International Journal of Control
Vol. 00, No. 00, Month 200x, 1–7

RESEARCH ARTICLE

Efficient Parameterization of Nonlinear System Models: a Comment on Nöel and Schoukens (2018)

Peter C. Young[1], and A. Janot[2]
[1]Systems and Control Group, Lancaster Environment Centre, Lancaster University, UK
[2]ONERA, Toulouse, France

(*Received 00 Month 200x; final version received 00 Month 200x*)

Nöel and Schoukens (2018) discuss a methodology for the discrete-time state-space identification of nonlinear systems and apply this to experimental data from the well known Silverbox nonlinear circuit, producing a model characterised by 13 parameters. This model explains the data very well but the parameter estimates are not well defined in the optimization results, with the very large confidence bounds suggesting that the model is over-parameterized. This comment shows that this is indeed the case and that the data can be explained equally well by an alternative continuous-time, State-Dependent Parameter (SDP) transfer function model with only 6 parameters, the estimates of which are well defined with very tight confidence bounds. The comment also raises questions about how the model form for nonlinear systems such as the Silverbox should be identified and suggests that the Data-Based Mechanistic (DBM) approach to modelling has some advantages in this regard.

Keywords: system identification, silverbox system, nonlinear modelling, continuous-time model, efficient parameterization.

1 Introduction

In their paper *Grey-box State-space Identification of Nonlinear Mechanical Vibrations*, Nöel and Schoukens (2018) introduce a powerful method of 'grey-box' nonlinear modelling based on a general discrete-time state-space model form and show how a 13 parameter model of this type is able to explain experimental data from the Silverbox benchmark system (see e.g. Wigren and Schoukens 2013, Marconato et al. 2012) very well and in a parametrically more efficient manner than previous models. This paper prompted us to look at the data analysed by the authors and the present comment follows from our own analysis of these data. It shows that it is possible to achieve still greater parametric efficiency by following the *Data-Based Mechanistic* (DBM) approach to modelling nonlinear dynamic systems using a continuous-time *State Dependent Parameter Transfer Function* (SDPTF) model. The main modelling results are outlined in this comment but the full details of the analysis are available in an associated technical note Young (2018)[1].

2 Re-examination of the Nöel and Schoukens (2018) Model

Nöel and Schoukens (2018) prescribe the equation of motion for the Silverbox system on the basis that the circuit is intended to mimic the behaviour of a mechanical system and suggest the following equation,

$$M\frac{d^2y(t)}{dt^2}+C_v\frac{dy(t)}{dt}+Ky(t)+c_1y^2(t)+c_2y^3(t)=p(t) \tag{1}$$

[1]This can be downloaded from the 'Publication Downloads' section of `http://captaintoolbox.co.uk/Captain_Toolbox.html/Captain_Toolbox.html`: PCY Technical Report 2018 TN5

ISSN: 0020-7179 print/ISSN 1366-5820 online

DOI: 10.1080/0020717YYxxxxxxxx
http://www.informaworld.com

that reflects this mechanical engineering heritage. For later reference, however, this equation can be written in the alternative 'reduced form':

$$\frac{d^2x(t)}{dt^2} + a_1\frac{dx(t)}{dt} + k_1x(t) + k_2x^2(t) + k_3x^3(t) = u(t); \quad y(k) = x(k) + \xi(k) \tag{2}$$

where $a_1 = C_v/M$, $k_1 = K/M$, $k_2 = c_1/M$, $k_3 = c_2/M$, $u(t) = p(t)$ and $y(k)$ is the sampled measurement of $x(t)$, with $\xi(k)$ representing the model error and any noise on the measurement. One assumes that this differential equation is the initial hypothesis about the nature of the model and expects initially that the modelling will proceed on the basis of this equation. But this is not the case; rather the authors immediately move away from this standard SISO differential equation to a discrete-time state-space alternative, following a modelling approach they have described previously in the paper. It is reasonable ask, therefore, whether the parameterization of the resulting model is satisfactory in relation to the initially hypothesised model (2). Here, the term 'parameterization' is being used to mean the number and location of the parameters in the model structure, so that the term 'efficient parameterization' in the title of this comment means that the model is explained by as few, statistically well-defined, parameters as are necessary to explain the data well *and* meet the objectives of the modelling study.

The full details of the Nöel and Schouken's 13 parameter state space model are not specified in their paper but, on request, Dr. Nöel kindly supplied these. The model has the following form:

$$\begin{aligned} x(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k) + \mathbf{E}\begin{bmatrix} x^2(k) \\ x^3(k) \end{bmatrix} \\ y(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{D}u(k) \end{aligned} \tag{3}$$

This was simulated in Simulink^{TM} and used as a basis for nonlinear least squares optimization, using the Matlab^{TM} routine lsqnonlin, based on the first 10000 samples from the data set but omitting the first 2500 samples to avoid the quite long-term initial condition effects. The resulting optimized parameters do not change very much from those supplied by Dr. Nöel and the coefficient of determination (see e.g. Young 2011, page 176), based on the error between the measured output $y(k)$ and the similarly sampled output $x(k)$ of the model is $R_T^2 = 0.99998$; i.e. the simulated model output explains 99.998% of the measured output variance. The optimized parameters are as follows[1]:

$$\begin{aligned} &\mathbf{A} = \begin{bmatrix} 0.9712 & 0.1828 \\ -0.1665 & 0.9806 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} -0.04952 \\ -0.1388 \end{bmatrix} \\ &\mathbf{E} = \begin{bmatrix} -0.01228 & 0.1955 \\ -0.03453 & 0.5477 \end{bmatrix}; \mathbf{C} = \begin{bmatrix} -1.1375 & 0.1833 \end{bmatrix}; \\ &\mathbf{D} = .002649 \end{aligned} \tag{4}$$

It is clear that this model performs very well in explaining the data. But is it efficiently parameterized in the above sense?

One clue to answering this question is to consider the confidence bounds provided by the confint routine in Matlab^{TM} based on the results from the lsqnonlin optimization. Unfortunately, although the specially designed input stimulus $u(k)$ is clearly sufficient to ensure identifiability, the Jacobian returned by confint indicates conditioning problems and these bounds are extremely wide. This is hardly surprising because it is well known that linear and nonlinear state space models (see e.g. Ljung 1987, Wigren 2006) are inherently over-parameterized unless specified in some canonical form (particularly if, as in Nöel and Schoukens (2018), the optimization is initiated using a completely black box, nonlinear subspace method of model identification that does not yield a model in such a canonical form).

In order to both investigate the possibility of over-parameterization and compare the results from this

[1]Note that the parameter estimates shown here are rounded to four significant figures and this degrades the R_T^2.

type of model with those obtained later in section 3 using an alternative continuous-time transfer function model, it makes sense to develop and optimize a similar nonlinear state space model in differential equation form. This model explains the data to about the same level as the discrete-time model, with $R_T^2 = 0.99998$; and, as expected, the confidence bounds provided by the confint routine are again extremely wide, suggesting that the 13 parameters are not well defined statistically and that the model is over-parameterized.

In order to solve this over-parameterization problem, the most obvious approach is to consider a canonical form for the continuous-time state space model where, in the linear part of the state equations, the second state is defined as the derivative of the first state. This not only reduces the number of parameters, it also provides a model where these parameters and the associated state variables have an immediate, more physical interpretation. The 8 optimized parameters in this case are as follows[1]:

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} -0.01744 & -0.03134 \\ 1.0 & 0.0 \end{bmatrix}; \; \mathbf{B} = \begin{bmatrix} 1.0 \\ 0.03086 \end{bmatrix} \\
\mathbf{E} &= \begin{bmatrix} 0.2458 & -3.8966 \\ 0.1236 & -2.0637 \end{bmatrix}; \; \mathbf{C} = \begin{bmatrix} -0.01631 & 0.03106 \end{bmatrix};
\end{aligned} \tag{5}$$

with D=0, i.e. no direct effect of the input on the output (note that the discrete-time and unconstrained continuous-time state space models have a very small values for this parameter). The explanation of the data is virtually the same as in the unconstrained model case, with $R_T^2 = 0.99998$, even though the number of parameters is reduced almost 40%, from 13 to 8. Moreover, the confidence bounds obtained from the confint routine are now very tight around the estimates and the associated standard errors on the parameters are very small indeed, demonstrating how well they are defined in statistical terms.

3 State-Dependent Parameter Transfer Function (SDPTF) Modelling

The model (5) is admirable in explaining the Silverbox data with only 8 parameters. In this section, however, it is shown that there are other *State Dependent Parameter Transfer Function* (SDPTF) model forms that are even more efficiently parameterized and, just as importantly in this example, have associated differential equation descriptions that relate directly to the originally hypothesised differential equation model form (2).

SDP modelling evolved many years ago (Young 1968, Mendel 1969, Young 1981) and Priestley (1980), who first used the name. The more advanced SDP methods used in the present technical note were developed in the 1990s (see e.g. Young 2000, 2001, Young et al. 2001) and, since then, these have been applied successfully to a large number of practical systems, in diverse areas of study. The SDP modelling procedure consists of three main steps. First, the sdp routine in CAPTAIN exploits optimal fixed interval smoothing estimation to identify, in non-parametric (graphical) terms, the nonlinear model structure, as defined by the location and nature of the statistically significant SDP nonlinearities. This is then followed by the parameterization of the identified SDP nonlinearities, using whatever method is most appropriate, if possible one that is transparent and has a useful physical interpretation. Finally, the parameters in the parameterized model are optimized in a similar manner to those of the state space models discussed in the previous section.

The non-parametric estimation results for the present Silverbox example are reported fully in Young (2018). The parameterized SDPTF model structure is identified from these non-parametric results and takes the following form:

$$x(t) = \frac{b_0 s + b_1}{s^2 + a_1 s + a_{sdp}} u(t); \quad a_{sdp} = k_1 + k_2 x(t) + k_3 x^2(t) \tag{6}$$

[1]Note that the $b_2 = 0.03086$ is not optimized because it only controls the gain of the system and can be set arbitrarily to any value, here to a value that relates to the transfer function modelling described in section 3.

or, in a differential equation form that can be linked directly to the initially hypothesised model (2),

$$\frac{d^2x(t)}{dt^2} + a_1\frac{dx(t)}{dt} + k_1x(t) + k_2x^2(t) + k_3x^3(t) = b_0\frac{du(t)}{dt} + b_1u(t) \tag{7}$$

From this, we see that the only difference is in the terms associated with the input variable (i.e. the numerator of the SDPTF model) and, in particular, the need for a $du(t)/dt$ term.

For convenience, the optimization of the model was carried out using the Simulink model with a normalised sampling interval of unity. Also, in order to avoid the initial transient behaviour and the associated need to remove the first 2500 samples, the initial conditions I_1 and I_2 on the two integrators in the model were optimized simultaneously with the model parameters. The resulting estimates are given below with the estimated standard errors in parentheses:

$$\begin{aligned}
&\hat{a}_1 = 0.01751(1.62\times10^{-6}); \hat{b}_0 = -0.1684(0.00015)\\
&b_1 = 0.03111(2.49\times10^{-6})); \ \hat{k}_1 = 0.03136(7.72\times10^{-7})\\
&\hat{k}_2 = -0.007884(9.17\times10^{-6}); \ \hat{k}_3 = 0.1231(1.96\times10^{-5})\\
&\hat{I}_1 = -0.06827(2.91\times10^{-5}); \hat{I}_2 = 0.03045(0.00037)
\end{aligned} \tag{8}$$

Note that the estimate of $\hat{b}_0 = -0.1684$ shows that the model has significant non-minimum phase characteristics. As this makes a quite large improvement to the explanatory power of the model, it appears to represent an important element in the system dynamics that, as far as we know, has not been recognised *explicitly* in previous Silverbox modelling studies. However, as J. P Nöel has pointed out to us (personal communication, 26th June, 2018), it is probably the reason why better estimation performance was obtained by Paduart et al. (2010) when they considered adding a multivariate polynomial in both the state and input variables, rather than a simpler description based on the states alone.

Finally, although the sampled residual error $\xi(k) = y(k) - x(k)$ in the model (7) is very small, it has significant serial correlation and is also correlated with the sampled input variable $u(k)$. If $\xi(k)$ is investigated further, the *Akaike Information Criterion* (AIC) identifies a very high AR(88) model that characterises $\xi(k)$ very well, with serially uncorrelated, white residuals $e(k)$ (one-step-ahead prediction errors) and very small residual variance. On the other hand, $e(k)$ is still correlated with the sampled input variable $u(k)$. This suggests that there is probably an additional, extremely low level of nonlinearity present in the data that has not been captured by the SDP nonlinear model (or, it would appear, by any previous models of the Silverbox). This is supported to some degree by the initial non-parametric SDP analysis (see the full details in Young 2018) which identifies some evidence of similar nonlinear effects in the first SDPTF denominator parameter a_1 (see next sub-section 3.1).

It might be argued theoretically that, because $e(k)$ is still correlated with the sampled input variable $u(k)$, the covariance matrix returned by the confint routine is not a valid representation of the uncertainty in the estimated parameters and so the standard error bounds in (8) are not strictly accurate estimates. Even if this argument is accepted, however, these bounds illustrate the relative uncertainties in the parameters and there is no doubt at all that they are very small indeed.

3.1 *Is there justification for a more complex model?*

Given the nature of the residual error in the 6 parameter SDP model, it makes sense to investigate whether the introduction of state-dependency into a_1 results in any significant improvement in the explanatory ability of the resulting parametric model (providing, of course, that there is no sign of over-parameterization that might question any such improvement). The full details of this analysis are given in (Young 2018), so it will suffice here to report that the optimized parameters in the resulting 8 parameter model are very well defined, with no signs of over-parameterization, and there is a small improvement in the variance of the model error when compared with the 6 parameter model (7) and the two state space models (3) and (5). Once again, however, this residual series $\xi(k)$, as well as the white residual $e(k)$ of an associated, AIC identified, AR(87) noise model, are still correlated with the input variable $u(k)$.

Disappointingly, therefore, we cannot claim that the enlargement of the model has purged much additional nonlinearity from the residual series, although the small improvement in the explanatory ability arising from the additional two parameters means that the resulting 8 parameter model is the best of all the models considered here in this regard.

It would be possible to continue to look for other SDP components in the model in order to try to explain the remaining, extremely small level of unexplained nonlinearity in the data. However, it seems reasonable to conclude that we have reached a stage of diminishing returns. To paraphrase Karl Popper (1959), we would contend that these DBM SDPTF models are conditionally validated and can be considered satisfactory for the most modelling objectives until they may be falsified by further experiment and analysis.

3.2 *The physical nature of the SDPTF model*

There are several advantages to the SDPTF model (7). For instance, although it exploits 'black-box' methodology, it does this using a DBM modelling approach, where the objective is to obtain a model that can be interpreted in the mechanistic terms that are most appropriate to the nature of the dynamic system. In this sense, therefore, the finally estimated model is the equivalent of a 'white-box' model because the optimized differential equation (7) relates directly to the original hypothesised model (2) and can be interpreted in these terms. Importantly, however, this model has been identified by a process direct inference from the experimental data without any prior assumptions about the model, other than it is in the form of a differential equation. This approach has been referred to as *Hypothetico-Inductive Data-Based Mechanistic* (HI-DBM) modelling (see e.g. Young 2013).

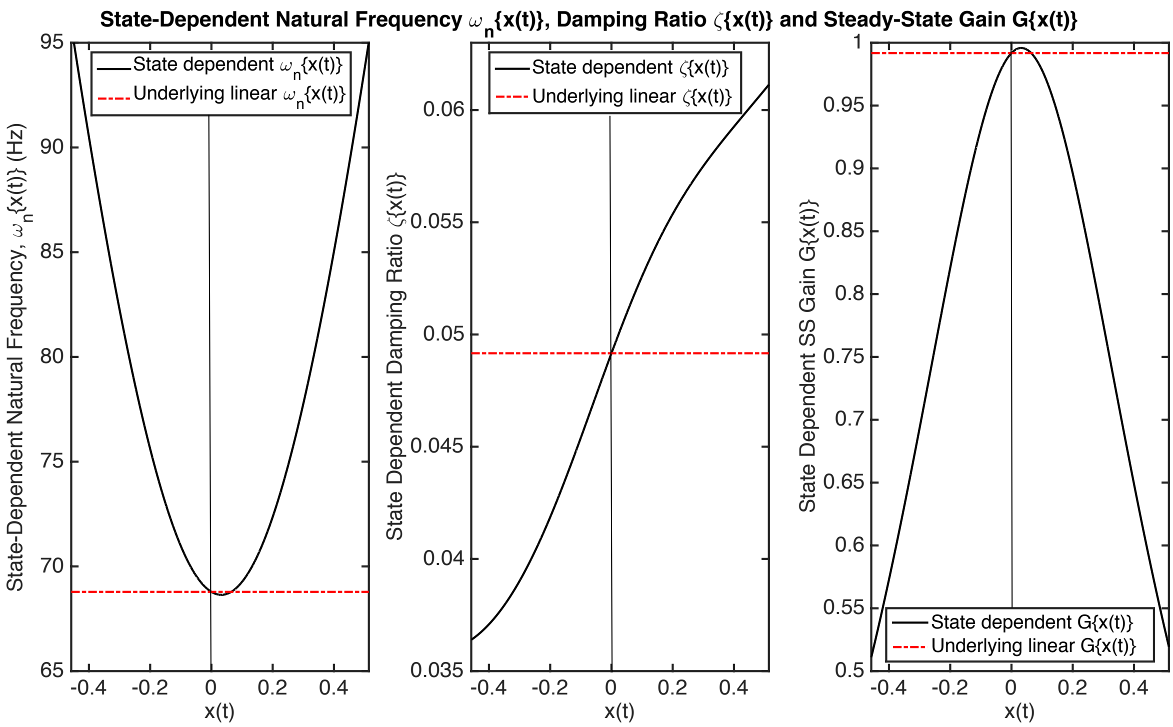

Figure 1. State dependent natural frequency (left panel), damping ratio (centre panel) and steady state gain (right panel) of the 8 parameter SDPTF model.

One interesting and useful interpretation of the SDPTF model is obtained by considering the state-dependent effect on specified physical properties of the system. For example, the natural frequency ω_n, damping ratio ζ and steady state gain G can be computed as for a linear system but defined by the SDPs and so dependent on the output $x(t)$. This is illustrated in Fig.1, which shows plots of the resulting state dependent characteristics $\omega_n\{x(t)\}$, $\zeta\{x(t)\}$ and $G\{x(t)\}$, based on the actual sampling interval of 1/2441 secs. The value of these properties at any value of $x(t)$ effectively provides a 'snapshot' of the system in these terms that describes the small perturbation behaviour of the system at this output level. Also shown by the red dash-dot lines are the underlying values of these properties defined by the linear part of the model. Extracting this kind of information would not be nearly so easy in the case of a discrete-time state space model and it would be even more difficult in the case of most full black-box models.

4 Conclusions

Different modelling objectives often demand different types of model, so there is rarely only one model of a dynamic system. This is certainly the case with the Silverbox system, where Marconato et al. (2012) list and discuss 8 models of different types ranging from two 'white box' models with 5 and 10 parameters, respectively, to 'grey-box' and 'black-box' models with between 12 and in excess of 700 parameters! Given that the silverbox is an electronic circuit constructed specially to mimic the behaviour of a single input, single output mechanical system, however, we assume that the most desirable model should be a nonlinear differential equation that can be related directly to the mechanical system model (2), which represents the original modelling hypothesis.

Such a mechanistically meaningful model can be identified in various ways. For instance, because the Silverbox circuit was developed in part to simulate a mechanical system, an identification methodology developed specially for such systems could be used. A companion technical note (Janot et al. 2016) uses this approach to model the Silverbox system, based on the 'inverse dynamic model' (IDM) identification methodology developed in Janot et al. (2017). The present comment has outlined an alternative data-based mechanistic modelling approach to the identification of a 'direct dynamic model' (DDM), where the specific objective is to obtain a state-dependent parameter differential equation that can be interpreted directly in terms of the originally hypothesised differential equation model form (2).

The resulting SDPTF models are not only efficiently parameterized and simple to understand, they are also interpretable directly in terms of this original model, as required. We believe that these models are, therefore, not only more consistent with the original aims behind the construction of the Silverbox circuit system, but are also in a form that is more immediately understandable and useful to many practicing engineers working on electro-mechanical systems than the discrete-time, state space model identified by Nöel and Schoukens (2018). Of course, this does not negate the more general utility of such an alternative model and the powerful modelling methodology that underlies its identification.

References

Janot, A., P. C. Young, and M. Gautier (2017). Identification and control of electro-mechanical systems using state-dependent parameter estimation. *International Journal of Control 90*(4), 643–660.

Janot, A., P. C. Young, and J. Noël (2016). On the identification of the silverbox system in a electromechanical context. Confidential Technical Report RT 1/25302 DCSD, Systems Control and Flight Dynamics, ONERA, Toulouse, France.

Ljung, L. (1987). *System Identification. Theory for the User*. Englewood Cliffs, NJ.: Prentice Hall.

Marconato, A., J. Sjöberg, J. Suykens, and J. Schoukens (2012). Identification of the silverbox benchmark using nonlinear state-space models. In *Proceedings of the 16th IFAC Symposium on System Identification*, Brussels, Belgium.

Mendel, J. M. (1969). A priori and a posteriori identification of time varying parameters. In *Prodeedings, 2nd Hawaii Conference on System Sciences.*

Nöel, J. P. and J. Schoukens (2018). Grey-box state-space identification of nonlinear mechanical vibrations. *International Journal of Control 91*, 1–22.

Paduart, J., L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. R. Pintelon (2010). Identification of nonlinear systems using polynomial nonlinear state space models. *Automatica 46*, 647–656.

Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Priestley, M. B. (1980). State-dependent models: A general approach to non-linear time series analysis. *Journal of Time Series Analysis 1*, 47–71.

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

Wigren, T. and J. Schoukens (2013). Three free data sets for development and benchmarking in nonlinear system identification. In *Proceedings 2013 European Control Conference*, Zurich, Switzerland, pp. 2933–2938.

Young, P. C. (1968). Adaptive pitch autostabilization of an air-surface missile,. In *Proc. Southwestern IEEE Conf.,San Antonio, Texas*, New York. IEEE.

Young, P. C. (1981). A second generation adaptive autostabilization system for airborne vehicles. *Automatica 17*, 459–470.

Young, P. C. (2000). Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In W. J. Fitzgerald, A. Walden, R. Smith, and P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing*, pp. 74–114. Cambridge University Press: Cambridge.

Young, P. C. (2001). The identification and estimation of nonlinear stochastic systems. In A. I. Mees (Ed.), *Nonlinear Dynamics and Statistics*, pp. 127–166. Birkhauser: Boston.

Young, P. C. (2011). *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Springer-Verlag, Berlin.

Young, P. C. (2013). Hypothetico-inductive data-based mechanistic modeling of hydrological systems. *Water Resources Research 49*(2), 915–935.

Young, P. C. (2018). Efficiently parameterized models of the Silverbox nonlinear circuit. Technical Report TN/PCY/5 (2018) (download from CAPTAIN Toolbox site), Lancaster Environment Centre, Lancaster University, UK.

Young, P. C., P. McKenna, and J. Bruun (2001). Identification of nonlinear stochastic systems by state dependent parameter estimation. *International Journal of Control 74*, 1837–1857.