The Brexitiers’ View of England’s Green & Pleasant Land:
Nostalgia for the Lost Empire of 1906?
State-Dependent Parameter Nonlinear Models
and a Hydrological Identification Benchmark

Peter Young
Lancaster University, U.K.  p.young@lancaster.ac.uk

***

With thanks to Alex Janot and Mathieu Brunot of ONERA
for their help in checking the presentation and demo m-file

Workshop on Nonlinear System Identification Benchmarks,
Liege 2018
My Background and Philosophical Approach

• One of my greatest pleasures is that I live in a multi-disciplinary world: academia; control and systems engineering; statistics; time-series analysis; forecasting; environmental science; and, most recently, electro-mechanical engineering, particularly the modelling and control of robotic systems.

• Predominantly, I am not a mathematician; rather I use mathematics to evolve computational methods that relate to the defined application objectives, most of which are concerned with the modelling, forecasting and control (including management) of systems in the natural, engineering and socio-economic worlds.

• My philosophy for doing this is enshrined in an approach I call Data-Based Mechanistic (DBM) modelling. This is an inductive method of modelling directly from time series data that tries to avoid prejudicial hypotheses and, while it exploits ‘black-box’ modelling techniques, also requires the resulting model to have a clear and scientifically acceptable mechanistic interpretation.
Data-Based Mechanistic (DBM) Modelling

- DBM modelling is a ‘method theory’, developed over many years. Its name emphasizes my contention that, while the model should be inferred from the analysis of data in an objective manner, it should also have an obvious mechanistic interpretation and be as simple as possible, consistent with the modelling objectives.

- In other words, I believe that a model should not just explain the time series data well, it should also provide a clear, transparent and scientifically acceptable mechanistic description of the system under investigation; a description that further enhances confidence in its ability to approximate reality in a meaningful manner and helps to achieve the application objectives.

- Note: ‘benchmark’ data sets, such as those we are discussing here, that emphasise only the modelling of the data, without necessarily considering of the objectives, are not consistent with the DBM philosophy.
In a very real sense, therefore, although DBM modelling exploits some of the methodology used in ‘black-box’ modelling, it is partly a reaction against the notion and, I believe, over-use of pure black-box models particularly when they are being used in research that is attempting to investigate the real nature of dynamic systems in the natural and man-made world.

In the case of nonlinear systems, this means that the nonlinearities in the model should be clearly identified in some form, such as a graphical portrayal, that makes physical sense to practitioners in the scientific or engineering discipline in which the model is being used.

One nonlinear model identification procedure that facilitates this is State-Dependent Parameter (SDP) estimation, an approach evolved many years ago (Young, 1968; Mendel, 1969; Young, 1981) and Priestley (1980), who first used the name. The SDP methods used in the later example were developed in the 1990s (see e.g. Young, 2000, 2001b) and these have been applied successfully to a large number of practical systems, in diverse areas of study, since then.
State-Dependent Parameter (SDP) Models

An SDP example that will be well known to you is the model of the ‘Silverbox’. My analysis of these data (Young, 2016) suggests that an acceptable system model is the following second order, nonlinear differential equation:

\[
\frac{d^2x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + k_1 x(t) + k_2 x^2(t) + k_3 x^3(t) = b_0 u(t)
\]

or, in SDP form:

\[
\frac{d^2x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + a_2 \{x(t)\} \cdot x(t) = b_0 u(t)
\]

where

\[
a_2 \{x(t)\} = k_1 + k_2 x(t) + k_3 x^2(t)
\]

The associated SDP Transfer Function (SDPTF) model is represented as follows:

\[
x(t) = \frac{b_0}{s^2 + a_1 s + a_2 \{x(t)\}} u(t)
\]

This explains 99.999% of the variance in silverbox output \(R_T^2 = 0.999999\), which should be good enough for almost all applications.
General SDP Transfer Function Model

The simplest SISO discrete-time SDPTF model takes the following form:

\[ x(k) = \frac{B(z^{-1}, v(k))}{A(z^{-1}, v(k))} u(k - \delta); \quad y(k) = x(k) + \xi(k) \]

\[ A(z^{-1}, v(k)) = 1 + a_1\{v(1, k)\}z^{-1} + a_2\{v(2, k)\}z^{-2} + \ldots + a_n v(n, k) \]

\[ B(z^{-1}, v(k)) = b_0\{v(n + 1, k)\} + b_1\{v(n + 2, k)\}z^{-1} + \ldots + b_m\{v(n + m + 1, k)\}z^{-m} \]

and \( v_t \) is a vector of measured variables (states) on which the parameters may be dependent. Or, in SDARX estimation equation terms:

\[ y(k) = z(k)^T p(k) + e(k); \quad \text{where} \quad p(k) = [a_1\{v_{1,t}\} \ldots b_m\{v_{n+m+1,k}\}]^T, \]

\[ z_t^T = [-y(k-1) \ldots - y(k-n) \ u(k-\delta) \ldots u(k-\delta-m)] \]

The continuous-time SDP model is defined in a similar manner.
IDENTIFICATION and ESTIMATION

1. Initial linear and Time-Variable Parameter (TVP) model identification, to establish whether the model is TVP or SDP, using the CAPTAIN Toolbox discrete-time rivbjid, dtfmopt, dtfm routines; and, if appropriate, the continuous-time rivcbjid routine (see later example and Young, 2011, 2015).

2. Initial **Non-Parametric Identification** of the nonlinear model structure (i.e. in the sense of the location and nature of the statistically significant SDP nonlinearities) using the CAPTAIN sdp routine. This is then followed by:

3. **Parameterization** of the SDP nonlinearities using whatever method is most appropriate, if possible one that is transparent and has a physical interpretation.

4. **Final Parameter Estimation**: statistically optimal estimation applied to this parameterized model (e.g. using NLS or PEM optimization, often applied to a Simulink version of the model, as in the later example).
Non-Parametric SDP Identification Algorithm

1. Estimate the starting parameters $p_{i,k}^0, i = 1, 2, ..., n + m + 1$ in the above SDARX model using LS (i.e. constant LS ARX parameter estimates);

2. Backfitting Algorithm: Iterate, $i - 1, 2, ..., n + m + 1; t = 1, 2, ..., t_c$
   (i) form the modified 'dependent' variable $y_k^i = y_k - \sum_{j \neq i} z_{j,k} \hat{p}_{j,k|N}^t$;
   (ii) sort both $y_k^i$ and $z_{i,k}$ according to the ascending order of the state $s_{i,k}$ associated with $z_{i,k}$;
   (iii) obtain an ML optimized Fixed Interval Smoothing (FIS) estimate $\hat{p}_{i,k|N}^t$ of $p_{i,k}$ in the modified single state dependent variable relationship $y_k^i = p_{i,k} \cdot z_{i,k}$

3. Continue 2 until iteration $t = t_c$, when the individual SDPs (which are each time series of length $N$) have not changed significantly according to some chosen criterion. Note: this algorithm produces a limited SDP ARX model.
Parametric SDP TF Model Optimization

This can be carried out in two stages, although the first is not essential:

1. Initial nonlinear model estimation using a multiple input, quasi-linear model estimated by CAPTAIN rivcbj or rivbj routines\(^1\), where the additional inputs are the parameterized nonlinear functions (possibly a series of ‘basis functions’).

2. Final model optimization using NLS or PEM, initialised with the parameters from stage 1. I normally optimize the parameters in a Simulink model, which then serves for subsequent forecasting and control/management studies.

The following example (see also the supplied Matlab command-line Demo.), illustrates the identification strategy outlined here and in the previous slides.

\(^1\)The CAPTAIN Toolbox is available free via [http://captaintoolbox.co.uk/Captain_Toolbox.html](http://captaintoolbox.co.uk/Captain_Toolbox.html)
A Hydrological Benchmark Example: DBM Modelling of Rainfall-Flow

- A important environmental topic is the modelling and forecasting of flow or level changes in a river system on the basis of upstream rainfall measurements.

- It is well known that the relationship between rainfall and flow is nonlinear, although the nature of this nonlinearity is still a topic of research.

- The benchmark example considered here involves daily rainfall and flow measurements made on the ‘ephemeral’ River Canning in Western Australia (i.e. a river that stops flowing in Summer: see also Young, 2008).

- I will outline the sequential stages of DBM modelling based entirely on these data, with no prior hypotheses about the nature of the model form and structure, other than that it can be described by a linear or nonlinear transfer function in discrete or continuous-time form.
The Canning River, WA: Location
The Canning River When Flowing
The Best Linear Models: All Unacceptable
Off-Line Time Variable Parameter Estimation

- Off-line estimation of time variable parameters is more powerful than the normal on-line ‘filtering’ (RLS, RPEM RRIV) methods because it can exploit Fixed Interval Smoothing to yield zero lag, lower variance TVP estimates.

- The dtfmopt and dtfm routines in CAPTAIN use this approach for discrete-time models. In the present example the simplest first order \([1 \ 1 \ 1]\), discrete-time TVP model is considered:

\[
y(k) = \frac{b_1(k)}{1 + a_1(k)} u(k - 1) + \xi(k)
\]

- The first results (see next slide) are obtained allowing both parameters to vary. However, this suggests that only \(b_1(k)\) varies significantly, so the subsequent TVP estimation is based on this, with \(a_1\) assumed to be constant.
DTFM: Time Variable Parameter Estimation

\begin{align*}
\text{TVP Estimate } a_1(t) & \\
& \begin{array}{cccc}
0.87 & 0.865 & 0.86 & 0.855 & 0.85 & 0.845 & 0.84 & 0.835 & 0.83 \\
\end{array} \\
\text{TVP Estimate } b_1(t) & \\
& \begin{array}{cccc}
0 & 0.01 & 0.02 & 0.03 & 0.04 & 0.05 & 0.06 & 0.07 & 0.08 \\
\end{array}
\end{align*}

\text{Date: } 1985.5, 1986, 1986.5, 1987
TVP Model with $b_1(k)$ Varying; $a_1$ Constant
**TVP $b_1(t)$ Highly Correlated with Flow**

Cross Correlation Between TVP Estimate and Flow

---

Cross Correlation Between TVP Estimate and Flow

---
State-Dependent Parameter Nonlinear Model: Non-parametric (NP) Identification

- Based on a comparison $[1\ 1\ 1]$ and $[1\ 1\ 0]$ SDP models and later results, the non-parametric SDP estimation is based on the $[1\ 1\ 0]$ SDP model with an SDP input parameter $b_0(y(t))$ and constant $a_1$ parameter.

- The next plot compares the SDP estimates with those obtained by ‘free’ estimation with both parameters considered as functions of $y(t)$.

- This confirms that $b_0(y(t))$ is the most significant SDP parameter with $-\hat{a}_1$ estimated as constant at 0.894. This is a ‘Hammerstein’ Nonlinear Model, i.e. with only an input nonlinearity.

- **NB:** this SDP estimation is mainly for nonlinear structure identification purposes. The NP estimates can be used directly (e.g. as a look-up table) but are mainly a step to the final parametric stage of SDP model estimation.
Initial Non-Parametric SDP Estimates

Comparison of Free (black) and Constrained (red) SDP estimates
There are various ways in which the SDP $b_0 \{ y(t) \}$, shown in the right hand panel of the previous plot, could be parameterized (see e.g. Beven et al., 2011) but I always feel that the simpler and more physically plausible it is, the better.

This is because complex ‘black box’ parameterizations are often more difficult and dangerous to extrapolate into regions not explored during SDP estimation (or any other nonlinear modelling, for that matter) based on limited data.

In this case simple power (i.e. $b_0 = y(t)^\gamma$) and exponential (i.e. $b_0 = 1 - exp\{ -\gamma y(t) \}$) laws, defined by only a single parameter, yield good results and are acceptable on hydrological grounds. However, it is clear that other simple, or a little more complex, parameterizations are possible.
• At this stage, it should be recalled that a simple $[1 \ 1 \ 0]$ model was used to explore non-parametric SDP possibilities. But now, in the final stage of SDP estimation, it makes sense to evaluate whether other, higher order, models may provide a superior explanation of the flow changes.

• This is made simpler by the Hammerstein nature of the identified SDP model because the ‘effective’ rainfall input can be defined as $u_e(t) = b_0 \{ y(t) \} u(t)$ and then this can be used as the input for linear estimation using either DT or CT models. In this case, the former is identified as a $[2 \ 3 \ 0 \ 5 \ 0]$ linear model.

• I have estimated the final SDP model with $u_e(t) = \{ y(t)^\gamma \} u(t)$ incorporated into a lsqnonlin optimization routine, where the rivbj routine is used for DT linear model estimation based this $u_e(t)$ (later, I consider alternative CT model estimation of the same kind using the rivcbj routine).

• The plot below compares the $\hat{b}_0 \{ y(t) \}$, estimated in this way, with the non-parametric estimate and the estimate obtained with an exponential SDP law.
Non-Parametric and Parametric SDP Estimates

SDP Estimation of Effective Rainfall Coefficient

- Confidence interval
- Nonparametric (NP) estimate
- NP estimate: exact location
- Exponential law
- Power law

Flow (acting as surrogate for catchment storage)
DBM Nonlinear DT Model Response: 1985-1987

DBM Model Estimation: 1985.2-1987.1: 95.6% of Flow Explained

Note: the residuals of the AR(5) noise model are serially uncorrelated and not significantly correlated with the input signal or the temperature series, as required.
First Validation: 1977-1978

Predictive Validation: 1977-1978.5: 95.1% of Flow Explained
Second Validation: 1979-1980

Cross-Validation: 1979-1980.5: 92.6% of Flow Explained

Date


Flow (cumecs)

0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Model
Measured
The Modelling Objective: Explaining the Catchment Characteristics

- The optimized SDP model has no time-delay and the rainfall is very difficult to forecast into the future, so it is not particularly useful as a forecasting tool (although the introduction of a ‘false’ time delay would allow this).

- Its main use is as an indicator of the catchment dynamic behaviour and associated hydrological interpretation. A Continuous-Time Model is more appropriate for this: it is (i) unique (not dependent on the sampling interval) and (ii) immediately interpretable in physically meaningful terms.

- As a final exercise, therefore, the model is estimated in a continuous time stochastic form, using the same lsqnonlin optimization routine but with a CT model. The DBM model obtained in this manner is shown on the next slide.
Continuous-Time State-Dependent Parameter Nonlinear Stochastic Model

\[ x(t) = \frac{-0.06459s^2 + 0.2473s + 0.0214}{s^2 + 0.439s + 0.02084}ue(t) \]

\[ ue(t) = \{y^s(t)^{0.8}\}u(t) \]

\[ \xi(k) = \frac{1}{1 - 0.8674z^{-1} + 0.3238z^{-2} - 0.1670z^{-3}}e(k); \]

\[ e(k) = N(0, \sigma^2\{y(k)\}), \text{ where } \sigma^2\{y(k)\} \text{ is an SD function of } y(k) \]

\[ y(k) = x(k) + \xi(k) \quad R^2_T = 0.958 \]

Note: \( y^s(t) \) denotes here \( y(t) \) as a ‘surrogate’ for catchment wetness (CW) and is not a feedback of \( y(t) \). Consequently, the model cannot be used for \textbf{on-line} simulation unless \( y^s(t) \) is replaced by some estimate of CW based, for example, on rainfall and/or temperature: see ‘HI-DBM modelling’, Young (2003, 2013).
Credibility Demonstrated by Model Decomposition

**Quick Surface Processes**
- First Order TF with Residence Time
  \[ T_q = 2.6 \text{ days} \]
  \[
  \frac{0.252}{s + 0.385}
  \]

**Instantaneous Effect**
- Negative Gain \(-0.065\) suggesting Losses in system

**Sub-Surface Processes**
- First Order TF with Residence Time
  \[ T_s = 18.47 \text{ days} \]
  \[
  \frac{0.0236}{s + 0.0541}
  \]

**AR(3) Noise Model**
\[
\frac{1}{1-0.867z^{-1} + 0.324z^{-2} - 0.167z^{-3}}
\]

**Measured Noisy Flow Output**
\[ e(k) \]

**Heteroscedastic White Noise**
\[ \xi(k) \]

**Measured Rainfall Input**
\[ u(k) \]

**Effective Rainfall SDP Nonlinearity**

**Effective Rainfall**
\[ u_e(k) \]

**Sampler**
\[ x(k) \]

**Noise-Free Output**
\[ x(t) \]

**Xq(t) Xs(t) XI(t) + x(k) y(k) \]
Monte Carlo Uncertainty Analysis

- Given the above stochastic model, Monte Carlo uncertainty analysis can be used for estimating the uncertainty in parameters and the model response.

- In particular, this allows us to evaluate the uncertainty in the derived, physically meaningful parameters (the residence times and partition percentages) and the uncertainty bounds on the model response. This is in the form of the histograms obtained from the stochastic realizations of the derived parameters; and the 95 percentile uncertainty bounds on the response.

- One problem with the latter computation in this example is the heteroscedasticity in the estimated noise sequence $\xi(k)$, whose variations are clearly proportional to the variations in the flow. This is handled by estimating a linear relationship between FIS estimated variations in the standard deviations of the final residuals $e(k)$ and the flow $y(k)$; effectively defining a state-dependent white noise input with variance $\sigma^2\{y(k)\}$. 
MCS Analysis: Residence Time Distributions

Slow Residence Time (hours)

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
<tr>
<td>250</td>
</tr>
<tr>
<td>300</td>
</tr>
</tbody>
</table>

Mean Estimate = 18.5 days

Quick Residence Time (hours)

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>150</td>
</tr>
<tr>
<td>200</td>
</tr>
<tr>
<td>250</td>
</tr>
<tr>
<td>300</td>
</tr>
</tbody>
</table>

Mean Estimate = 2.6 days
MCS Analysis: Partition % Distributions

Groundwater Partition (%)

Surface Partition (%)

Mean Estimate=42.5%

Mean Estimate=63.8%
MCS Analysis: Uncertainty in Model Response

Date
Flow (cumecs)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
95% confidence bounds
Model output
Measured Flow
Zoom into Section of Previous Plot
‘Transparent SDP’ or ‘Black Box’ Models

• An important aspect of DBM modelling is the definition of the modelling objectives. I feel the concentration on the identification of overly-complex black-box nonlinear models that can explain the data too well, without evaluation of how they function in terms of these modelling objectives, is questionable and needs to be justified in practical terms.

• Based on experience in multiple different disciplines, I believe that many (most?) scientists and engineers mistrust totally ‘black-box’ models because they are unable to visualise and interpret the nature of the nonlinearities.

• They also feel that the models can be over-parameterised, as typified by the published research using data from experiments on a hydraulic actuator controlling a robot arm (see e.g. Sjöberg et al., 1995; Hu et al., 2001), where most models appear to be totally black box, with too many parameters.
• The SDP Hammerstein model of these robot data (see Young, 2001a, 2006b) has only 15 parameters: a 10 parameter radial basis function model for the nonlinearity and 5 in the [3 2 1] linear TF. The model has reasonable explanatory power, cf other models, with $R_T^2 = 0.96$ (error RMS=0.318).

• Finally, there is evidence that the system is not only nonlinear but also non-stationary so that an adaptive SDP model, in which just the parameters of the linear component in the Hammerstein model parameters are allowed to vary, might well be required. An adaptive extension is quite straightforward because of the model’s simplicity. This is discussed in Young (2006b) where the dtfmo/dtfm routines are applied in on-line ‘filtering’ mode, producing an adaptive model that shows a significant improvement in validation terms.

• Even without this, the validation error for the constant parameter SDP model is marginally less than Hu et al’s 102 parameter black box ‘Quasi-ARMAX’ model and, in contrast to the black-box model, it provides a graphical visualisation that makes rather obvious physical sense, as we see on the next slide.
A ‘Simple’ but Flawed Alternative Approach to SDP Model Estimation

- As mentioned previously, an alternative method of SDP TF model estimation is to use standard rivbj or rivcbj estimation with additional, nonlinearly defined ‘basis function’ inputs. While inherently flawed (e.g. biased basis function parameter estimates; can be over-parameterized; etc.), it can be helpful.

- In this rainfall-flow example, an exponential law can be represented as a series expansion in a finite number of terms. This produces reasonable CT model results with five terms in the expansion, yielding a model with better, but possibly misleading, explanatory power of $R_T^2 = 0.966$ (see the Matlab Demo).

- However, this 5-input TF model, with the additional TFs having first and second order numerator polynomials, and the common denominator polynomial, has 12 parameters, twice as many as the 6 parameters of the simple exponential model; and extrapolation of the resulting input nonlinearity is questionable.
Concluding Comments

- The DBM approach to system identification is to identify the simplest, mechanistically transparent model that satisfies the defined objectives (normally connected with scientific understanding, forecasting or control/management system design).

- In the multi-disciplines where I work, the scientists, practicing engineers and economists are not normally impressed by ‘black-box' models, no matter how well they perform, unless these can be interpreted in a manner that is credible to them. In other words, a nonlinear SDP model needs to have nonlinearities that, wherever possible, have a clear and physically meaningful form.

- I have found that SDP nonlinear models are widely applicable and can normally satisfy such scientific and practical engineering requirements, including nonlinear SDP control (see e.g. Young, 2001b, 2006b,a; Janot et al., 2017).

- The current SDP algorithms need to be improved: (e.g. a Multi-State Dependent Parameter (MSDP) tool is in development (Sadeghi et al., 2010; Mindham et al., 2018)).
References


