS.mat	u2	z2	Wigren, 1990
Drives	DATAPRBS.mat	u3	z3	
Drives	DATAUNIF.mat	u11	z11	Wigren, 1990
Drives	DATAUNIF.mat	u12	z12	
Tanks	Tank1.mat	u	y	Wigren, 2006
Tanks	Tank2.mat	u	y	

Table 1. Summary of downloadable data.

5. Previous identification results

The coupled electrical drives process was identified using a recursive prediction error method based on the nonlinear Wiener model in Wigren, (1990). The cascaded tank process was recursively identified with a prediction error method, based on a general state space model in Wigren, (2006). The reader is referred to those publications for details of the results. It is expected that these results can be significantly improved since they are based on recursive algorithms.

6. Conclusions

This report has presented two freely downloadable sets of data, suitable for support of the development of new nonlinear system identification algorithms. More contributions of downloadable data is needed to support researchers with diverse data addressing the very wide class of nonlinear dynamic systems.

References

Billings, S. A. and S. Y. Fakhouri (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, **18**, 15-26.

Bohlin, T, (1994). A case study of grey box identification, *Automatica*, **30**, pp. 307-318.

Ljung, L. (1997). Non-linear black box models in system identification. *Proc. IFAC Symposium on Advanced Control of Chemical Processes, ADCHEM'97,* Banff, Canada, 1-13.

Sjöberg, J. and L. Ljung (1992). Overtraining, regularization, and searching for minimum in neural networks. *Proc. Symp. On Adaptive systems in Control and Signal Processing.* Grenoble, Switzerland.

The Mathworks, (2010). MATLAB System Identification Toolbox. *Available:* http://www.mathworks.com/products/sysid/ .

Wellstead, P. E. (1979). Introduction to Physical System Modeling. Academic Press, London, UK.

Wigren, T. (1990). Recursive identification based on the nonlinear Wiener model. Ph.D thesis, Acta Universitatis Upsaliensis, Uppsala Dissertations from the Faculty of Science 31, Uppsala University, Uppsala, Sweden. Available: http://user.it.uu.se/~tw/

Wigren, T. (2006). Recursive prediction error identifiction and scaling of non-linear state space models using a restricted black-box parameterization, *Automatica*, **42**, 159-168.